

Third Edition

Sponsored by the Optical Society of America

HANDBOOK OF OPTICS

Volume V

*Atmospheric Optics, Modulators, Fiber Optics,
X-Ray and Neutron Optics*



Editor-in-Chief:
Michael Bass

Associate Editors:
Casimer M. DeCusatis
Jay M. Enoch
Vasudevan Lakshminarayanan
Guifang Li
Carolyn MacDonald
Virendra N. Mahajan
Eric Van Stryland

OSA[®]

HANDBOOK OF OPTICS

ABOUT THE EDITORS

Editor-in-Chief: Dr. Michael Bass is professor emeritus at CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida.

Associate Editors:

Dr. Casimer M. DeCusatis is a distinguished engineer and technical executive with IBM Corporation, Poughkeepsie, New York.

Dr. Jay M. Enoch is dean emeritus and professor at the School of Optometry at the University of California, Berkeley.

Dr. Vasudevan Lakshminarayanan is professor of Optometry, Physics, and Electrical Engineering at the University of Waterloo, Ontario, Canada.

Dr. Guifang Li is a professor at CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida.

Dr. Carolyn MacDonald is professor and chair of physics at the University at Albany, SUNY, and the director of the Albany Center for X-Ray Optics.

Dr. Virendra N. Mahajan is a distinguished scientist at The Aerospace Corporation.

Dr. Eric Van Stryland is a professor at CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida.

HANDBOOK OF OPTICS

Volume V

Atmospheric Optics, Modulators, Fiber Optics, X-Ray and Neutron Optic

THIRD EDITION

Sponsored by the
OPTICAL SOCIETY OF AMERICA

Michael Bass Editor-in-Chief

*CREOL, The College of Optics and Photonics
University of Central Florida
Orlando, Florida*

Carolyn MacDonald Associate Editor

*Department of Physics
University at Albany
Albany, New York*

Guifang Li Associate Editor

*CREOL, The College of Optics and Photonics
University of Central Florida
Orlando, Florida*

Casimer M. DeCusatis Associate Editor

*IBM Corporation
Poughkeepsie, New York*

Virendra N. Mahajan Associate Editor

*Aerospace Corporation
El Segundo, California*



New York Chicago San Francisco Lisbon London Madrid
Mexico City Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

Copyright © 2010 by The McGraw-Hill Companies, Inc. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-163314-7

MHID: 0-07-163314-6

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-163313-0, MHID: 0-07-163313-8.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative please e-mail us at bulksales@mcgraw-hill.com.

Information contained in this work has been obtained by The McGraw-Hill Companies, Inc. ("McGraw-Hill") from sources believed to be reliable. However, neither McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. ("McGraw-Hill") and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

COVER ILLUSTRATIONS

Boadband supercontinuum. Generated in a photonic crystal fiber using a mode-locked Ti:Sapphire laser as pump source. The spectrum is much broader than seen in the photograph, extending from 400 nm to beyond 2 μm . (*Photo courtesy of the Optoelectronics Group, University of Bath.*)

Supernova remnant. A Chandra X-Ray Space Telescope image of the supernova remnant G292.0+1.8. The colors in the image encode the X-ray energies emitted by the supernova remnant; the center of G292.0+1.8 contains a region of high energy X-ray emission from the magnetized bubble of high-energy particles that surrounds the pulsar, a rapidly rotating neutron star that remained behind after the original, massive star exploded. (*This image is from NASA/CXC/Penn State/S.Park et al. and more detailed information can be found on the Chandra website: <http://chandra.harvard.edu/photo/2007/g292/>.*)

Crab Nebula. A Chandra X-Ray Space Telescope image of the Crab Nebula—the remains of a nearby supernova explosion first seen on Earth in 1054 AD. At the center of the bright nebula is a rapidly spinning neutron star, or pulsar, that emits pulses of radiation 30 times a second. (*This image is from NASA/CXC/ASU/J.Hester et al. and more detailed information can be found on the Chandra website: <http://chandra.harvard.edu/photo/2002/0052/>.*)

This page intentionally left blank

CONTENTS

Contributors	xix
Brief Contents of All Volumes	xxiii
Editors' Preface	xxix
Preface to Volume V	xxx
Glossary and Fundamental Constants	xxxiii

Part 1. Measurements

Chapter 1. Scatterometers	<i>John C. Stover</i>	1.3

- 1.1 Glossary / 1.3
- 1.2 Introduction / 1.3
- 1.3 Definitions and Specifications / 1.5
- 1.4 Instrument Configurations and Component Descriptions / 1.7
- 1.5 Instrumentation Issues / 1.11
- 1.6 Measurement Issues / 1.13
- 1.7 Incident Power Measurement, System Calibration, and Error Analysis / 1.14
- 1.8 Summary / 1.16
- 1.9 References / 1.16

Chapter 2. Spectroscopic Measurements	<i>Brian Henderson</i>	2.1

- 2.1 Glossary / 2.1
- 2.2 Introductory Comments / 2.2
- 2.3 Optical Absorption Measurements of Energy Levels / 2.2
- 2.4 The Homogeneous Lineshape of Spectra / 2.13
- 2.5 Absorption, Photoluminescence, and Radiative Decay Measurements / 2.19
- 2.6 References / 2.24

Part 2. Atmospheric Optics

Chapter 3. Atmospheric Optics	<i>Dennis K. Killinger, James H. Churnside, and Laurence S. Rothman</i>	3.3

- 3.1 Glossary / 3.3
- 3.2 Introduction / 3.4
- 3.3 Physical and Chemical Composition of the Standard Atmosphere / 3.6
- 3.4 Fundamental Theory of Interaction of Light with the Atmosphere / 3.11
- 3.5 Prediction of Atmospheric Optical Transmission: Computer Programs and Databases / 3.22
- 3.6 Atmospheric Optical Turbulence / 3.26
- 3.7 Examples of Atmospheric Optical Remote Sensing / 3.36
- 3.8 Meteorological Optics / 3.40
- 3.9 Atmospheric Optics and Global Climate Change / 3.43
- 3.10 Acknowledgments / 3.45
- 3.11 References / 3.45

Chapter 4. Imaging through Atmospheric Turbulence 4.1

Virendra N. Mahajan and Guang-ming Dai

- Abstract / 4.1
- 4.1 Glossary / 4.1
- 4.2 Introduction / 4.2
- 4.3 Long-Exposure Image / 4.3
- 4.4 Kolmogorov Turbulence and Atmospheric Coherence Length / 4.7
- 4.5 Application to Systems with Annular Pupils / 4.10
- 4.6 Modal Expansion of Aberration Function / 4.17
- 4.7 Covariance and Variance of Expansion Coefficients / 4.20
- 4.8 Angle of Arrival Fluctuations / 4.23
- 4.9 Aberration Variance and Approximate Strehl Ratio / 4.27
- 4.10 Modal Correction of Atmospheric Turbulence / 4.28
- 4.11 Short-Exposure Image / 4.31
- 4.12 Adaptive Optics / 4.35
- 4.13 Summary / 4.36
- 4.14 Acknowledgments / 4.37
- 4.15 References / 4.37

Chapter 5. Adaptive Optics 5.1

Robert Q. Fugate

- 5.1 Glossary / 5.1
- 5.2 Introduction / 5.2
- 5.3 The Adaptive Optics Concept / 5.2
- 5.4 The Nature of Turbulence and Adaptive Optics Requirements / 5.5
- 5.5 AO Hardware and Software Implementation / 5.21
- 5.6 How to Design an Adaptive Optical System / 5.38
- 5.7 Acknowledgments / 5.46
- 5.8 References / 5.47

PART 3. Modulators

Chapter 6. Acousto-Optic Devices 6.1

I-Cheng Chang

- 6.1 Glossary / 6.3
- 6.2 Introduction / 6.4
- 6.3 Theory of Acousto-Optic Interaction / 6.5
- 6.4 Acousto-Optic Materials / 6.16
- 6.5 Acousto-Optic Deflector / 6.22
- 6.6 Acousto-Optic Modulator / 6.31
- 6.7 Acousto-Optic Tunable Filter / 6.35
- 6.8 References / 6.45

Chapter 7. Electro-Optic Modulators 7.1

Georgeanne M. Purvinis and Theresa A. Maldonado

- 7.1 Glossary / 7.1
- 7.2 Introduction / 7.3
- 7.3 Crystal Optics and the Index Ellipsoid / 7.3
- 7.4 The Electro-Optic Effect / 7.6
- 7.5 Modulator Devices / 7.16
- 7.6 Applications / 7.36
- 7.7 Appendix: Euler Angles / 7.39
- 7.8 References / 7.40

Chapter 8. Liquid Crystals 8.1

Sebastian Gauza and Shin-Tson Wu

- Abstract / 8.1
- 8.1 Glossary / 8.1

8.2	Introduction to Liquid Crystals /	8.2
8.3	Types of Liquid Crystals /	8.4
8.4	Liquid Crystals Phases /	8.8
8.5	Physical Properties /	8.13
8.6	Liquid Crystal Cells /	8.25
8.7	Liquid Crystals Displays /	8.29
8.8	Polymer/Liquid Crystal Composites /	8.36
8.9	Summary /	8.37
8.10	References /	8.38
8.11	Bibliography /	8.39

Part 4. Fiber Optics

Chapter 9. Optical Fiber Communication Technology and System Overview *Ira Jacobs* 9.3

9.1	Introduction /	9.3
9.2	Basic Technology /	9.4
9.3	Receiver Sensitivity /	9.8
9.4	Bit Rate and Distance Limits /	9.12
9.5	Optical Amplifiers /	9.13
9.6	Fiber-Optic Networks /	9.14
9.7	Analog Transmission on Fiber /	9.15
9.8	Technology and Applications Directions /	9.17
9.9	References /	9.17

Chapter 10. Nonlinear Effects in Optical Fibers *John A. Buck* 10.1

10.1	Key Issues in Nonlinear Optics in Fibers /	10.1
10.2	Self- and Cross-Phase Modulation /	10.3
10.3	Stimulated Raman Scattering /	10.4
10.4	Stimulated Brillouin Scattering /	10.7
10.5	Four-Wave Mixing /	10.9
10.6	Conclusion /	10.11
10.7	References /	10.12

Chapter 11. Photonic Crystal Fibers *Philip St. J. Russell and Greg J. Pearce* 11.1

11.1	Glossary /	11.1
11.2	Introduction /	11.2
11.3	Brief History /	11.2
11.4	Fabrication Techniques /	11.4
11.5	Modeling and Analysis /	11.6
11.6	Characteristics of Photonic Crystal Cladding /	11.7
11.7	Linear Characteristics of Guidance /	11.11
11.8	Nonlinear Characteristics of Guidance /	11.22
11.9	Intrafiber Devices, Cutting, and Joining /	11.26
11.10	Conclusions /	11.28
11.11	Appendix /	11.28
11.12	References /	11.28

Chapter 12. Infrared Fibers *James A. Harrington* 12.1

12.1	Introduction /	12.1
12.2	Nonoxide and Heavy-Metal Oxide Glass IR Fibers /	12.3
12.3	Crystalline Fibers /	12.7
12.4	Hollow Waveguides /	12.10
12.5	Summary and Conclusions /	12.13
12.6	References /	12.13

Chapter 13. Sources, Modulators, and Detectors for Fiber Optic Communication Systems *Elsa Garmire* 13.1

- 13.1 Introduction / 13.1
- 13.2 Double Heterostructure Laser Diodes / 13.3
- 13.3 Operating Characteristics of Laser Diodes / 13.8
- 13.4 Transient Response of Laser Diodes / 13.13
- 13.5 Noise Characteristics of Laser Diodes / 13.18
- 13.6 Quantum Well and Strained Lasers / 13.24
- 13.7 Distributed Feedback and Distributed Bragg Reflector Lasers / 13.28
- 13.8 Tunable Lasers / 13.32
- 13.9 Light-Emitting Diodes / 13.36
- 13.10 Vertical Cavity Surface-Emitting Lasers / 13.42
- 13.11 Lithium Niobate Modulators / 13.48
- 13.12 Electroabsorption Modulators / 13.55
- 13.13 Electro-Optic and Electrorefractive Modulators / 13.61
- 13.14 PIN Diodes / 13.63
- 13.15 Avalanche Photodiodes, MSM Detectors, and Schottky Diodes / 13.71
- 13.16 References / 13.74

Chapter 14. Optical Fiber Amplifiers *John A. Buck* 14.1

- 14.1 Introduction / 14.1
- 14.2 Rare-Earth-Doped Amplifier Configuration and Operation / 14.2
- 14.3 EDFA Physical Structure and Light Interactions / 14.4
- 14.4 Other Rare-Earth Systems / 14.7
- 14.5 Raman Fiber Amplifiers / 14.8
- 14.6 Parametric Amplifiers / 14.10
- 14.7 References / 14.11

Chapter 15. Fiber Optic Communication Links (Telecom, Datacom, and Analog) *Casimer DeCusatis and Guifang Li* 15.1

- 15.1 Figures of Merit / 15.2
- 15.2 Link Budget Analysis: Installation Loss / 15.6
- 15.3 Link Budget Analysis: Optical Power Penalties / 15.8
- 15.4 References / 15.18

Chapter 16. Fiber-Based Couplers *Daniel Nolan* 16.1

- 16.1 Introduction / 16.1
- 16.2 Achromaticity / 16.3
- 16.3 Wavelength Division Multiplexing / 16.4
- 16.4 $1 \times N$ Power Splitters / 16.4
- 16.5 Switches and Attenuators / 16.4
- 16.6 Mach-Zehnder Devices / 16.4
- 16.7 Polarization Devices / 16.5
- 16.8 Summary / 16.6
- 16.9 References / 16.6

Chapter 17. Fiber Bragg Gratings *Kenneth O. Hill* 17.1

- 17.1 Glossary / 17.1
- 17.2 Introduction / 17.1
- 17.3 Photosensitivity / 17.2
- 17.4 Properties of Bragg Gratings / 17.3
- 17.5 Fabrication of Fiber Gratings / 17.4
- 17.6 The Application of Fiber Gratings / 17.8
- 17.7 References / 17.9

Chapter 18. Micro-Optics-Based Components for Networking **18.1** *Joseph C. Palais*

- 18.1 Introduction / 18.1
- 18.2 Generalized Components / 18.1
- 18.3 Network Functions / 18.2
- 18.4 Subcomponents / 18.5
- 18.5 Components / 18.9
- 18.6 References / 18.12

Chapter 19. Semiconductor Optical Amplifiers **19.1** *Jay M. Wiesenfeld and Leo H. Spiekman*

- 19.1 Introduction / 19.1
- 19.2 Device Basics / 19.2
- 19.3 Fabrication / 19.15
- 19.4 Device Characterization / 19.17
- 19.5 Applications / 19.22
- 19.6 Amplification of Signals / 19.22
- 19.7 Switching and Modulation / 19.28
- 19.8 Nonlinear Applications / 19.29
- 19.9 Final Remarks / 19.36
- 19.10 References / 19.36

Chapter 20. Optical Time-Division Multiplexed Communication Networks **20.1** *Peter J. Delfyett*

- 20.1 Glossary / 20.1
- 20.2 Introduction / 20.3
- 20.3 Multiplexing and Demultiplexing / 20.3
- 20.4 Introduction to Device Technology / 20.12
- 20.5 Summary and Future Outlook / 20.24
- 20.6 Bibliography / 20.25

Chapter 21. WDM Fiber-Optic Communication Networks **21.1** *Alan E. Willner, Changyuan Yu, Zhongqi Pan, and Yong Xie*

- 21.1 Introduction / 21.1
- 21.2 Basic Architecture of WDM Networks / 21.4
- 21.3 Fiber System Impairments / 21.13
- 21.4 Optical Modulation Formats for WDM Systems / 21.27
- 21.5 Optical Amplifiers in WDM Networks / 21.37
- 21.6 Summary / 21.44
- 21.7 Acknowledgments / 21.44
- 21.8 References / 21.44

Chapter 22. Solitons in Optical Fiber Communication Systems **22.1** *Pavel V. Mamyshev*

- 22.1 Introduction / 22.1
- 22.2 Nature of the Classical Soliton / 22.2
- 22.3 Properties of Solitons / 22.4
- 22.4 Classical Soliton Transmission Systems / 22.5
- 22.5 Frequency-Guiding Filters / 22.7
- 22.6 Sliding Frequency-Guiding Filters / 22.8
- 22.7 Wavelength Division Multiplexing / 22.9
- 22.8 Dispersion-Managed Solitons / 22.12
- 22.9 Wavelength-Division Multiplexed Dispersionmanaged Soliton Transmission / 22.15
- 22.10 Conclusion / 22.17
- 22.11 References / 22.17

Chapter 23. Fiber-Optic Communication Standards **23.1**
Casimer DeCusatis

- 23.1 Introduction / 23.1
- 23.2 ESCON / 23.1
- 23.3 FDDI / 23.2
- 23.4 Fibre Channel Standard / 23.4
- 23.5 ATM/SONET / 23.6
- 23.6 Ethernet / 23.7
- 23.7 Infiniband / 23.8
- 23.8 References / 23.8

Chapter 24. Optical Fiber Sensors **24.1**
Richard O. Claus, Ignacio Matias, and Francisco Arregui

- 24.1 Introduction / 24.1
- 24.2 Extrinsic Fabry-Perot Interferometric Sensors / 24.2
- 24.3 Intrinsic Fabry-Perot Interferometric Sensors / 24.4
- 24.4 Fiber Bragg Grating Sensors / 24.5
- 24.5 Long-Period Grating Sensors / 24.8
- 24.6 Comparison of Sensing Schemes / 24.13
- 24.7 Conclusion / 24.13
- 24.8 References / 24.13
- 24.9 Further Reading / 24.14

Chapter 25. High-Power Fiber Lasers and Amplifiers **25.1**
Timothy S. McComb, Martin C. Richardson, and Michael Bass

- 25.1 Glossary / 25.1
- 25.2 Introduction / 25.3
- 25.3 Fiber Laser Limitations / 25.6
- 25.4 Fiber Laser Fundamentals / 25.7
- 25.5 Fiber Laser Architectures / 25.9
- 25.6 LMA Fiber Designs / 25.18
- 25.7 Active Fiber Dopants / 25.22
- 25.8 Fiber Fabrication and Materials / 25.26
- 25.9 Spectral and Temporal Modalities / 25.29
- 25.10 Conclusions / 25.33
- 25.11 References / 25.33

PART 5. X-Ray and Neutron Optics

SUBPART 5.1. INTRODUCTION AND APPLICATIONS

Chapter 26. An Introduction to X-Ray and Neutron Optics **26.5**
Carolyn MacDonald

- 26.1 History / 26.5
- 26.2 X-Ray Interaction with Matter / 26.6
- 26.3 Optics Choices / 26.7
- 26.4 Focusing and Collimation / 26.9
- 26.5 References / 26.11

Chapter 27. Coherent X-Ray Optics and Microscopy *Qun Shen* **27.1**

- 27.1 Glossary / 27.1
- 27.2 Introduction / 27.2
- 27.3 Fresnel Wave Propagation / 27.2
- 27.4 Unified Approach for Near- and Far-Field Diffraction / 27.2

- 27.5 Coherent Diffraction Microscopy / 27.4
 27.6 Coherence Preservation in X-Ray Optics / 27.5
 27.7 References / 27.5

Chapter 28. Requirements for X-Ray Diffraction *Scott T. Misture* 28.1

- 28.1 Introduction / 28.1
 28.2 Slits / 28.1
 28.3 Crystal Optics / 28.3
 28.4 Multilayer Optics / 28.5
 28.5 Capillary and Polycapillary Optics / 28.5
 28.6 Diffraction and Fluorescence Systems / 28.5
 28.7 X-Ray Sources and Microsources / 28.7
 28.8 References / 28.7

Chapter 29. Requirements for X-Ray Fluorescence *Walter Gibson and George Havrilla* 29.1

- 29.1 Introduction / 29.1
 29.2 Wavelength-Dispersive X-Ray Fluorescence (WDXRF) / 29.2
 29.3 Energy-Dispersive X-Ray Fluorescence (EDXRF) / 29.3
 29.4 References / 29.12

Chapter 30. Requirements for X-Ray Spectroscopy *Dirk Lützenkirchen-Hecht and Ronald Frahm* 30.1

- 30.1 References / 30.5

Chapter 31. Requirements for Medical Imaging and X-Ray Inspection *Douglas Pfeiffer* 31.1

- 31.1 Introduction to Radiography and Tomography / 31.1
 31.2 X-Ray Attenuation and Image Formation / 31.1
 31.3 X-Ray Detectors and Image Receptors / 31.4
 31.4 Tomography / 31.5
 31.5 Computed Tomography / 31.5
 31.6 Digital Tomosynthesis / 31.7
 31.7 Digital Displays / 31.8
 31.8 Conclusion / 31.9
 31.9 References / 31.10

Chapter 32. Requirements for Nuclear Medicine *Lars R. Furenlid* 32.1

- 32.1 Introduction / 32.1
 32.2 Projection Image Acquisition / 32.2
 32.3 Information Content in SPECT / 32.3
 32.4 Requirements for Optics For SPECT / 32.4
 32.5 References / 32.4

Chapter 33. Requirements for X-Ray Astronomy *Scott O. Rohrbach* 33.1

- 33.1 Introduction / 33.1
 33.2 Trade-Offs / 33.2
 33.3 Summary / 33.4

Chapter 34. Extreme Ultraviolet Lithography *Franco Cerrina and Fan Jiang* 34.1

- 34.1 Introduction / 34.1
 34.2 Technology / 34.2

- 34.3 Outlook / 34.5
- 34.4 Acknowledgments / 34.6
- 34.5 References / 34.7

Chapter 35. Ray Tracing of X-Ray Optical Systems *Franco Cerrina and Manuel Sanchez del Rio* **35.1**

- 35.1 Introduction / 35.1
- 35.2 The Conceptual Basis of SHADOW / 35.2
- 35.3 Interfaces and Extensions of SHADOW / 35.3
- 35.4 Examples / 35.4
- 35.5 Conclusions and Future / 35.5
- 35.6 References / 35.6

Chapter 36. X-Ray Properties of Materials *Eric M. Gullikson* **36.1**

- 36.1 X-Ray and Neutron Optics / 36.2
- 36.2 Electron Binding Energies, Principal K- and L-Shell Emission Lines,
and Auger Electron Energies / 36.3
- 36.3 References / 36.10

SUBPART 5.2. REFRACTIVE AND INTERFERENCE OPTICS

Chapter 37. Refractive X-Ray Lenses *Bruno Lengeler and Christian G. Schroer* **37.3**

- 37.1 Introduction / 37.3
- 37.2 Refractive X-Ray Lenses with Rotationally Parabolic Profile / 37.4
- 37.3 Imaging with Parabolic Refractive X-Ray Lenses / 37.6
- 37.4 Microfocusing with Parabolic Refractive X-Ray Lenses / 37.7
- 37.5 Prefocusing and Collimation with Parabolic Refractive X-Ray Lenses / 37.8
- 37.6 Nanofocusing Refractive X-Ray Lenses / 37.8
- 37.7 Conclusion / 37.11
- 37.8 References / 37.11

Chapter 38. Gratings and Monochromators in the VUV and Soft X-Ray Spectral Region *Malcolm R. Howells* **38.1**

- 38.1 Introduction / 38.1
- 38.2 Diffraction Properties / 38.1
- 38.3 Focusing Properties / 38.3
- 38.4 Dispersion Properties / 38.6
- 38.5 Resolution Properties / 38.7
- 38.6 Efficiency / 38.8
- 38.7 References / 38.8

Chapter 39. Crystal Monochromators and Bent Crystals *Peter Siddons* **39.1**

- 39.1 Crystal Monochromators / 39.1
- 39.2 Bent Crystals / 39.5
- 39.3 References / 39.6

Chapter 40. Zone Plates *Alan Michette* **40.1**

- 40.1 Introduction / 40.1
- 40.2 Geometry of a Zone Plate / 40.1
- 40.3 Zone Plates as Thin Lenses / 40.3
- 40.4 Diffraction Efficiencies of Zone Plates / 40.4
- 40.5 Manufacture of Zone Plates / 40.8

- 40.6 Bragg-Fresnel Lenses / 40.9
 40.7 References / 40.10

Chapter 41. Multilayers *Eberhard Spiller* 41.1

- 41.1 Glossary / 41.1
 41.2 Introduction / 41.1
 41.3 Calculation of Multilayer Properties / 41.3
 41.4 Fabrication Methods and Performance / 41.4
 41.5 Multilayers for Diffractive Imaging / 41.9
 41.6 References / 41.10

Chapter 42. Nanofocusing of Hard X-Rays with Multilayer Laue Lenses *Albert T. Macrander, Hanfei Yan, Hyon Chol Kang, Jörg Maser, Chian Liu, Ray Conley, and G. Brian Stephenson* 42.1

- Abstract / 42.1
 42.1 Introduction / 42.2
 42.2 MLL Concept and Volume Diffraction Calculations / 42.4
 42.3 Magnetron-Sputtered MLLs / 42.5
 42.4 Instrumental Beamline Arrangement and Measurements / 42.9
 42.5 Takagi-Taupin Calculations / 42.12
 42.6 Wedged MLLs / 42.12
 42.7 MMLs with Curved Interfaces / 42.14
 42.8 MLL Prospects / 42.15
 42.9 Summary / 42.17
 42.10 Acknowledgments / 42.17
 42.11 References / 42.18

Chapter 43. Polarizing Crystal Optics *Qun Shen* 43.1

- 43.1 Introduction / 43.1
 43.2 Linear Polarizers / 43.2
 43.3 Linear Polarization Analyzers / 43.4
 43.4 Phase Plates for Circular Polarization / 43.5
 43.5 Circular Polarization Analyzers / 43.6
 43.6 Acknowledgments / 43.8
 43.7 References / 43.8

SUBPART 5.3. REFLECTIVE OPTICS

Chapter 44. Image Formation with Grazing Incidence Optics *James E. Harvey* 44.3

- 44.1 Glossary / 44.3
 44.2 Introduction to X-Ray Mirrors / 44.3
 44.3 Optical Design and Residual Aberrations of Grazing Incidence Telescopes / 44.6
 44.4 Image Analysis for Grazing Incidence X-Ray Optics / 44.12
 44.5 Validation of Image Analysis for Grazing Incidence X-Ray Optics / 44.16
 44.6 References / 44.18

Chapter 45. Aberrations for Grazing Incidence Optics *Timo T. Saha* 45.1

- 45.1 Grazing Incidence Telescopes / 45.1
 45.2 Surface Equations / 45.1
 45.3 Transverse Ray Aberration Expansions / 45.3
 45.4 Curvature of the Best Focal Surface / 45.5
 45.5 Aberration Balancing / 45.5
 45.6 On-Axis Aberrations / 45.6
 45.7 References / 45.8

Chapter 46. X-Ray Mirror Metrology *Peter Z. Takacs* 46.1

- 46.1 Glossary / 46.1
- 46.2 Introduction / 46.1
- 46.3 Surface Finish Metrology / 46.2
- 46.4 Surface Figure Metrology / 46.3
- 46.5 Practical Profile Analysis Considerations / 46.6
- 46.6 References / 46.12

Chapter 47. Astronomical X-Ray Optics *Marshall K. Joy and Brian D. Ramsey* 47.1

- 47.1 Introduction / 47.1
- 47.2 Wolter X-Ray Optics / 47.2
- 47.3 Kirkpatrick-Baez Optics / 47.7
- 47.4 Hard X-Ray Optics / 47.9
- 47.5 Toward Higher Angular Resolution / 47.10
- 47.6 References / 47.11

Chapter 48. Multifoil X-Ray Optics *Ladislav Pina* 48.1

- 48.1 Introduction / 48.1
- 48.2 Grazing Incidence Optics / 48.1
- 48.3 Multifoil Lobster-Eye Optics / 48.2
- 48.4 Multifoil Kirkpatrick-Baez Optics / 48.3
- 48.5 Summary / 48.4
- 48.6 References / 48.4

Chapter 49. Pore Optics *Marco W. Beijersbergen* 49.1

- 49.1 Introduction / 49.1
- 49.2 Glass Micropore Optics / 49.1
- 49.3 Silicon Pore Optics / 49.6
- 49.4 Micromachined Silicon / 49.7
- 49.5 References / 49.7

Chapter 50. Adaptive X-Ray Optics *Ali Khounsary* 50.1

- 50.1 Introduction / 50.1
- 50.2 Adaptive Optics in X-Ray Astronomy / 50.2
- 50.3 Active and Adaptive Optics for Synchrotron- and Lab-Based X-Ray Sources / 50.2
- 50.4 Conclusions / 50.8
- 50.5 References / 50.8

Chapter 51. The Schwarzschild Objective *Franco Cerrina* 51.1

- 51.1 Introduction / 51.1
- 51.2 Applications to X-Ray Domain / 51.3
- 51.3 References / 51.5

Chapter 52. Single Capillaries *Donald H. Bilderback and Sterling W. Cornaby* 52.1

- 52.1 Background / 52.1
- 52.2 Design Parameters / 52.1
- 52.3 Fabrication / 52.4
- 52.4 Applications of Single-Bounce Capillary Optics / 52.5
- 52.5 Applications of Condensing Capillary Optics / 52.6
- 52.6 Conclusions / 52.6

52.7	Acknowledgments /	52.6
52.8	References /	52.6

Chapter 53. Polycapillary X-Ray Optics	<i>Carolyn MacDonald and Walter Gibson</i>	53.1

53.1	Introduction /	53.1
53.2	Simulations and Defect Analysis /	53.3
53.3	Radiation Resistance /	53.5
53.4	Alignment and Measurement /	53.5
53.5	Collimation /	53.8
53.6	Focusing /	53.9
53.7	Applications /	53.10
53.8	Summary /	53.19
53.9	Acknowledgments /	53.19
53.10	References /	53.19

SUBPART 5.4. X-RAY SOURCES

Chapter 54. X-Ray Tube Sources	<i>Susanne M. Lee and Carolyn MacDonald</i>	54.3

54.1	Introduction /	54.3
54.2	Spectra /	54.4
54.3	Cathode Design and Geometry /	54.10
54.4	Effect of Anode Material, Geometry, and Source Size on Intensity and Brightness /	54.11
54.5	General Optimization /	54.15
54.6	References /	54.17

Chapter 55. Synchrotron Sources	<i>Steven L. Hulbert and Gwyn P. Williams</i>	55.1

55.1	Introduction /	55.1
55.2	Theory of Synchrotron Radiation Emission /	55.2
55.3	Insertion Devices (Undulators and Wigglers) /	55.9
55.4	Coherence of Synchrotron Radiation Emission in the Long Wavelength Limit /	55.17
55.5	Conclusion /	55.20
55.6	References /	55.20

Chapter 56. Laser-Generated Plasmas	<i>Alan Michette</i>	56.1

56.1	Introduction /	56.1
56.2	Characteristic Radiation /	56.2
56.3	Bremsstrahlung /	56.8
56.4	Recombination Radiation /	56.10
56.5	References /	56.10

Chapter 57. Pinch Plasma Sources	<i>Victor Kantsyrev</i>	57.1

57.1	Introduction /	57.1
57.2	Types of Z-Pinch Radiation Sources /	57.2
57.3	Choice of Optics for Z-Pinch Sources /	57.4
57.4	References /	57.5

Chapter 58. X-Ray Lasers	<i>Greg Tallents</i>	58.1

58.1	Free-Electron Lasers /	58.1
58.2	High Harmonic Production /	58.2
58.3	Plasma-Based EUV Lasers /	58.2
58.4	References /	58.4

Chapter 59. Inverse Compton X-Ray Sources *Frank Carroll* 59.1

- 59.1 Introduction / 59.1
- 59.2 Inverse Compton Calculations / 59.2
- 59.3 Practical Devices / 59.2
- 59.4 Applications / 59.3
- 59.5 Industrial/Military/Crystallographic Uses / 59.4
- 59.6 References / 59.4

SUBPART 5.5. X-RAY DETECTORS

Chapter 60. Introduction to X-Ray Detectors *Walter Gibson and Peter Siddons* 60.3

- 60.1 Introduction / 60.3
- 60.2 Detector Type / 60.3
- 60.3 Summary / 60.9
- 60.4 References / 60.10

Chapter 61. Advances in Imaging Detectors *Aaron Couture* 61.1

- 61.1 Introduction / 61.1
- 61.2 Flat-Panel Detectors / 61.3
- 61.3 CCD Detectors / 61.7
- 61.4 Conclusion / 61.8
- 61.5 References / 61.8

Chapter 62. X-Ray Spectral Detection and Imaging *Eric Lifshin* 62.1

- 62.1 References / 62.6

SUBPART 5.6. NEUTRON OPTICS AND APPLICATIONS

Chapter 63. Neutron Optics *David Mildner* 63.3

- 63.1 Neutron Physics / 63.3
- 63.2 Scattering Lengths and Cross Sections / 63.5
- 63.3 Neutron Sources / 63.12
- 63.4 Neutron Optical Devices / 63.15
- 63.5 Refraction and Reflection / 63.19
- 63.6 Diffraction and Interference / 63.23
- 63.7 Polarization Techniques / 63.27
- 63.8 Neutron Detection / 63.31
- 63.9 References / 63.35

Chapter 64. Grazing-Incidence Neutron Optics *Mikhail Gubarev and Brian Ramsey* 64.1

- 64.1 Introduction / 64.1
- 64.2 Total External Reflection / 64.1
- 64.3 Diffractive Scattering and Mirror Surface Roughness Requirements / 64.2
- 64.4 Imaging Focusing Optics / 64.3
- 64.5 References / 64.7

CONTRIBUTORS

- Francisco Arregui** *Public University Navarra, Pamplona, Spain* (CHAP. 24)
- Michael Bass** *CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida* (CHAP. 25)
- Marco W. Beijersbergen** *Cosine Research B.V./Cosine Science & Computing B.V., Leiden University, Leiden, Netherlands* (CHAP. 49)
- Donald H. Bilderback** *Cornell High Energy Synchrotron Source, School of Applied and Engineering Physics, Cornell University, Ithaca, New York* (CHAP. 52)
- John A. Buck** *Georgia Institute of Technology, School of Electrical and Computer Engineering, Atlanta, Georgia* (CHAPS. 10, 14)
- Frank Carroll** *MXISystems, Nashville, Tennessee* (CHAP. 59)
- Franco Cerrina** *Department of Electrical and Computer Engineering, University of Wisconsin, Madison, Wisconsin* (CHAPS. 34, 35, 51)
- I-Cheng Chang** *Accord Optics, Sunnyvale, California* (CHAP. 6)
- James H. Churnside** *National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Boulder, Colorado* (CHAP. 3)
- Richard O. Claus** *Virginia Tech, Blacksburg, Virginia* (CHAP. 24)
- Ray Conley** *X-Ray Science Division, Argonne National Laboratory, Argonne, Illinois, and National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, New York* (CHAP. 42)
- Sterling W. Cornaby** *Cornell High Energy Synchrotron Source, School of Applied and Engineering Physics, Cornell University Ithaca, New York* (CHAP. 52)
- Aaron Couture** *GE Global Research Center, Niskayuna, New York* (CHAP. 61)
- Guang-ming Dai** *Laser Vision Correction Group, Advanced Medical Optics, Milpitas, California* (CHAP. 4)
- Casimer DeCusatis** *IBM Corporation, Poughkeepsie, New York* (CHAPS. 15, 23)
- Peter J. Delfyett** *CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida* (CHAP. 20)
- Ronald Frahm** *Bergische Universität Wuppertal, Wuppertal, Germany* (CHAP. 30)
- Robert Q. Fugate** *Starfire Optical Range, Directed Energy Directorate, Air Force Research Laboratory, Kirtland Air Force Base, New Mexico* (CHAP. 5)
- Lars R. Furenlid** *University of Arizona, Tucson, Arizona* (CHAP. 32)
- Elsa Garmire** *Dartmouth College, Hanover, New Hampshire* (CHAP. 13)
- Sebastian Gauza** *CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida* (CHAP. 8)
- Walter Gibson** *X-Ray Optical Systems, Inc., East Greenbush, New York* (CHAPS. 29, 53, 60)
- Mikhail Gubarev** *NASA/Marshall Space Flight Center, Huntsville, Alabama* (CHAP. 64)
- Eric M. Gullikson** *Center for X-Ray Optics, Lawrence Berkeley National Laboratory, Berkeley, California* (CHAP. 36)
- James A. Harrington** *Rutgers University, Piscataway, New Jersey* (CHAP. 12)
- James E. Harvey** *CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida* (CHAP. 44)
- George Havrilla** *Los Alamos National Laboratory, Los Alamos, New Mexico* (CHAP. 29)

Brian Henderson *Department of Physics and Applied Physics, University of Strathclyde, Glasgow, United Kingdom (CHAP. 2)*

Kenneth O. Hill *Communications Research Centre, Ottawa, Ontario, Canada, and Nu-Wave Photonics, Ottawa, Ontario, Canada (CHAP. 17)*

Malcolm R. Howells *Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California (CHAP. 38)*

Steven L. Hulbert *National Synchrotron Light Source, Brookhaven National Laboratory, Upton, New York (CHAP. 55)*

Ira Jacobs *The Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia (CHAP. 9)*

Fan Jiang *Electrical and Computer Engineering & Center for Nano Technology, University of Wisconsin, Madison (CHAP. 34)*

Marshall K. Joy *National Aeronautics and Space Administration, Marshall Space Flight Center, Huntsville, Alabama (CHAP. 47)*

Hyon Chol Kang *Materials Science Division, Argonne National Laboratory, Argonne, Illinois, and Advanced Materials Engineering Department, Chosun University, Gwangju, Republic of Korea (CHAP. 42)*

Victor Kantsyrev *Physics Department, University of Nevada, Reno, Nevada (CHAP. 57)*

Ali Khounsary *Argonne National Laboratory, Argonne, Illinois (CHAP. 50)*

Dennis K. Killinger *Center for Laser Atmospheric Sensing, Department of Physics, University of South Florida, Tampa, Florida (CHAP. 3)*

Susanne M. Lee *GE Global Research, Nikayuna, New York (CHAP. 54)*

Bruno Lengeler *Physikalisches Institut, RWTH Aachen University, Aachen, Germany (CHAP. 37)*

Guifang Li *CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida (CHAP. 15)*

Eric Lifshin *College of Nanoscale Science and Engineering, University at Albany, Albany, New York (CHAP. 62)*

Chian Liu *X-Ray Science Division, Argonne National Laboratory, Argonne, Illinois (CHAP. 42)*

Dirk Lützenkirchen-Hecht *Bergische Universität Wuppertal, Wuppertal, Germany (CHAP. 30)*

Carolyn MacDonald *University at Albany, Albany, New York (CHAPS. 26, 53, 54)*

Albert T. Macrander *X-Ray Science Division, Argonne National Laboratory, Argonne, Illinois (CHAP. 42)*

Virendra N. Mahajan *The Aerospace Corporation, El Segundo, California (CHAP. 4)*

Theresa A. Maldonado *Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas (CHAP. 7)*

Pavel V. Mamyshev *Bell Laboratories—Lucent Technologies, Holmdel, New Jersey (CHAP. 22)*

Jörg Maser *X-Ray Science Division, Argonne National Laboratory, Argonne, Illinois, and Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois (CHAP. 42)*

Ignacio Matias *Public University Navarra, Pamplona, Spain (CHAP. 24)*

Timothy S. McComb *CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida (CHAP. 25)*

Alan Michette *King's College, London, United Kingdom (CHAPS. 40, 56)*

David Mildner *NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland (CHAP. 63)*

Scott T. Misture *Kazuo Inamori School of Engineering, Alfred University, Alfred, New York (CHAP. 28)*

Daniel Nolan *Corning Inc., Corning, New York (CHAP. 16)*

Joseph C. Palais *Ira A. Fulton School of Engineering, Arizona State University, Tempe, Arizona (CHAP. 18)*

- Zhongqi Pan** *University of Louisiana at Lafayette, Lafayette, Louisiana (CHAP. 21)*
- Greg J. Pearce** *Max-Planck Institute for the Science of Light, Erlangen, Germany (CHAP. 11)*
- Douglas Pfeiffer** *Boulder Community Hospital, Boulder, Colorado (CHAP. 31)*
- Ladislav Pina** *Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague, Holesovickach (CHAP. 48)*
- Georgeanne M. Purvinis** *The Battelle Memorial Institute, Columbus, Ohio (CHAP. 7)*
- Brian D. Ramsey** *National Aeronautics and Space Administration, Marshall Space Flight Center, Huntsville, Alabama (CHAPS. 47, 64)*
- Martin C. Richardson** *CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida (CHAP. 25)*
- Scott O. Rohrbach** *Optics Branch, Goddard Space Flight Center, NASA, Greenbelt, Maryland (CHAP. 33)*
- Laurence S. Rothman** *Harvard-Smithsonian Center for Astrophysics, Atomic and Molecular Physics Division, Cambridge, Massachusetts (CHAP. 3)*
- Philip St. J. Russell** *Max-Planck Institute for the Science of Light, Erlangen, Germany (CHAP. 11)*
- Timo T. Saha** *NASA/Goddard Space Flight Center, Greenbelt, Maryland (CHAP. 45)*
- Manuel Sanchez del Rio** *European Synchrotron Radiation Facility, Grenoble, France (CHAP. 35)*
- Christian G. Schroer** *Institute of Structural Physics, TU Dresden, Dresden, Germany (CHAP. 37)*
- Qun Shen** *National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, New York (CHAPS. 27, 43)*
- Peter Siddons** *National Synchrotron Light Source, Brookhaven National Laboratory, Upton, New York (CHAPS. 39, 60)*
- Leo H. Spiekman** *Alphion Corp., Princeton Junction, New Jersey (CHAP. 19)*
- Eberhard Spiller** *Spiller X-Ray Optics, Livermore, California (CHAP. 41)*
- G. Brian Stephenson** *Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois, Materials Science Division, Argonne National Laboratory, Argonne, Illinois (CHAP. 42)*
- John C. Stover** *The Scatter Works, Inc., Tucson, Arizona (CHAP. 1)*
- Peter Z. Takacs** *Brookhaven National Laboratory, Upton, New York (CHAP. 46)*
- Greg Tallents** *University of York, York, United Kingdom (CHAP. 58)*
- Jay M. Wiesenfeld** *Bell Laboratories, Alcatel-Lucent, Murray Hill, New Jersey (CHAP. 19)*
- Gwyn P. Williams** *Free Electron Laser, Thomas Jefferson National Accelerator Facility, Newport News, Virginia (CHAP. 55)*
- Alan E. Willner** *University of Southern California, Los Angeles, California (CHAP. 21)*
- Shin-Tson Wu** *CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida (CHAP. 8)*
- Yong Xie** *Texas Instruments Inc., Dallas, Texas (CHAP. 21)*
- Hanfei Yan** *Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois, and National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, New York (CHAP. 42)*
- Changyuan Yu** *National University of Singapore, and A *STAR Institute for Infocomm Research, Singapore (CHAP. 21)*

This page intentionally left blank

BRIEF CONTENTS OF ALL VOLUMES

VOLUME I. GEOMETRICAL AND PHYSICAL OPTICS, POLARIZED LIGHT, COMPONENTS AND INSTRUMENTS

PART 1. GEOMETRICAL OPTICS

Chapter 1. General Principles of Geometrical Optics *Douglas S. Goodman*

PART 2. PHYSICAL OPTICS

Chapter 2. Interference *John E. Greivenkamp*

Chapter 3. Diffraction *Arvind S. Marathay and John F. McCalmont*

Chapter 4. Transfer Function Techniques *Glenn D. Boreman*

Chapter 5. Coherence Theory *William H. Carter*

Chapter 6. Coherence Theory: Tools and Applications *Gisele Bennett, William T. Rhodes, and J. Christopher James*

Chapter 7. Scattering by Particles *Craig F. Bohren*

Chapter 8. Surface Scattering *Eugene L. Church and Peter Z. Takacs*

Chapter 9. Volume Scattering in Random Media *Aristide Dogariu and Jeremy Ellis*

Chapter 10. Optical Spectroscopy and Spectroscopic Lineshapes *Brian Henderson*

Chapter 11. Analog Optical Signal and Image Processing *Joseph W. Goodman*

PART 3. POLARIZED LIGHT

Chapter 12. Polarization *Jean M. Bennett*

Chapter 13. Polarizers *Jean M. Bennett*

Chapter 14. Mueller Matrices *Russell A. Chipman*

Chapter 15. Polarimetry *Russell A. Chipman*

Chapter 16. Ellipsometry *Rasheed M. A. Azzam*

PART 4. COMPONENTS

Chapter 17. Lenses *R. Barry Johnson*

Chapter 18. Afocal Systems *William B. Wetherell*

Chapter 19. Nondispersive Prisms *William L. Wolfe*

Chapter 20. Dispersive Prisms and Gratings *George J. Zissis*

Chapter 21. Integrated Optics *Thomas L. Koch, Frederick J. Leonberger, and Paul G. Suchoski*

Chapter 22. Miniature and Micro-Optics *Tom D. Milster and Tomasz S. Tkaczyk*

Chapter 23. Binary Optics *Michael W. Farn and Wilfrid B. Veldkamp*

Chapter 24. Gradient Index Optics *Duncan T. Moore*

PART 5. INSTRUMENTS

Chapter 25. Cameras *Norman Goldberg*

Chapter 26. Solid-State Cameras *Gerald C. Holst*

Chapter 27. Camera Lenses *Ellis Betensky, Melvin H. Kreitzer, and Jacob Moskovich*

Chapter 28. Microscopes *Rudolf Oldenbourg and Michael Shribak*

- Chapter 29. Reflective and Catadioptric Objectives *Lloyd Jones*
Chapter 30. Scanners *Leo Beiser and R. Barry Johnson*
Chapter 31. Optical Spectrometers *Brian Henderson*
Chapter 32. Interferometers *Parameswaran Hariharan*
Chapter 33. Holography and Holographic Instruments *Lloyd Huff*
Chapter 34. Xerographic Systems *Howard Stark*
Chapter 35. Principles of Optical Disk Data Storage *Masud Mansuripur*

VOLUME II. DESIGN, FABRICATION, AND TESTING; SOURCES AND DETECTORS; RADIOMETRY AND PHOTOMETRY

PART 1. DESIGN

- Chapter 1. Techniques of First-Order Layout *Warren J. Smith*
Chapter 2. Aberration Curves in Lens Design *Donald C. O'Shea and Michael E. Harrigan*
Chapter 3. Optical Design Software *Douglas C. Sinclair*
Chapter 4. Optical Specifications *Robert R. Shannon*
Chapter 5. Tolerancing Techniques *Robert R. Shannon*
Chapter 6. Mounting Optical Components *Paul R. Yoder, Jr.*
Chapter 7. Control of Stray Light *Robert P. Breault*
Chapter 8. Thermal Compensation Techniques *Philip J. Rogers and Michael Roberts*

PART 2. FABRICATION

- Chapter 9. Optical Fabrication *Michael P. Mandina*
Chapter 10. Fabrication of Optics by Diamond Turning *Richard L. Rhorer and Chris J. Evans*

PART 3. TESTING

- Chapter 11. Orthonormal Polynomials in Wavefront Analysis *Virendra N. Mahajan*
Chapter 12. Optical Metrology *Zacarias Malacara and Daniel Malacara-Hernández*
Chapter 13. Optical Testing *Daniel Malacara-Hernández*
Chapter 14. Use of Computer-Generated Holograms in Optical Testing *Katherine Creath and James C. Wyant*

PART 4. SOURCES

- Chapter 15. Artificial Sources *Anthony LaRocca*
Chapter 16. Lasers *William T. Silfvast*
Chapter 17. Light-Emitting Diodes *Roland H. Haitz, M. George Craford, and Robert H. Weissman*
Chapter 18. High-Brightness Visible LEDs *Winston V. Schoenfeld*
Chapter 19. Semiconductor Lasers *Pamela L. Derry, Luis Figueroa, and Chi-shain Hong*
Chapter 20. Ultrashort Optical Sources and Applications *Jean-Claude Diels and Ladan Arissian*
Chapter 21. Attosecond Optics *Zenghu Chang*
Chapter 22. Laser Stabilization *John L. Hall, Matthew S. Taubman, and Jun Ye*
Chapter 23. Quantum Theory of the Laser *János A. Bergou, Berthold-Georg Englert, Melvin Lax, Marian O. Scully, Herbert Walther, and M. Suhail Zubairy*

PART 5. DETECTORS

- Chapter 24. Photodetectors *Paul R. Norton*
Chapter 25. Photodetection *Abhay M. Joshi and Gregory H. Olsen*
Chapter 26. High-Speed Photodetectors *John E. Bowers and Yih G. Wey*
Chapter 27. Signal Detection and Analysis *John R. Willison*
Chapter 28. Thermal Detectors *William L. Wolfe and Paul W. Kruse*

PART 6. IMAGING DETECTORS

- Chapter 29. Photographic Films *Joseph H. Altman*
Chapter 30. Photographic Materials *John D. Baloga*

- Chapter 31. Image Tube Intensified Electronic Imaging *C. Bruce Johnson and Larry D. Owen*
 Chapter 32. Visible Array Detectors *Timothy J. Tredwell*
 Chapter 33. Infrared Detector Arrays *Lester J. Kozlowski and Walter F. Kosonocky*

PART 7. RADIOMETRY AND PHOTOMETRY

- Chapter 34. Radiometry and Photometry *Edward F. Zalewski*
 Chapter 35. Measurement of Transmission, Absorption, Emission, and Reflection *James M. Palmer*
 Chapter 36. Radiometry and Photometry: Units and Conversions *James M. Palmer*
 Chapter 37. Radiometry and Photometry for Vision Optics *Yoshi Ohno*
 Chapter 38. Spectroradiometry *Carolyn J. Sher DeCusatis*
 Chapter 39. Nonimaging Optics: Concentration and Illumination *William Cassarly*
 Chapter 40. Lighting and Applications *Anurag Gupta and R. John Koschel*

VOLUME III. VISION AND VISION OPTICS

- Chapter 1. Optics of the Eye *Neil Charman*
 Chapter 2. Visual Performance *Wilson S. Geisler and Martin S. Banks*
 Chapter 3. Psychophysical Methods *Denis G. Pelli and Bart Farell*
 Chapter 4. Visual Acuity and Hyperacuity *Gerald Westheimer*
 Chapter 5. Optical Generation of the Visual Stimulus *Stephen A. Burns and Robert H. Webb*
 Chapter 6. The Maxwellian View with an Addendum on Apodization *Gerald Westheimer*
 Chapter 7. Ocular Radiation Hazards *David H. Sliney*
 Chapter 8. Biological Waveguides *Vasudevan Lakshminarayanan and Jay M. Enoch*
 Chapter 9. The Problem of Correction for the Stiles-Crawford Effect of the First Kind in Radiometry and Photometry, a Solution *Jay M. Enoch and Vasudevan Lakshminarayanan*
 Chapter 10. Colorimetry *David H. Brainard and Andrew Stockman*
 Chapter 11. Color Vision Mechanisms *Andrew Stockman and David H. Brainard*
 Chapter 12. Assessment of Refraction and Refractive Errors and Their Influence on Optical Design *B. Ralph Chou*
 Chapter 13. Binocular Vision Factors That Influence Optical Design *Clifton Schor*
 Chapter 14. Optics and Vision of the Aging Eye *John S. Werner, Brooke E. Scheffrin, and Arthur Bradley*
 Chapter 15. Adaptive Optics in Retinal Microscopy and Vision *Donald T. Miller and Austin Roorda*
 Chapter 16. Refractive Surgery, Correction of Vision, PRK and LASIK *L. Diaz-Santana and Harilaos Giniis*
 Chapter 17. Three-Dimensional Confocal Microscopy of the Living Human Cornea *Barry R. Masters*
 Chapter 18. Diagnostic Use of Optical Coherence Tomography in the Eye *Johannes F. de Boer*
 Chapter 19. Gradient Index Optics in the Eye *Barbara K. Pierscionek*
 Chapter 20. Optics of Contact Lenses *Edward S. Bennett*
 Chapter 21. Intraocular Lenses *Jim Schwiegerling*
 Chapter 22. Displays for Vision Research *William Cowan*
 Chapter 23. Vision Problems at Computers *Jeffrey Anshel and James E. Sheedy*
 Chapter 24. Human Vision and Electronic Imaging *Bernice E. Rogowitz, Thrasyvoulos N. Pappas, and Jan P. Allebach*
 Chapter 25. Visual Factors Associated with Head-Mounted Displays *Brian H. Tsou and Martin Shenker*

VOLUME IV. OPTICAL PROPERTIES OF MATERIALS, NONLINEAR OPTICS, QUANTUM OPTICS

PART 1. PROPERTIES

- Chapter 1. Optical Properties of Water *Curtis D. Mobley*
 Chapter 2. Properties of Crystals and Glasses *William J. Tropf, Michael E. Thomas, and Eric W. Rogala*
 Chapter 3. Polymeric Optics *John D. Lytle*
 Chapter 4. Properties of Metals *Roger A. Paquin*

- Chapter 5. Optical Properties of Semiconductors *David G. Seiler, Stefan Zollner, Alain C. Diebold, and Paul M. Amirtharaj*
- Chapter 6. Characterization and Use of Black Surfaces for Optical Systems *Stephen M. Pompea and Robert P. Breault*
- Chapter 7. Optical Properties of Films and Coatings *Jerzy A. Dobrowolski*
- Chapter 8. Fundamental Optical Properties of Solids *Alan Miller*
- Chapter 9. Photonic Bandgap Materials *Pierre R. Villeneuve*

PART 2. NONLINEAR OPTICS

- Chapter 10. Nonlinear Optics *Chung L. Tang*
- Chapter 11. Coherent Optical Transients *Paul R. Berman and Duncan G. Steel*
- Chapter 12. Photorefractive Materials and Devices *Mark Cronin-Golomb and Marvin Klein*
- Chapter 13. Optical Limiting *David J. Hagan*
- Chapter 14. Electromagnetically Induced Transparency *Jonathan P. Marangos and Thomas Halfmann*
- Chapter 15. Stimulated Raman and Brillouin Scattering *John Reintjes and Mark Bashkansky*
- Chapter 16. Third-Order Optical Nonlinearities *Mansoor Sheik-Bahae and Michael P. Hasselbeck*
- Chapter 17. Continuous-Wave Optical Parametric Oscillators *Majid Ebrahim-Zadeh*
- Chapter 18. Nonlinear Optical Processes for Ultrashort Pulse Generation *Uwe Siegner and Ursula Keller*
- Chapter 19. Laser-Induced Damage to Optical Materials *Marion J. Soileau*

PART 3. QUANTUM AND MOLECULAR OPTICS

- Chapter 20. Laser Cooling and Trapping of Atoms *Harold J. Metcalf and Peter van der Straten*
- Chapter 21. Strong Field Physics *Todd Ditmire*
- Chapter 22. Slow Light Propagation in Atomic and Photonic Media *Jacob B. Khurgin*
- Chapter 23. Quantum Entanglement in Optical Interferometry *Hwang Lee, Christoph F. Wildfeuer, Sean D. Huver, and Jonathan P. Dowling*

VOLUME V. ATMOSPHERIC OPTICS, MODULATORS, FIBER OPTICS, X-RAY AND NEUTRON OPTICS

PART 1. MEASUREMENTS

- Chapter 1. Scatterometers *John C. Stover*
- Chapter 2. Spectroscopic Measurements *Brian Henderson*

PART 2. ATMOSPHERIC OPTICS

- Chapter 3. Atmospheric Optics *Dennis K. Killinger, James H. Churnside, and Laurence S. Rothman*
- Chapter 4. Imaging through Atmospheric Turbulence *Virendra N. Mahajan and Guang-ming Dai*
- Chapter 5. Adaptive Optics *Robert Q. Fugate*

PART 3. MODULATORS

- Chapter 6. Acousto-Optic Devices *I-Cheng Chang*
- Chapter 7. Electro-Optic Modulators *Georgianne M. Purvinis and Theresa A. Maldonado*
- Chapter 8. Liquid Crystals *Sebastian Gauza and Shin-Tson Wu*

PART 4. FIBER OPTICS

- Chapter 9. Optical Fiber Communication Technology and System Overview *Ira Jacobs*
- Chapter 10. Nonlinear Effects in Optical Fibers *John A. Buck*
- Chapter 11. Photonic Crystal Fibers *Philip St. J. Russell and Greg J. Pearce*
- Chapter 12. Infrared Fibers *James A. Harrington*
- Chapter 13. Sources, Modulators, and Detectors for Fiber Optic Communication Systems *Elsa Garmire*
- Chapter 14. Optical Fiber Amplifiers *John A. Buck*

- Chapter 15. Fiber Optic Communication Links (Telecom, Datacom, and Analog) *Casimer DeCusatis and Guifang Li*
- Chapter 16. Fiber-Based Couplers *Daniel Nolan*
- Chapter 17. Fiber Bragg Gratings *Kenneth O. Hill*
- Chapter 18. Micro-Optics-Based Components for Networking *Joseph C. Palais*
- Chapter 19. Semiconductor Optical Amplifiers *Jay M. Wiesenfeld and Leo H. Spiekman*
- Chapter 20. Optical Time-Division Multiplexed Communication Networks *Peter J. Delfyett*
- Chapter 21. WDM Fiber-Optic Communication Networks *Alan E. Willner, Changyuan Yu, Zhongqi Pan, and Yong Xie*
- Chapter 22. Solitons in Optical Fiber Communication Systems *Pavel V. Mamyshev*
- Chapter 23. Fiber-Optic Communication Standards *Casimer DeCusatis*
- Chapter 24. Optical Fiber Sensors *Richard O. Claus, Ignacio Matias, and Francisco Arregui*
- Chapter 25. High-Power Fiber Lasers and Amplifiers *Timothy S. McComb, Martin C. Richardson, and Michael Bass*

PART 5. X-RAY AND NEUTRON OPTICS

Subpart 5.1. Introduction and Applications

- Chapter 26. An Introduction to X-Ray and Neutron Optics *Carolyn MacDonald*
- Chapter 27. Coherent X-Ray Optics and Microscopy *Qun Shen*
- Chapter 28. Requirements for X-Ray Diffraction *Scott T. Misture*
- Chapter 29. Requirements for X-Ray Fluorescence *George J. Havrilla*
- Chapter 30. Requirements for X-Ray Spectroscopy *Dirk Lützenkirchen-Hecht and Ronald Frahm*
- Chapter 31. Requirements for Medical Imaging and X-Ray Inspection *Douglas Pfeiffer*
- Chapter 32. Requirements for Nuclear Medicine *Lars R. Furenlid*
- Chapter 33. Requirements for X-Ray Astronomy *Scott O. Rohrbach*
- Chapter 34. Extreme Ultraviolet Lithography *Franco Cerrina and Fan Jiang*
- Chapter 35. Ray Tracing of X-Ray Optical Systems *Franco Cerrina and M. Sanchez del Rio*
- Chapter 36. X-Ray Properties of Materials *Eric M. Gullikson*

Subpart 5.2. Refractive and Interference Optics

- Chapter 37. Refractive X-Ray Lenses *Bruno Lengeler and Christian G. Schroer*
- Chapter 38. Gratings and Monochromators in the VUV and Soft X-Ray Spectral Region *Malcolm R. Howells*
- Chapter 39. Crystal Monochromators and Bent Crystals *Peter Siddons*
- Chapter 40. Zone Plates *Alan Michette*
- Chapter 41. Multilayers *Eberhard Spiller*
- Chapter 42. Nanofocusing of Hard X-Rays with Multilayer Laue Lenses *Albert T. Macrander, Hanfei Yan, Hyon Chol Kang, Jörg Maser, Chian Liu, Ray Conley, and G. Brian Stephenson*
- Chapter 43. Polarizing Crystal Optics *Qun Shen*

Subpart 5.3. Reflective Optics

- Chapter 44. Image Formation with Grazing Incidence Optics *James Harvey*
- Chapter 45. Aberrations for Grazing Incidence Optics *Timo T. Saha*
- Chapter 46. X-Ray Mirror Metrology *Peter Z. Takacs*
- Chapter 47. Astronomical X-Ray Optics *Marshall K. Joy and Brian D. Ramsey*
- Chapter 48. Multifoil X-Ray Optics *Ladislav Pina*
- Chapter 49. Pore Optics *Marco Beijersbergen*
- Chapter 50. Adaptive X-Ray Optics *Ali Khounsary*
- Chapter 51. The Schwarzschild Objective *Franco Cerrina*
- Chapter 52. Single Capillaries *Donald H. Bilderback and Sterling W. Cornaby*
- Chapter 53. Polycapillary X-Ray Optics *Carolyn MacDonald and Walter Gibson*

Subpart 5.4. X-Ray Sources

- Chapter 54. X-Ray Tube Sources *Susanne M. Lee and Carolyn MacDonald*
Chapter 55. Synchrotron Sources *Steven L. Hulbert and Gwyn P. Williams*
Chapter 56. Laser-Generated Plasmas *Alan Michette*
Chapter 57. Pinch Plasma Sources *Victor Kantsyrev*
Chapter 58. X-Ray Lasers *Greg Tallents*
Chapter 59. Inverse Compton X-Ray Sources *Frank Carroll*

Subpart 5.5. X-Ray Detectors

- Chapter 60. Introduction to X-Ray Detectors *Walter M. Gibson and Peter Siddons*
Chapter 61. Advances in Imaging Detectors *Aaron Couture*
Chapter 62. X-Ray Spectral Detection and Imaging *Eric Lifshin*

Subpart 5.6. Neutron Optics and Applications

- Chapter 63. Neutron Optics *David Mildner*
Chapter 64. Grazing-Incidence Neutron Optics *Mikhail Gubarev and Brian Ramsey*

EDITORS' PREFACE

The third edition of the *Handbook of Optics* is designed to pull together the dramatic developments in both the basic and applied aspects of the field while retaining the archival, reference book value of a handbook. This means that it is much more extensive than either the first edition, published in 1978, or the second edition, with Volumes I and II appearing in 1995 and Volumes III and IV in 2001. To cover the greatly expanded field of optics, the *Handbook* now appears in five volumes. Over 100 authors or author teams have contributed to this work.

Volume I is devoted to the fundamentals, components, and instruments that make optics possible. Volume II contains chapters on design, fabrication, testing, sources of light, detection, and a new section devoted to radiometry and photometry. Volume III concerns vision optics only and is printed entirely in color. In Volume IV there are chapters on the optical properties of materials, nonlinear, quantum and molecular optics. Volume V has extensive sections on fiber optics and x ray and neutron optics, along with shorter sections on measurements, modulators, and atmospheric optical properties and turbulence. Several pages of color inserts are provided where appropriate to aid the reader. A purchaser of the print version of any volume of the *Handbook* will be able to download a digital version containing all of the material in that volume in PDF format to one computer (see download instructions on bound-in card). The combined index for all five volumes can be downloaded from www.HandbookofOpticsOnline.com.

It is possible by careful selection of what and how to present that the third edition of the *Handbook* could serve as a text for a comprehensive course in optics. In addition, students who take such a course would have the *Handbook* as a career-long reference.

Topics were selected by the editors so that the *Handbook* could be a desktop (bookshelf) general reference for the parts of optics that had matured enough to warrant archival presentation. New chapters were included on topics that had reached this stage since the second edition, and existing chapters from the second edition were updated where necessary to provide this compendium. In selecting subjects to include, we also had to select which subjects to leave out. The criteria we applied were: (1) was it a specific application of optics rather than a core science or technology and (2) was it a subject in which the role of optics was peripheral to the central issue addressed. Thus, such topics as medical optics, laser surgery, and laser materials processing were not included. While applications of optics are mentioned in the chapters there is no space in the *Handbook* to include separate chapters devoted to all of the myriad uses of optics in today's world. If we had, the third edition would be much longer than it is and much of it would soon be outdated. We designed the third edition of the *Handbook of Optics* so that it concentrates on the principles of optics that make applications possible.

Authors were asked to try to achieve the dual purpose of preparing a chapter that was a worthwhile reference for someone working in the field and that could be used as a starting point to become acquainted with that aspect of optics. They did that and we thank them for the outstanding results seen throughout the *Handbook*. We also thank Mr. Taisuke Soda of McGraw-Hill for his help in putting this complex project together and Mr. Alan Tourtlotte and Ms. Susannah Lehman of the Optical Society of America for logistical help that made this effort possible.

We dedicate the third edition of the *Handbook of Optics* to all of the OSA volunteers who, since OSA's founding in 1916, give their time and energy to promoting the generation, application, archiving, and worldwide dissemination of knowledge in optics and photonics.

Michael Bass, Editor-in-Chief

Associate Editors:

Casimer M. DeCusatis

Jay M. Enoch

Vasudevan Lakshminarayanan

Guifang Li

Carolyn MacDonald

Virendra N. Mahajan

Eric Van Stryland

This page intentionally left blank

PREFACE TO VOLUME V

Volume V begins with Measurements, Atmospheric Optics, and Optical Modulators. There are chapters on scatterometers, spectroscopic measurements, transmission through the atmosphere, imaging through turbulence, and adaptive optics to overcome distortions as well as chapters on electro- and acousto-optic modulators and liquid crystal spatial light modulators. These are followed by the two main parts of this volume—Fiber Optics and X-Ray and Neutron Optics.

Optical fiber technology is truly an interdisciplinary field, incorporating aspects of solid-state physics, material science, and electrical engineering, among others. In the section on fiber optics, we introduce the fundamentals of optical fibers and cable assemblies, optical connectors, light sources, detectors, and related components. Assembly of the building blocks into optical networks required discussion of the unique requirements of digital versus analog links and telecommunication versus data communication networks. Issues such as optical link budget calculations, dispersion- or attenuation-limited links, and compliance with relevant industry standards are all addressed. Since one of the principle advantages of fiber optics is the ability to create high-bandwidth, long-distance interconnections, we also discuss the design and use of optical fiber amplifiers for different wavelength transmission windows. This leads to an understanding of the different network components which can be fabricated from optical fiber itself, such as splitters, combiners, fiber Bragg gratings, and other passive optical networking elements. We then provide a treatment of other important devices, including fiber sensors, fibers optimized for use in the infrared, micro-optic components for fiber networks and fiber lasers. Note that micro-optics for other applications are covered in Volume I of this *Handbook*. The physics of semiconductor lasers and photodetectors are presented in Volume II. Applications such as time or wavelength-division multiplexing networks provide their own challenges and are discussed in detail. High optical power applications lead us to a consideration of nonlinear optical fiber properties. Advanced topics for high speed, future networks are described in this section, including polarization mode dispersion; readers interested in the physical optics underlying dispersion should consult Volume I of this *Handbook*. This section includes chapters on photonic crystal fibers (for a broader treatment of photonic bandgap materials, see Volume IV) and on the growing applications of optical fiber networks.

Part 5 of this volume discusses a variety of X-Ray and Neutron Optics and their use in a wide range of applications. Part 5.1 is an introduction to the use and properties of x rays. It begins with a short chapter summarizing x-ray interactions and optics, followed by a discussion of coherence effects, and then illustrations of application constraints to the use of optics in seven applications, ranging from materials analysis to medicine, astronomy, and chip manufacturing. Because modeling is an important tool for both optics development and system design, Part 5.1 continues with a discussion of optics simulations, followed by tables of materials properties in the x-ray regime. Parts 5.2 and 5.3 are devoted to the discussion of the three classes of x-ray optics. Part 5.2 covers refractive, interference, and diffractive optics, including gratings, crystals (flat, bent, and polarizing), zone plates, and Laue lenses. It also includes a discussion of multilayer coatings, which are based on interference, but often added to reflective x-ray optics. Reflective optics is the topic of Part 5.3. Since reflective optics in the x-ray regime are used primarily in grazing incidence, the first three chapters of Part 5.3 cover the theory of image formation, aberrations, and metrology of grazing incidence mirrors. This is followed with descriptions of mirrors for astronomy and microscopy, adaptive optics for high heat load synchrotron beam lines, glass capillary reflective optics, also generally used for beam lines, and array optics such as multifoils, pore optics, and polycapillaries. The best choice of optic for a particular function depends on the application requirements, but is also influenced by the properties of the available sources and detectors. Part 5.4 describes six different types of x-ray sources. This is followed by Part 5.5, which includes an introduction to detectors and in-depth

discussions of imaging and spectral detectors. Finally, Part 5.6 describes the similarities and differences in the use of comparable optics technologies with neutrons.

In 1998, Walter Gibson designed the expansion of the x-ray and neutron section of the second edition of the *Handbook* from its original single chapter form. The third edition of this section is dedicated to his memory.

Guifang Li, Casimer M. DeCusatis, Virendra N. Mahajan, and Carolyn MacDonald
Associate Editors

In Memoriam Walter Maxwell Gibson (November 11, 1930–May 15, 2009)

After a childhood in Southern Utah working as a shepherd and stunt rider, Walt received his Ph.D. in nuclear chemistry under Nobel Laureate Glen Seaborg in 1956. He then spent 20 years at Bell Labs, where he did groundbreaking research in semiconductor detectors, particle-solid interactions, and in the development of ion beam techniques for material analysis. His interest in materials analysis and radiation detection naturally led him to an early and ongoing interest in developing x-ray analysis techniques, including early synchrotron beam line development. In 1970, he was named a fellow of the American Physical Society. In 1976, Walter was invited to chair the physics department of the University at Albany, SUNY (where he was fond of noting that they must have been confused as he had been neither an academic nor a physicist). He remained with the department for more than 25 years and was honored with the university's first named professorship, the James W. Corbett Distinguished Service Professor of Physics, in 1998. He later retired from the university to become the full-time chief technical officer of X-Ray Optical Systems, Inc., which he had cofounded coincident with UAlbany's Center for X-Ray Optics in 1991. He was the author of more than 300 technical articles and mentor to more than 48 doctoral graduates.

Walter Gibson's boundless energy, enthusiasm, wisdom, caring, courage, and vision inspired multiple generations of scientists.

GLOSSARY AND FUNDAMENTAL CONSTANTS

Introduction

This glossary of the terms used in the *Handbook* represents to a large extent the language of optics. The symbols are representations of numbers, variables, and concepts. Although the basic list was compiled by the author of this section, all the editors have contributed and agreed to this set of symbols and definitions. Every attempt has been made to use the same symbols for the same concepts throughout the entire *Handbook*, although there are exceptions. Some symbols seem to be used for many concepts. The symbol α is a prime example, as it is used for absorptivity, absorption coefficient, coefficient of linear thermal expansion, and more. Although we have tried to limit this kind of redundancy, we have also bowed deeply to custom.

Units

The abbreviations for the most common units are given first. They are consistent with most of the established lists of symbols, such as given by the International Standards Organization ISO¹ and the International Union of Pure and Applied Physics, IUPAP.²

Prefixes

Similarly, a list of the numerical prefixes¹ that are most frequently used is given, along with both the common names (where they exist) and the multiples of ten that they represent.

Fundamental Constants

The values of the fundamental constants³ are listed following the sections on SI units.

Symbols

The most commonly used symbols are then given. Most chapters of the *Handbook* also have a glossary of the terms and symbols specific to them for the convenience of the reader. In the following list, the symbol is given, its meaning is next, and the most customary unit of measure for the quantity is presented in brackets. A bracket with a dash in it indicates that the quantity is unitless. Note that there is a difference between units and dimensions. An angle has units of degrees or radians and a solid angle square degrees or steradians, but both are pure ratios and are dimensionless. The unit symbols as recommended in the SI system are used, but decimal multiples of some of the dimensions are sometimes given. The symbols chosen, with some cited exceptions, are also those of the first two references.

RATIONALE FOR SOME DISPUTED SYMBOLS

The choice of symbols is a personal decision, but commonality improves communication. This section explains why the editors have chosen the preferred symbols for the *Handbook*. We hope that this will encourage more agreement.

Fundamental Constants

It is encouraging that there is almost universal agreement for the symbols for the fundamental constants. We have taken one small exception by adding a subscript B to the k for Boltzmann's constant.

Mathematics

We have chosen i as the imaginary almost arbitrarily. IUPAP lists both i and j , while ISO does not report on these.

Spectral Variables

These include expressions for the wavelength λ , frequency ν , wave number σ , ω for circular or radian frequency, k for circular or radian wave number and dimensionless frequency x . Although some use f for frequency, it can be easily confused with electronic or spatial frequency. Some use $\bar{\nu}$ for wave number, but, because of typography problems and agreement with ISO and IUPAP, we have chosen σ ; it should not be confused with the Stefan-Boltzmann constant. For spatial frequencies we have chosen ξ and η , although f_x and f_y are sometimes used. ISO and IUPAP do not report on these.

Radiometry

Radiometric terms are contentious. The most recent set of recommendations by ISO and IUPAP are L for radiance [$\text{Wcm}^{-2}\text{sr}^{-1}$], M for radiant emittance or exitance [Wcm^{-2}], E for irradiance or incidence [Wcm^{-2}], and I for intensity [Wsr^{-2}]. The previous terms, W , H , N , and J , respectively, are still in many texts, notably Smith⁴ and Lloyd⁵ but we have used the revised set, although there are still shortcomings. We have tried to deal with the vexatious term *intensity* by using *specific intensity* when the units are $\text{Wcm}^{-2}\text{sr}^{-1}$, *field intensity* when they are Wcm^{-2} , and *radiometric intensity* when they are Wsr^{-1} .

There are two sets to terms for these radiometric quantities, which arise in part from the terms for different types of reflection, transmission, absorption, and emission. It has been proposed that the *ion* ending indicate a process, that the *ance* ending indicate a value associated with a particular sample, and that the *ivity* ending indicate a generic value for a "pure" substance. Then one also has reflectance, transmittance, absorptance, and emittance as well as reflectivity, transmissivity, absorptivity, and emissivity. There are now two different uses of the word emissivity. Thus the words *exitance*, *incidence*, and *sterance* were coined to be used in place of emittance, irradiance, and radiance. It is interesting that ISO uses radiance, exitance, and irradiance whereas IUPAP uses radiance, exitance [*sic*], and irradiance. We have chosen to use them both, i.e., emittance, irradiance, and radiance will be followed in square brackets by exitance, incidence, and sterance (or vice versa). Individual authors will use the different endings for transmission, reflection, absorption, and emission as they see fit.

We are still troubled by the use of the symbol E for irradiance, as it is so close in meaning to electric field, but we have maintained that accepted use. The spectral concentrations of these quantities, indicated by a wavelength, wave number, or frequency subscript (e.g., L_λ) represent partial differentiations; a subscript q represents a photon quantity; and a subscript v indicates a quantity normalized to the response of the eye. Thereby, L_v is luminance, E_v illuminance, and M_v and I_v luminous emittance and luminous intensity. The symbols we have chosen are consistent with ISO and IUPAP.

The refractive index may be considered a radiometric quantity. It is generally complex and is indicated by $\bar{n} = n - ik$. The real part is the relative refractive index and k is the extinction coefficient. These are consistent with ISO and IUPAP, but they do not address the complex index or extinction coefficient.

Optical Design

For the most part ISO and IUPAP do not address the symbols that are important in this area.

There were at least 20 different ways to indicate focal ratio; we have chosen FN as symmetrical with NA; we chose f and efl to indicate the effective focal length. Object and image distance, although given many different symbols, were finally called s_o and s_i since s is an almost universal symbol for distance. Field angles are θ and ϕ ; angles that measure the slope of a ray to the optical axis are u ; u can also be $\sin u$. Wave aberrations are indicated by W_{ijk} , while third-order ray aberrations are indicated by σ_i and more mnemonic symbols.

Electromagnetic Fields

There is no argument about **E** and **H** for the electric and magnetic field strengths, **Q** for quantity of charge, ρ for volume charge density, σ for surface charge density, etc. There is no guidance from Refs. 1 and 2 on polarization indication. We chose $\perp$ and $\parallel$ rather than p and s , partly because s is sometimes also used to indicate scattered light.

There are several sets of symbols used for reflection transmission, and (sometimes) absorption, each with good logic. The versions of these quantities dealing with field amplitudes are usually specified with lower case symbols: r , t , and a . The versions dealing with power are alternately given by the uppercase symbols or the corresponding Greek symbols: R and T versus ρ and τ . We have chosen to use the Greek, mainly because these quantities are also closely associated with Kirchhoff's law that is usually stated symbolically as $\alpha = \epsilon$. The law of conservation of energy for light on a surface is also usually written as $\alpha + \rho + \tau = 1$.

Base SI Quantities

length	m	meter
time	s	second
mass	kg	kilogram
electric current	A	ampere
temperature	K	kelvin
amount of substance	mol	mole
luminous intensity	cd	candela

Derived SI Quantities

energy	J	joule
electric charge	C	coulomb
electric potential	V	volt
electric capacitance	F	farad
electric resistance	Ω	ohm
electric conductance	S	siemens
magnetic flux	Wb	weber
inductance	H	henry
pressure	Pa	pascal
magnetic flux density	T	tesla
frequency	Hz	hertz
power	W	watt
force	N	newton
angle	rad	radian
angle	sr	steradian

Prefixes

<i>Symbol</i>	<i>Name</i>	<i>Common name</i>	<i>Exponent of ten</i>
F	exa		18
P	peta		15
T	tera	trillion	12
G	giga	billion	9
M	mega	million	6
k	kilo	thousand	3
h	hecto	hundred	2
da	deca	ten	1
d	deci	tenth	−1
c	centi	hundredth	−2
m	milli	thousandth	−3
μ	micro	millionth	−6
n	nano	billionth	−9
p	pico	trillionth	−12
f	femto		−15
a	atto		−18

Constants

c	speed of light vacuo [$299792458 \text{ ms}^{-1}$]
c_1	first radiation constant $= 2\pi c^2 h = 3.7417749 \times 10^{-16} [\text{Wm}^2]$
c_2	second radiation constant $= hc/k = 0.014838769 [\text{mK}]$
e	elementary charge [$1.60217733 \times 10^{-19} \text{ C}$]
g_n	free fall constant [9.80665 ms^{-2}]
h	Planck's constant [$6.6260755 \times 10^{-34} \text{ Js}$]
k_B	Boltzmann constant [$1.380658 \times 10^{-23} \text{ JK}^{-1}$]
m_e	mass of the electron [$9.1093897 \times 10^{-31} \text{ kg}$]
N_A	Avogadro constant [$6.0221367 \times 10^{23} \text{ mol}^{-1}$]
R_∞	Rydberg constant [$10973731.534 \text{ m}^{-1}$]
ϵ_0	vacuum permittivity [$\mu_0^{-1} \text{ C}^{-2}$]
σ	Stefan-Boltzmann constant [$5.67051 \times 10^{-8} \text{ Wm}^{-1} \text{ K}^{-4}$]
μ_0	vacuum permeability [$4\pi \times 10^{-7} \text{ NA}^{-2}$]
μ_B	Bohr magneton [$9.2740154 \times 10^{-24} \text{ JT}^{-1}$]

General

B	magnetic induction [$\text{Wbm}^{-2}, \text{kg s}^{-1} \text{ C}^{-1}$]
C	capacitance [$\text{f}, \text{C}^2 \text{ s}^2 \text{ m}^{-2} \text{ kg}^{-1}$]
C	curvature [m^{-1}]
c	speed of light in vacuo [ms^{-1}]
c_1	first radiation constant [Wm^2]
c_2	second radiation constant [mK]
D	electric displacement [Cm^{-2}]
E	incidence [irradiance] [Wm^{-2}]
e	electronic charge [coulomb]
E_v	illuminance [$\text{lux}, \text{lmm}^{-2}$]
E	electrical field strength [Vm^{-1}]
E	transition energy [J]
E^g	band-gap energy [eV]
f^g	focal length [m]
f_c	Fermi occupation function, conduction band
f_v	Fermi occupation function, valence band

FN	focal ratio (f/number) [—]
g	gain per unit length [m^{-1}]
g_{th}	gain threshold per unit length [m^{-1}]
$\mathbf{H}$	magnetic field strength [Am^{-1} , $\text{Cs}^{-1} \text{m}^{-1}$]
h	height [m]
I	irradiance (see also E) [Wm^{-2}]
I	radiant intensity [Wsr^{-1}]
I	nuclear spin quantum number [—]
I	current [A]
i	$\sqrt{-1}$
$\text{Im}()$	imaginary part of
J	current density [Am^{-2}]
$\mathbf{j}$	total angular momentum [$\text{kg m}^2 \text{s}^{-1}$]
$J_1()$	Bessel function of the first kind [—]
k	radian wave number $= 2\pi/\lambda$ [rad cm^{-1}]
$\mathbf{k}$	wave vector [rad cm^{-1}]
k	extinction coefficient [—]
L	sterance [radiance] [$\text{Wm}^{-2} \text{sr}^{-1}$]
L_v	luminance [cdm^{-2}]
L	inductance [h , $\text{m}^2 \text{kg C}^2$]
L	laser cavity length
L, M, N	direction cosines [—]
M	angular magnification [—]
M	radiant exitance [radiant emittance] [Wm^{-2}]
m	linear magnification [—]
m	effective mass [kg]
MTF	modulation transfer function [—]
N	photon flux [s^{-1}]
N	carrier (number) density [m^{-3}]
n	real part of the relative refractive index [—]
$\tilde{n}$	complex index of refraction [—]
NA	numerical aperture [—]
OPD	optical path difference [m]
P	macroscopic polarization [C m^{-2}]
$\text{Re}()$	real part of [—]
R	resistance [Ω]
$\mathbf{r}$	position vector [m]
S	Seebeck coefficient [VK^{-1}]
s	spin quantum number [—]
s	path length [m]
S_o	object distance [m]
S_i	image distance [m]
T	temperature [K , C]
t	time [s]
t	thickness [m]
u	slope of ray with the optical axis [rad]
V	Abbe reciprocal dispersion [—]
V	voltage [V , $\text{m}^2 \text{kgs}^{-2} \text{C}^{-1}$]
x, y, z	rectangular coordinates [m]
Z	atomic number [—]

Greek Symbols

α	absorption coefficient [cm^{-1}]
α	(power) absorptance (absorptivity)

ϵ	dielectric coefficient (constant) [—]
ϵ	emittance (emissivity) [—]
ϵ	eccentricity [—]
ϵ_1	Re (ϵ)
ϵ_2	Im (ϵ)
τ	(power) transmittance (transmissivity) [—]
ν	radiation frequency [Hz]
ω	circular frequency = $2\pi\nu$ [rads ⁻¹]
ω	plasma frequency [Hz]
λ	wavelength [μm , nm]
σ	wave number = $1/\lambda$ [cm ⁻¹]
σ	Stefan Boltzmann constant [Wm ⁻² K ⁻¹]
ρ	reflectance (reflectivity) [—]
θ, ϕ	angular coordinates [rad, °]
ξ, η	rectangular spatial frequencies [m ⁻¹ , r ⁻¹]
ϕ	phase [rad, °]
ϕ	lens power [m ⁻²]
Φ	flux [W]
χ	electric susceptibility tensor [—]
Ω	solid angle [sr]

Other

$\Re$	responsivity
$\exp(x)$	e^x
$\log_a(x)$	log to the base a of x
$\ln(x)$	natural log of x
$\log(x)$	standard log of x : $\log_{10}(x)$
Σ	summation
Π	product
Δ	finite difference
δx	variation in x
dx	total differential
∂x	partial derivative of x
$\delta(x)$	Dirac delta function of x
δ_{ij}	Kronecker delta

REFERENCES

1. Anonymous, *ISO Standards Handbook 2: Units of Measurement*, 2nd ed., International Organization for Standardization, 1982.
2. Anonymous, *Symbols, Units and Nomenclature in Physics*, Document U.I.P. 20, International Union of Pure and Applied Physics, 1978.
3. E. Cohen and B. Taylor, "The Fundamental Physical Constants," *Physics Today*, 9 August 1990.
4. W. J. Smith, *Modern Optical Engineering*, 2nd ed., McGraw-Hill, 1990.
5. J. M. Lloyd, *Thermal Imaging Systems*, Plenum Press, 1972.

William L. Wolfe
College of Optical Sciences
University of Arizona
Tucson, Arizona

PART

1

MEASUREMENTS

This page intentionally left blank

SCATTEROMETERS

John C. Stover

*The Scatter Works, Inc.
Tucson, Arizona*

1.1 GLOSSARY

BRDF	bidirectional reflectance distribution function
BTDF	bidirectional transmittance distribution function
BSDF	bidirectional scatter distribution function
f	focal length
L	distance
P	power
R	length
r	radius
TIS	total integrated scatter
θ	angle
θ_N	vignetting angle
θ_{spec}	specular angle
λ	wavelength
σ	rms roughness
Ω	solid angle

1.2 INTRODUCTION

In addition to being a serious source of noise, scatter reduces throughput, limits resolution, and has been the unexpected source of practical difficulties in many optical systems. On the other hand, its measurement has proved to be an extremely sensitive method of providing metrology information for components used in many diverse applications. Measured scatter is a good indicator of surface quality and can be used to characterize surface roughness as well as locate and size

discrete defects. It is also used to measure the quality of optical coatings and bulk optical materials. This chapter reviews basic issues associated with scatter metrology and touches on various industrial applications.

The pioneering scattering instrumentation^{1–32} work started in the 1960s and extended into the 1990s. This early work (reviewed in 1995)¹¹ resulted in commercially available lab scatterometers and written standards in SEMI and ASTM detailing measurement, calibration and reporting.^{33–36} Understanding the measurements and the ability to repeat results and communicate them led to an expansion of industrial applications, scatterometry has become an increasingly valuable source of noncontact metrology in industries where surface inspection is important. For example, each month millions of silicon wafers (destined to be processed into computer chips) are inspected for point defects (pits and particles) with “particle scanners,” which are essentially just scatterometers. These rather amazing instruments (now costing more than \$1 million each) map wafer defects smaller than 50 nm and can distinguish between pits and particles. In recent years their manufacture has matured to the point where system specifications and calibration are now also standardized in SEMI.^{37–40} Scatter metrology is also found in industries as diverse as medicine, sheet metal production and even the measurement of appearance—where it has been noted that while beauty is in the eye of the beholder, what we see is scattered light. The polarization state of scatter signals has also been exploited^{25–28, 41–44} and is providing additional product information. Many more transitions from lab scatterometer to industry application are expected. They depend on understanding the basic measurement concepts outlined in this chapter.

Although it sounds simple, the instrumentation required for these scatter measurements is fairly sophisticated. Scatter signals are generally small compared to the specular beam and can vary by several orders of magnitude in just a few degrees. Complete characterization may require measurement over a large fraction of the sphere surrounding the scatter source. For many applications, a huge array of measurement decisions (incident angle, wavelength, source and receiver polarization, scan angles, etc.) faces the experimenter. The instrument may faithfully record a signal, but is it from the sample alone? Or, does it also include light from the instrument, the wall behind the instrument, and even the experimenter’s shirt? These are not easy questions to answer at nanowatt levels in the visible and get even harder in the infrared and ultraviolet. It is easy to generate scatter data—lots of it. Obtaining accurate values of appropriate measurements and communicating them requires knowledge of the instrumentation as well as insight into the problem being addressed.

In 1961, Bennett and Porteus¹ reported measurement of signals obtained by integrating scatter over the reflective hemisphere. They defined a parameter called the *total integrated scatter* (TIS) as the integrated reflected scatter normalized by the total reflected light. Using a scalar diffraction theory result drawn from the radar literature,² they related the TIS to the reflector root mean square (rms) roughness. By the mid-1970s, several scatterometers had been built at various university, government, and industry labs that were capable of measuring scatter as a function of angle; however, instrument operation and data manipulation were not always well automated.^{3–6} Scattered power per unit solid angle (sometimes normalized by the incident power) was usually measured. Analysis of scatter data to characterize sample surface roughness was the subject of many publications.^{7–11} Measurement comparison between laboratories was hampered by instrument differences, sample contamination, and confusion over what parameters should be compared. A derivation of what is commonly called BRDF (for bidirectional reflectance distribution function) was published by Nicodemus and coworkers in 1970, but did not gain common acceptance as a way to quantify scatter measurements until after publication of their 1977 NBS monograph.¹² With the advent of small powerful computers in the 1980s, instrumentation became more automated. Increased awareness of scatter problems and the sensitivity of many end-item instruments increased government funding for better instrumentation.^{13–14} As a result, instrumentation became available that could measure and analyze as many as 50 to 100 samples a day instead of just a handful. Scatterometers became commercially available and the number (and sophistication) of measurement facilities increased.^{15–17} Further instrumentation improvements will include more out-of-plane capability, extended wavelength control, and polarization control at both source and receiver. As of 2008 there are written standards for BRDF and TIS in ASTM and SEMI.^{33–36}

This review gives basic definitions, instrument configurations, components, scatter specifications, measurement techniques, and briefly discusses calibration and error analysis.

1.3 DEFINITIONS AND SPECIFICATIONS

One of the difficulties encountered in comparing measurements made on early instruments was getting participants to calculate the same quantities. There were problems of this nature as late as 1988 in a measurement round-robin run at 633 nm.²⁰ But, there are other reasons for reviewing these basic definitions before discussing instrumentation. The ability to write useful scatter specifications (i.e., the ability to make use of quantified scatter information) depends just as much on understanding the defined quantity as it does on understanding the instrumentation and the specific scatter problem. In addition, definitions are often given in terms of mathematical abstractions that can only be approximated in the lab. This is the case for BRDF.

$$\text{BRDF} = \frac{\text{differential radiance}}{\text{differential irradiance}} \approx \frac{dP_s/d\Omega}{P_i \cos \theta_s} \approx \frac{P_s/\Omega}{P_i \cos \theta_s} \quad (1)$$

BRDF has been strictly defined as the ratio of the sample differential radiance to the differential irradiance under the assumptions of a collimated beam with uniform cross section incident on an isotropic surface reflector (no bulk scatter allowed). Under these conditions, the third quantity in Eq. (1) is found, where power P in watts instead of intensity I in W/m^2 has been used. The geometry is shown in Fig. 1. The value θ_s is the polar angle in the scatter direction measured from reflector normal and Ω is the differential solid angle (in steradians) through which dP_s (watts) scatters when P_i (watts) is incident on the reflector. The cosine comes from the definition of radiance and may be viewed as a correction from the actual size of the scatter source to the apparent size (or projected area) as the viewer rotates away from surface normal.

The details of the derivation do not impact scatter instrumentation, but the initial assumptions and the form of the result do. When scattered light power is measured, it is through a finite diameter

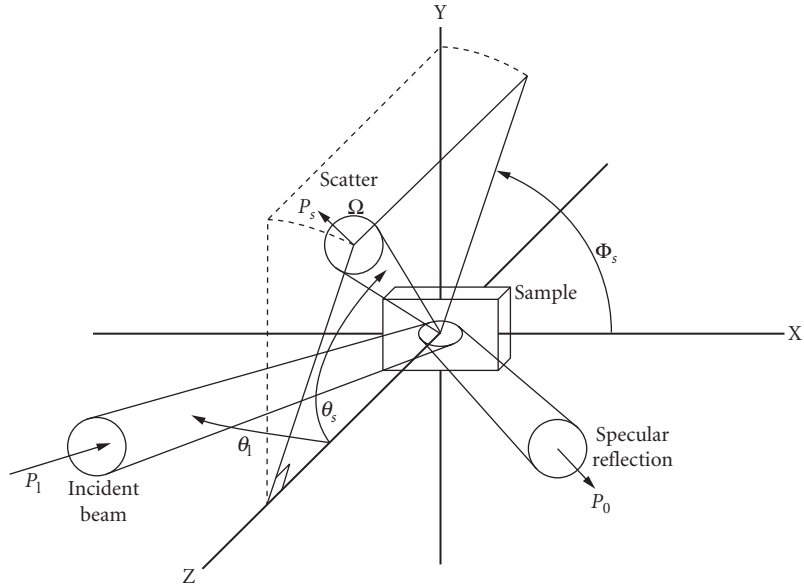


FIGURE 1 Geometry for the definition of BRDF.

aperture; as a result the calculation is for an average BRDF over the aperture. This is expressed in the final term of Eq. (1), where P_s is the measured power through the finite solid angle Ω defined by the receiver aperture and the distance to the scatter source. Thus, when the receiver aperture is swept through the scatter field to obtain angle dependence, the measured quantity is actually the convolution of the aperture over the differential BRDF. This does not cause serious distortion unless the scatter field has abrupt intensity changes, as it does near specular or near diffraction peaks associated with periodic surface structure. But there are even more serious problems between the strict definition of BRDF (as derived by Nicodemus) and practical measurements. There are no such things as uniform cross-section beams and isotropic samples that scatter only from surface structure. So, the third term of Eq. (1) is not exactly the differential radiance/irradiance ratio for the situations we create in the lab with our instruments. However, it makes perfect sense to measure normalized scattered power density as a function of direction [as defined in the fourth term of Eq. (1)] even though it cannot be exactly expressed in convenient radiometric terms.

A slightly less cumbersome definition (in terms of writing scatter specifications) is realized if the cosine term is dropped. This is referred to as “the cosine-corrected BRDF,” or sometimes, “the scatter function.” Its use has caused some of the confusion surrounding measurement differences found in scatter round robins. In accordance with the original definition, accepted practice, and BRDF Standards,^{33,36} the BRDF contains the cosine, as given in Eq. (1), and the cosine-corrected BRDF does not. It also makes sense to extend the definition to volume scatter sources and even make measurements on the transmissive side of the sample. The term BTDF (for bidirectional transmission distribution function) is used for transmissive scatter, and BSDF (bidirectional scatter distribution function) is all-inclusive.

The BSDF has units of inverse steradians and, unlike reflectance and transmission (which vary from 0.0 to 1.0), can take on very large values as well as very small values.^{1,21} For near-normal incidence, a measurement made at the specular beam of a high reflectance mirror results in a BSDF value of approximately $1/\Omega$, which is generally a large number. Measured values at the specular direction on the order of 10^6 sr^{-1} are common for a HeNe laser source. For low-scatter measurements, large apertures are generally used and values fall to the noise equivalent BSDF (or NEBSDF). This level depends on incident power and polar angle (position) as well as aperture size and detector noise, and typically varies from 10^{-4} sr^{-1} to 10^{-10} sr^{-1} . Thus, the measured BSDF can easily vary by over a dozen orders of magnitude in a given angle scan. This large variation results in challenges in instrumentation design as well as data storage, analysis, and presentation, and is another reason for problems with comparison measurements.

Instrument signature is the measured background scatter signal caused by the instrument and not the sample. It is caused by a combination of scatter created within the instrument and by the NEBSDF. Any instrument scatter that reaches the receiver field of view (FOV) will contribute to it. Common causes are scatter from source optics and the system beam dump. It is typically measured without a sample in place; however, careful attention has to be paid to the receiver FOV to ascertain that this is representative of the sample measurement situation. It is calculated as though the signal came from the sample (i.e., the receiver/sample solid angle is used) so that it can be compared to the measured sample BSDF. Near specular, the signal can be dominated by scatter (or diffraction) contributions from the source, called instrument signature. At high scatter angles it can generally be limited to NEBSDF levels. Sample measurements are always a combination of desired signal and instrument signature. Reduction of instrument signature, especially near specular, is a prime consideration in instrument design and use.

BSDF specifications always require inclusion of incident angle, source wavelength, and polarization as well as observation angles, scatter levels, and sample orientation. Depending on the sample and the measurement, they may also require aperture information to account for convolution effects. Specifications for scatter instrumentation should include instrument signature limits and the required NEBSDF. Specifications for the NEBSDF must include the polar angle, the solid angle, and the incident power to be meaningful.

TIS measurements are made by integrating the BSDF over a majority of either the reflective or transmissive hemispheres surrounding the scatter source. This is usually done with instrumentation that gathers (integrates) the scattered light signal. The TIS can sometimes be calculated from BSDF

data. If an isotropic sample is illuminated at near-normal incidence with circularly polarized light, data from a single measurement scan is enough to calculate a reasonably accurate TIS value for an entire hemisphere of scatter. The term “total integrated scatter” is a misnomer in that the integration is never actually “total,” as some scatter is never measured. Integration is commonly performed from a few degrees from specular to polar angles approaching 90° (approaching 1° to more than 45° in the SEMI Standard).³⁴ Measurements can be made of either transmissive or reflective scatter. TIS is calculated by ratioing the integrated scatter to the reflected (or transmitted) power as shown below in Eq. (2). For optically smooth components, the scatter signal is small compared to the specular beam and is often ignored. For reflective scatter the conversion to rms roughness (σ) under the assumption of an optically smooth, clean, reflective surface, via Davies’ scalar theory,² is also given. This latter calculation does not require gaussian surface statistics (as originally assumed by Davies) or even surface isotropy, but will work for other distributions, including gratings and machined optics.^{8,11} There are other issues (polarization and the assumption of mostly near specular scatter) that cause some error in this conversion. Comparison of TIS-generated roughness to profile-generated values is made difficult by a number of issues (bandwidth limits, one-dimensional profiling of a two-dimensional surface, etc.) that are beyond the scope of this section (see Ref. 11 for a complete discussion). One additional caution is that the literature and more than one stray radiation analysis program define TIS as scattered power normalized by incident power (which is essentially diffuse reflectance). This is a seriously incorrect distortion of TIS. Such a definition obviously cannot be related to surface roughness, as a change in reflectance (but not roughness) will change the ratio.

$$\text{TIS} = \frac{\text{integrated scattered power}}{\text{total reflected power}} \cong \left(\frac{4\pi\sigma}{\lambda} \right)^2 \quad (2)$$

TIS is one of three ratios that may be formed from the incident power, the specular reflected (or transmitted) power, and the integrated scatter. The other two ratios are the diffuse reflectance (or transmittance) and the specular reflectance (or transmittance). Typically, all three ratios may be obtained from measurements taken in TIS or BSDF instruments. Calculation, or specification, of any of these quantities that involve integration of scatter, also requires that the integration limits be given, as well as the wavelength, angle of incidence, source polarization, and sample orientation.

1.4 INSTRUMENT CONFIGURATIONS AND COMPONENT DESCRIPTIONS

The scatterometer shown in Fig. 2 is representative of the most common instrument configuration in use. The source is fixed in position. The sample is rotated to the desired incident angle, and the receiver is rotated about the sample in the plane of incidence. Although dozens of instruments have been built following this general design, other configurations are in use. For example, the source and receiver may be fixed and the sample rotated so that the scatter pattern moves past the receiver. This is easier mechanically than moving the receiver at the end of an arm, but complicates analysis because the incident angle and the observation angle change simultaneously. Another combination is to fix the source and sample together, at constant incident angle, and rotate this unit (about the point of illumination on the sample) so that the scatter pattern moves past a fixed receiver. This has the advantage that a long receiver/sample distance can be used without motorizing a long (heavy) receiver arm. It has the disadvantage that heavy (or multiple) sources are difficult to deal with. Other configurations, with everything fixed, have been designed that employ several receivers to merely sample the BSDF and display a curve fit of the resulting data. This is an economical solution if the BSDF is relatively uniform without isolated diffraction peaks. Variations on this last combination are common in industry where the samples are moved through the beam (sometimes during the manufacturing process) and the scatter measured in one or more directions.

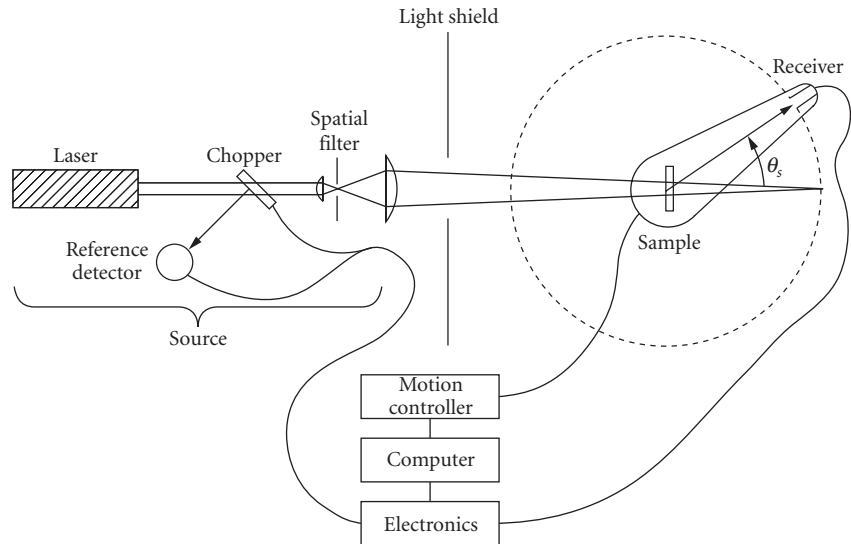


FIGURE 2 Components of a typical BSDF scatterometer.

Computer control of the measurement is essential to maximize versatility and minimize measurement time. The software required to control the measurement plus display and analyze the data can be expected to be a significant portion of total instrument development cost. The following paragraphs review typical design features (and issues) associated with the source, sample mount, and receiver components.

The source in Fig. 2 is formed by a laser beam that is chopped, spatially filtered, expanded, and finally brought to a focus on the receiver path. The beam is chopped to reduce both optical and electronic noise. This is accomplished through the use of lock-in detection in the electronics package which suppresses all signals except those at the chopping frequency. Low-noise, programmable gain electronics are essential to reducing NEBSDF. The reference detector is used to allow the computer to ratio out laser power fluctuations and, in some cases, to provide the necessary timing signal to the lock-in electronics. Polarizers, wave plates, and neutral density filters are also commonly placed prior to the spatial filter when required in the source optics. The spatial filter removes scatter from the laser beam and presents a point source which is imaged by the final focusing element to the detector zero position. Although a lens is shown in Fig. 2, the use of a mirror, which works over a larger range of wavelengths and generally scatters less light, is more common. For most systems the relatively large focal length of the final focusing element allows use of a spherical mirror and causes only minor aberration. Low-scatter spherical mirrors are easier to obtain than other conic sections. The incident beam is typically focused at the receiver to facilitate near specular measurement. Another option (a collimated beam at the receiver) is sometimes used and will be considered in the discussion on receivers. In either case, curved samples can be accommodated by adjusting the position of the spatial filter with respect to the final focusing optic. The spot size on the sample is obviously determined by elements of the system geometry and can be adjusted by changing the focal length of the first lens (often a microscope objective). The source region is completed by a shield that isolates stray laser light from the detector.

Lasers are convenient sources, but are not necessary. Broadband sources are often required to meet a particular application or to simulate the environment where a sample will be used. Monochromators and filters can be used to provide scatterometer sources of arbitrary wavelength.²⁰ Noise floor with these tunable incoherent sources increases dramatically as the spectral bandpass is narrowed, but they have the advantage that the scatter pattern does not contain laser speckle.

The sample mount can be very simple or very complex. In principle, 6° of mechanical freedom are required to fully adjust the sample. Three translational degrees of freedom allow the sample area (or volume) of interest to be positioned at the detector rotation axis and illuminated by the source. Three rotational degrees of freedom allow the sample to be adjusted for angle of incidence, out-of-plane tilt, and rotation about sample normal. The order in which these stages are mounted affects the ease of use (and cost) of the sample holder. In practice, it often proves convenient to either eliminate, or occasionally duplicate, some of these degrees of freedom. Exact requirements for these stages differ depending on whether the sample is reflective or transmissive, as well as with size and shape. In addition, some of these axes may be motorized to allow the sample area to be raster-scanned to automate sample alignment or to measure reference samples. The order in which these stages are mounted affects the ease of sample alignment. As a general rule, the scatter pattern is insensitive to small changes in incident angle but very sensitive to small angular deviations from specular. Instrumentation should be configured to allow location of the specular reflection (or transmission) very accurately.

The receiver rotation stage should be motorized and under computer control so that the input aperture may be placed at any position on the observation circle (dotted line in Fig. 2). Data scans may be initiated at any location. Systems vary as to whether data points are taken with the receiver stopped or “on the fly.” The measurement software is less complicated if the receiver is stopped. Unlike many TIS systems, the detector is always approximately normal to the incoming scatter signal. In addition to the indicated axis of rotation, some mechanical freedom is required to ensure that the receiver is at the correct height and pointed (tilted) at the illuminated sample. Sensitivity, low noise, linearity, and dynamic range are the important issues in choosing a detector element and designing the receiver housing. In general, these requirements are better met with photovoltaic detectors than photoconductive detectors. Small area detectors reduce the NEBSDF.

Receiver designs vary, but changeable apertures, bandpass filters, polarizers, lenses, and field stops are often positioned in front of the detector element. Figure 3 shows two receiver configurations, one designed for use with a converging source and one with a collimated source. In Fig. 3a, the illuminated sample spot is imaged on a field stop in front of the detector. This configuration is commonly

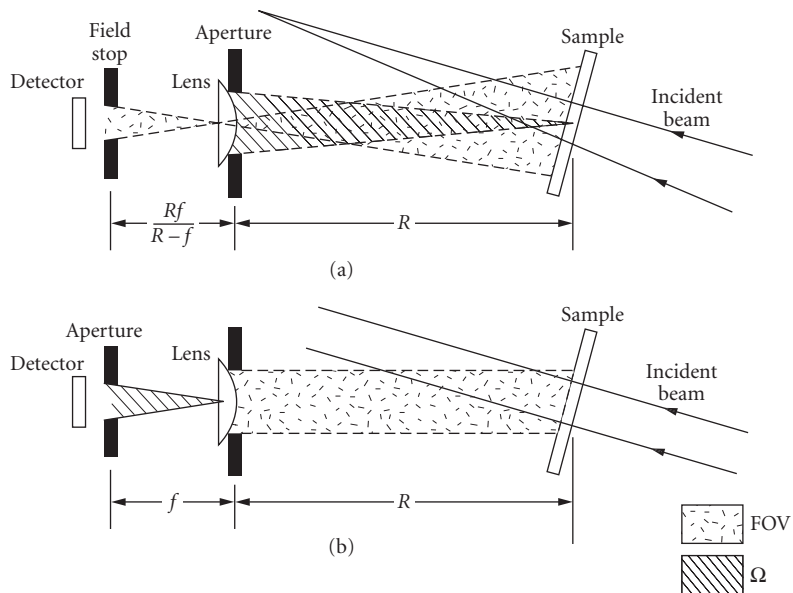


FIGURE 3 Receiver configurations: (a) converging source and (b) collimated source.

used with the source light converging on the receiver path. The field stop determines the receiver FOV. The aperture at the front of the receiver determines the solid angle over which scatter is gathered. Any light entering this aperture that originates from within the FOV will reach the detector and become part of the signal. This includes instrument signature contributions scattered through small angles by the source optics. It will also include light scattered by the receiver lens so that it appears to come from the sample. The configuration in Fig. 3a can be used to obtain near specular measurements by bringing a small receiver aperture close to the focused specular beam. With this configuration, reducing the front aperture does not limit the FOV. The receiver in Fig. 3b is in better accordance with the strict definition of BRDF in that a collimated source can be used. An aperture is located one focal length behind a collecting lens (or mirror) in front of the detector. The intent is to measure bundles of nearly parallel rays scattered from the sample. The angular spread of rays allowed to pass to the detector defines the receiver solid angle, which is equal to the aperture size divided by the focal length of the lens. This ratio (not the front aperture/sample distance) determines the solid angle of this receiver configuration. The FOV is determined by the clear aperture of the lens, which must be kept larger than the illuminated spot on the sample. The Fig. 3b design is unsuitable for near specular measurement because the relatively broad collimated specular beam will scatter from the receiver lens for several degrees from specular. It is also limited in measuring large incident angle situations where the elongated spot may exceed the FOV. If the detector (and its stop) can be moved in relation to the lens, receivers can be adjusted from one configuration to the other. Away from the specular beam, in low instrument signature regions, there is no difference in the measured BSDF values between the two systems. Commercially available research scatterometers are available that measure both in and out of the incident plane and from the mid-IR to the near UV.

The two common methods of approaching TIS measurements are shown in Fig. 4. The first one, employed by Bennett and Porteus in their early instrument,¹ uses a hemispherical mirror (or Coblentz sphere) to gather scattered light from the sample and image it onto a detector. The specular beam enters and leaves the hemisphere through a small circular hole. The diameter of that hole defines the near specular limit of the instrument. The reflected beam (not the incident beam) should be centered in the hole because the BSDF will be symmetrical about it. Alignment of the hemispherical mirror is critical in this approach. The second approach involves the use of an integrating sphere. A section of the sphere is viewed by a recessed detector. If the detector FOV is limited to a section of the sphere

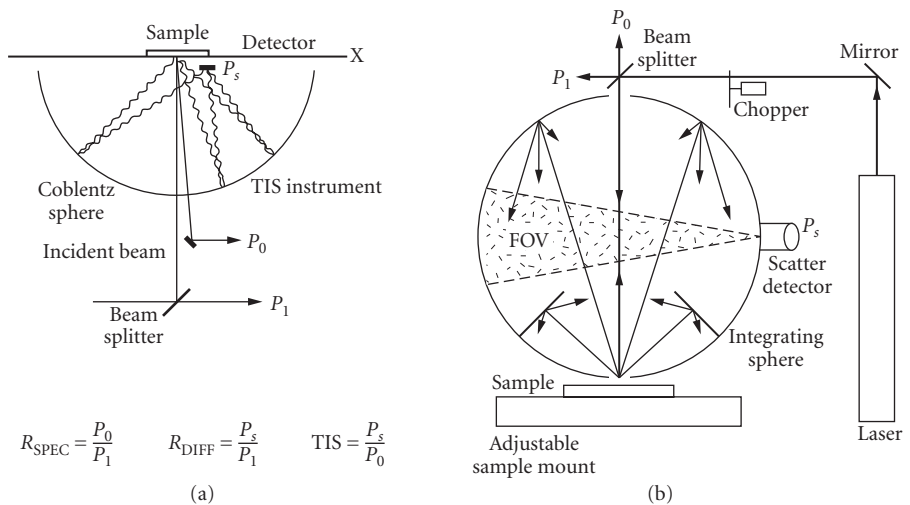


FIGURE 4 TIS measurement with a (a) Coblentz sphere and (b) diffuse integrating sphere.

closer to specular center, the measurement is dominated by the aperture convolution of the specular beam, and there is no opportunity to measure scatter. The value θ_N is easily calculated (via a small-angle approximation) using the instrument geometry and parameters identified in Fig. 5, where the receiver is shown at the θ_N position. The parameter F is the focal length of the sample.

$$\theta_N = (r_{\text{MIR}} + r_{\text{FOV}})/L + (r_{\text{FOV}} + r_{\text{apt}})/R - r_{\text{spot}}/F \quad (3)$$

It is easy to achieve values of θ_N below 10° and values as small as 1° can be realized with careful design. The offset angle from specular, θ_{spec} , at which the measurement is dominated by the specular beam, can be reduced to less than a tenth of a degree at visible wavelengths and is given by

$$\theta_{\text{spec}} = \frac{r_{\text{diff}} + r_{\text{apt}}}{R} \approx \frac{3\lambda}{D} + \frac{r_{\text{apt}}}{R} \quad (4)$$

Here, r_{diff} and r_{apt} are the radius of the focused spot and the receiver aperture, respectively (see Fig. 5 again). The value of r_{diff} can be estimated in terms of the diameter D of the focusing optic and its distance to the focused spot, $R + L$ (estimated as $2.5R$). The diffraction limit has been doubled in this estimate to allow for aberrations.

To take near specular measurements, both angles and the instrument signature need to be reduced. The natural reaction is to “increase R to increase angular resolution.” Although a lot of money has been spent doing this, it is an unnecessarily expensive approach. Angular resolution is achieved by reducing r_{apt} and by taking small steps. The radius r_{apt} can be made almost arbitrarily small so the economical way to reduce the r_{apt}/R terms is by minimizing r_{apt} —not by increasing R . A little thought about r_{FOV} and r_{diff} reveals that they are both proportional to R , so nothing is gained in the near specular game by purchasing large-radius rotary stages.

The reason for building a large-radius scatterometer is to accommodate a large FOV. This is often driven by the need to take measurements at large incident angles or by the use of broadband sources, both of which create larger spots on the sample. When viewing normal to the sample, the FOV requirements can be stringent. Because the maximum FOV is proportional to detector diameter (and limited at some point by minimum receiver lens FN), increasing R is the only open-ended design parameter available. It should be sized to accommodate the smallest detector likely to be used in the system. This will probably be in the mid-IR where uniform high-detectivity photovoltaic detectors are more difficult to obtain. On the other hand, a larger detector diameter means increased electronic noise and a larger NEBSDF.

Scatter sources of instrument signature can be reduced by these techniques.

1. Use the lowest-scatter focusing element in the source that you can afford and learn how to keep it clean. This will probably be a spherical mirror.
2. Keep the source area as “black” as possible. This especially includes the sample side of the spatial filter pinhole which is conjugate with the receiver aperture. Use a black pinhole.
3. Employ a specular beam dump that rides with your receiver and additional beam dumps to capture sample reflected and transmitted beams when the receiver has left the near specular area. Use your instrument to measure the effectiveness of your beam dumps.²⁶
4. Near specular scatter caused by dust in the air can be significantly reduced through the use of a filtered air supply over the specular beam path. A filtered air supply is essential for measuring optically smooth surfaces.

Away from specular, reduction of NEBSDF is the major factor in measuring low-scatter samples and increasing instrument quality. Measurement of visible scatter from a clean semiconductor wafer will take many instruments right down to instrument signature levels. Measurement of cross-polarized scatter requires a low NEBSDF for even high-scatter optics. For a given receiver solid angle, incident power, and scatter direction, the NEBSDF is limited by the noise equivalent power of the receiver (and associated electronics), once optical noise contributions are eliminated. The electronic contributions to NEBSDF are easily measured by simply covering the receiver aperture during a measurement.

TABLE 1 Comparison of Characteristics for Detectors Used at Different Wavelengths

Detector (2 mm dia.)	NEP (W/Hz)	Wavelength (nm)	P_i (W)	NEBSDF (sr ⁻¹)
PMT	10^{-15}	633	0.005	10^{-10}
Si	10^{-13}	633	0.005	10^{-8}
Ge	3×10^{-13}	1,320	0.001	10^{-7}
InSb	10^{-12} *	3,390	0.002	5×10^{-7}
HgMgTe	10^{-11} *	10,600	2.0	10^{-8}
Pyro	10^{-8}	10,600	2.0	10^{-5}

*Detector at 77 K.

Because the resulting signal varies in a random manner, NEBSDF should be expressed as an rms value (roughly equal to one-third of the peak level). An absolute minimum measurable scatter signal (in watts) can be found from the product of three terms: the required signal-to-noise ratio, the system noise equivalent power (or NEP given in watts per square root hertz), and the square root of the noise bandwidth (BW_n). The system NEP is often larger than the detector NEP and cannot be reduced below it. The detector NEP is a function of wavelength and increases with detector diameter. Typical detector NEP values (2-mm diameter) and wavelength ranges are shown as follows for several common detectors in Table 1. Notice that NEP tends to increase with wavelength. The noise bandwidth varies as the reciprocal of the sum of the system electronics time constant and the measurement integration time. Values of 0.1 to 10 Hz are commonly achieved. In addition to system NEP, the NEBSDF may be increased by contributions from stray source light, room lights, and noise in the reference signal. Table 1 also shows achievable rms NEBSDF values that can be realized at unity cosine, a receiver solid angle of 0.003 sr, 1-second integration, and the indicated incident powers. This column can be used as a rule of thumb in system design or to evaluate existing equipment. Simply adjust by the appropriate incident power, solid angle, and so on, to make the comparison. Adjusted values substantially higher than these indicate there is room for system improvement (don't worry about differences as small as a factor of 2). Further reduction of the instrument signature under these geometry and power conditions will require dramatically increased integration time (because of the square root dependence on noise bandwidth) and special attention to electronic dc offsets. Because the NEP tends to increase with wavelength, higher powers are needed in the mid-IR to reach the same NEBSDFs that can be realized in the visible. Because scatter from many sources tends to decrease at longer wavelengths, a knowledge of the instrument NEBSDF is especially critical in the mid-IR.

As a final configuration comment, the software package (both measurement and analysis) is crucial for an instrument that is going to be used for any length of time. Poor software will quickly cost work-years of effort due to errors, increased measurement and analysis time, and lost business. Expect to expend 1 to 2 man-years with experienced programmers writing a good package—it is worth it.

1.6 MEASUREMENT ISSUES

Sample measurement should be preceded (and sometimes followed) by a measurement of the instrument signature. This is generally accomplished by removing the sample and measuring the apparent BSDF from the sample as a transmissive scan. This is not an exact measure of instrument noise during sample measurement, but if the resulting BSDF is multiplied by sample reflectance (or transmission) before comparison to sample data, it can define some hard limits over which the sample data cannot be trusted. The signature should also be compared to the NEBSDF value obtained with the receiver aperture blocked. Obtaining the instrument signature also presents an opportunity to measure the incident power, which is required for calculation of the BSDF. The ability to see the data displayed as it is taken is an extremely helpful feature when it comes to reducing instrument signature and eliminating measurement setup errors.

Angle scans, which have dominated the preceding discussion, are an obvious way to take measurements. BSDF is also a function of position on the sample, source wavelength, and source polarization, and scans can also be taken at fixed angle (receiver position) as a function of these variables. Obviously, a huge amount of data is required to completely characterize scatter from a sample.

Raster scans are taken to measure sample uniformity or locate (map) discrete defects. A common method is to fix the receiver position and move the sample in its own x - y plane, recording the BSDF at each location. Faster approaches involve using multiple detectors (e.g., array cameras) with large area illumination, and scanning the source over the sample. Results can be presented using color maps or 3D isometric plots. Results can be further analyzed via histograms and various image-processing techniques.

There are three obvious choices for making wavelength scans. Filters (variable or discrete) can be employed at the source or receiver. A monochromator can be used as a source.²⁰ Finally, there is some advantage to using a Fourier transforming infrared spectrometer (FTIR) as a source in the mid-IR.²⁰ Details of these techniques are beyond the scope of this discussion; however, a couple of generalities will be mentioned. Even though these measurements often involve relatively large bandwidths at a given wavelength (compared to a laser), the NEBSDF is often larger by a few orders because of the smaller incident power. Further, because the bandwidths change differently between the various source types given above, meaningful measurement comparisons between instruments are often difficult to make.

Polarization scans are often limited to SS, SP, PS, and PP (source/receiver) combinations. However, complete polarization dependence of the sample requires the measurement of the sample Mueller matrix. This is found by creating a set of Stokes vectors at the source and measuring the resulting Stokes vector in the desired scatter direction.^{10,11,25–28} This is an area of instrumentation development that is the subject of increasing attention.^{41–44}

Speckle effects in the BSDF from a laser source can be eliminated in several ways. If a large receiver solid angle is used (generally several hundred speckles in size) there is not a problem. The sample can be rotated about its normal so that speckle is time averaged out of the measurement. This is still a problem when measuring very near the specular beam because sample rotation unavoidably moves the beam slightly during the measurement. In this case, the sample can be measured several times at slightly different orientations and the results averaged to form one speckle-free BSDF.

Scatter measurement in the retrodirection (back into the incident beam) has been of increasing interest in recent years and represents an interesting measurement challenge. Measurement requires the insertion of a beam splitter in the source. This also scatters light and, because it is closer to the receiver than the sample, dramatically raises the NEBSDF. Diffuse samples can be measured this way, but not much else. A clever (high tech) Doppler-shift technique, employing a moving sample, has been reported²⁹ that allows separation of beam-splitter scatter from sample scatter and allows measurement of mirror scatter. A more economical (low tech) approach simply involves moving the source chopper to a location between the receiver and the sample.³⁰ Beam-splitter scatter is now dc and goes unnoticed by the ac-sensitive receiver. Noise floor is now limited by scatter from the chopper which must be made from a low-scatter, specular, absorbing material. Noise floors as low as $3 \times 10^{-8} \text{ sr}^{-1}$ have been achieved.

1.7 INCIDENT POWER MEASUREMENT, SYSTEM CALIBRATION, AND ERROR ANALYSIS

Regardless of the type of BSDF measurement, the degree of confidence in the results is determined by instrument calibration, as well as by attention to the measurement limitations previously discussed. Scatter measurements have often been received with considerable skepticism. In part, this has been due to misunderstanding of the definition of BSDF and confusion about various measurement subtleties, such as instrument signature or aperture convolution. However, quite often the measurements have been wrong and the skepticism is justified.

Instrument calibration is often confused with the measurement of P_i , which is why these topics are covered in the same section. To understand the source of this confusion, it is necessary to first consider the various quantities that need to be measured to calculate the BSDF. From Eq. (1), they

are P_s , θ_s , Ω , and P_i . The first two require measurement over a wide range of values. In particular, P_s , which may vary over many orders of magnitude, is a problem. In fact, linearity of the receiver to obtain a correct value of P_s is a key calibration issue. Notice that an absolute measurement of P_i is not required, as long as the P_s/P_i ratio is correctly evaluated. P_i and Ω generally take on only one (or just a few) discrete values during a data scan. The value of Ω is determined by system geometry. The value of P_i is generally measured in one of two convenient ways.¹¹

The first technique, sometimes referred to as *the absolute method* makes use of the scatter detector (and sometimes a neutral density filter) to directly measure the power incident upon the sample. This method relies on receiver linearity (as does the overall calibration of BSDF) and on filter accuracy when one is used. The second technique, sometimes referred to as *the reference method* makes use of a known BSDF reference sample (usually a diffuse reflector and unfortunately often referred to as the “calibration sample”) to obtain the value of P_i . Scatter from the reference sample is measured and the result used to infer the value of P_i via Eq. (1). The $P_i\Omega$ product may be evaluated this way. This method depends on knowing the absolute BSDF of the reference. Both techniques become more difficult in the mid-IR, where “known” neutral density filters and “known” reference samples are difficult to obtain. Reference sample uniformity in the mid-IR is often the critical issue and care must be exercised. Variations at 10.6 μm as large as 7:1 have been observed across the face of a diffuse gold reference “of known BRDF.” The choice of measurement methods is usually determined by whether it is more convenient to measure the BSDF of a reference or the total power P_i . Both are equally valid methods of obtaining P_i . However, neither method constitutes a system calibration, because calibration issues such as an error analysis and a linearity check over a wide range of scatter values are not addressed over the full range of BSDF angles and powers when P_i is measured (or calculated). The use of a reference sample is an excellent system check regardless of how P_i is obtained.

System linearity is a key part of system calibration. In order to measure linearity, the receiver transfer characteristic, signal out as a function of light in, must be found. This may be done through the use of a known set of neutral density filters or through the use of a comparison technique³¹ that makes use of two data scans—with and without a single filter. However, there are other calibration problems than just linearity. The following paragraph outlines an error analysis for BSDF systems.

Because the calculation of BSDF is very straightforward, the sources of error can be examined through a simple analysis^{11,32} under the assumption that the four defining parameters are independent.

$$\frac{\Delta \text{BSDF}}{\text{BSDF}} = \left[\left(\frac{\Delta P_s}{P_s} \right)^2 + \left(\frac{\Delta P_i}{P_i} \right)^2 + \left(\frac{\Delta \Omega}{\Omega} \right)^2 + \left(\frac{\Delta \theta_s \sin \theta_s}{\cos^2 \theta_s} \right)^2 \right]^{1/2} \quad (5)$$

In similar fashion, each of these terms may be broken into the components that cause errors in it. When this is done, the total error may be found as a function of angle. Two high-error regions are identified. The first is the near specular region (inside 1°), where errors are dominated by the accuracy to which the receiver aperture can be located in the cross-section direction. Or, in other words, did the receiver scan exactly through the specular beam, or did it just miss it? The second relatively high error region is near the sample plane where $\cos \theta_s$ approaches zero. In this region, a small error in angular position results in a large error in calculated BSDF. These errors are often seen in BSDF data as an abrupt increase in calculated BSDF in the grazing scatter direction, the result of division by a very small cosine into the signal gathered by a finite receiver aperture (and/or a dc offset voltage in the detector electronics). This is another example where use of the cosine-corrected BSDF makes more sense.

Accuracy is system dependent; however, at signal levels well above the NEBSDF, uncertainties less than ± 10 percent can be obtained away from the near specular and grazing directions. With expensive electronics and careful error analysis, these inaccuracies can be reduced to the ± 1 percent level.

Full calibration is not required on a daily basis. Sudden changes in instrument signature are an indication of possible calibration problems. Measurement of a reference sample that varies over several orders of magnitude is a good system check. It is prudent to take such a reference scan with data sets in case the validity of the data is questioned at a later time. A diffuse sample, with nearly constant BRDF, is a good reference choice for the measurement of P_i but a poor one for checking system calibration.

1.8 SUMMARY

The art of scatter measurement has evolved to an established form of metrology within the optics industry. Because scatter measurements tend to be a little more complicated than many other optical metrology procedures, a number of key issues must be addressed to obtain useful information. System specifications and measurements need to be given in terms of accepted, well-defined (and understood) quantities (BSDF, TIS, etc.). All parameters associated with a measurement specification need to be given (such as angle limits, receiver solid angles, noise floors, wavelength, etc.). Measurement of near specular scatter and/or low BSDF values are particularly difficult and require careful attention to instrument signature values; however, if the standardized procedures are followed, the result will be repeatable, accurate data.

TIS and BSDF are widely accepted throughout the industry and their measurement is defined by SEMI and ASTM standards. Scatter measurements are used routinely as a quality check on optical components. BSDF specifications are now often used (as they should be) in place of scratch/dig or rms roughness, when scatter is the issue. Conversion of surface scatter data to other useful formats, such as surface roughness statistics, is commonplace. The sophistication of the instrumentation (and analysis) applied to these problems is still increasing. Out-of-plane measurements and polarization-sensitive measurements are two areas that are experiencing rapid advances. Measurement of scatter outside the optics community is also increasing. Although the motivation for scatter measurement differs in industrial situations, the basic measurement and instrumentation issues encountered are essentially the ones described here.

1.9 REFERENCES

1. H. E. Bennett and J. O. Porteus, "Relation between Surface Roughness and Specular Reflectance at Normal Incidence," *J. Opt. Soc. Am.* **51**:123 (1961).
2. H. Davies, "The Reflection of Electromagnetic Waves from a Rough Surface," *Proc. Inst. Elec. Engrs.* **101**:209 (1954).
3. J. C. Stover, "Roughness Measurement by Light Scattering," in A. J. Glass and A. H. Guenther (eds.), *Laser Induced Damage in Optical Materials*, U. S. Govt. Printing Office, Washington, D. C., 1974, p. 163.
4. J. E. Harvey, "Light Scattering Characteristics of Optical Surfaces," Ph.D. dissertation, University of Arizona, 1976.
5. E. L. Church, H. A. Jenkinson, and J. M. Zavada, "Measurement of the Finish of Diamond-Turned Metal Surfaces By Differential Light Scattering," *Opt. Eng.* **16**:360 (1977).
6. J. C. Stover and C. H. Gillespie, "Design Review of Three Reflectance Scatterometers," *Proc. SPIE* **362** (Scattering in Optical Materials):172 (1982).
7. J. C. Stover, "Roughness Characterization of Smooth Machined Surfaces by Light Scattering," *Appl. Opt.* **14** (N8):1796 (1975).
8. E. L. Church and J. M. Zavada, "Residual Surface Roughness of Diamond-Turned Optics," *Appl. Opt.* **14**:1788 (1975).
9. E. L. Church, H. A. Jenkinson, and J. M. Zavada, "Relationship between Surface Scattering and Microtopographic Features," *Opt. Eng.* **18**(2):125 (1979).
10. E. L. Church, "Surface Scattering," in M. Bass (ed.), *Handbook of Optics*, vol. I, 2d ed., McGraw-Hill, New York, 1994.
11. J. C. Stover, *Optical Scattering Measurement and Analysis*, SPIE Press, (1995—new edition to be published in 2009).
12. F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis, *Geometric Considerations and Nomenclature for Reflectance*, NBS Monograph 160, U. S. Dept. of Commerce, 1977.
13. W. L. Wolfe and F. O. Bartell, "Description and Limitations of an Automated Scatterometer," *Proc. SPIE* **362**:30 (1982).
14. D. R. Cheever, F. M. Cady, K. A. Klicker, and J. C. Stover, "Design Review of a Unique Complete Angle-Scatter Instrument (CASI)," *Proc. SPIE* **818** (Current Developments in Optical Engineering II):13 (1987).
15. P. R. Spyak and W. L. Wolfe, "Cryogenic Scattering Measurements," *Proc. SPIE* **967**:15 (1989).

16. W. L. Wolfe, K. Magee, and D. W. Wolfe, "A Portable Scatterometer for Optical Shop Use," *Proc. SPIE* **525**:160 (1985).
17. J. Rifkin, "Design Review of a Complete Angle Scatter Instrument," *Proc. SPIE* **1036**:15 (1988).
18. T. A. Leonard and M. A. Pantoliano, "BRDF Round Robin," *Proc. SPIE* **967**:22 (1988).
19. T. F. Schiff, J. C. Stover, D. R. Cheever, and D. R. Bjork, "Maximum and Minimum Limitations Imposed on BSDF Measurements," *Proc. SPIE* **967** (1988).
20. F. M. Cady, M. W. Knighton, D. R. Cheever, B. D. Swimley, M. E. Southwood, T. L. Hundtoft, and D. R. Bjork, "Design Review of a Broadband 3-D Scatterometer," *Proc. SPIE* **1753**:21 (1992).
21. K. A. Klicker, J. C. Stover, D. R. Cheever, and F. M. Cady, "Practical Reduction of Instrument Signature in Near Specular Light Scatter Measurements," *Proc. SPIE* **818**:26 (1987).
22. S. J. Wein and W. L. Wolfe, "Gaussian Apodized Apertures and Small Angle Scatter Measurements," *Opt. Eng.* **28**(3):273–280 (1989).
23. J. C. Stover and M. L. Bernt, "Very Near Specular Measurement via Incident Angle Scaling," *Proc. SPIE* **1753**:16 (1992).
24. F. M. Cady, D. R. Cheever, K. A. Klicker, and J. C. Stover, "Comparison of Scatter Data from Various Beam Dumps," *Proc. SPIE* **818**:21 (1987).
25. W. S. Bickle and G. W. Videen, "Stokes Vectors, Mueller Matrices and Polarized Light: Experimental Applications to Optical Surfaces and All Other Scatterers," *Proc. SPIE* **1530**:02 (1991).
26. T. F. Schiff, D. J. Wilson, B. D. Swimley, M. E. Southwood, D. R. Bjork, and J. C. Stover, "Design Review of a Unique Out-of-Plane Polarimetric Scatterometer," *Proc. SPIE* **1753**:33 (1992).
27. T. F. Schiff, D. J. Wilson, B. D. Swimley, M. E. Southwood, D. R. Bjork, and J. C. Stover, "Mueller Matrix Measurements with an Out-Of-Plane Polarimetric Scatterometer," *Proc. SPIE* **1746**:33 (1992).
28. T. F. Schiff, B. D. Swimley, and J. C. Stover, "Mueller Matrix Measurements of Scattered Light," *Proc. SPIE* **1753**:34 (1992).
29. Z. H. Gu, R. S. Dummer, A. A. Maradudin, and A. R. McGurn, "Experimental Study of the Opposition Effect in the Scattering of Light from a Randomly Rough Metal Surface," *Appl. Opt.* **28**(N3):537 (1989).
30. T. F. Schiff, D. J. Wilson, B. D. Swimley, M. E. Southwood, D. R. Bjork, and J. C. Stover, "Retroreflections on a Low Tech Approach to the Measurement of Opposition Effects," *Proc. SPIE* **1753**:35 (1992).
31. F. M. Cady, D. R. Bjork, J. Rifkin, and J. C. Stover, "Linearity in BSDF Measurement," *Proc. SPIE* **1165**:44 (1989).
32. F. M. Cady, D. R. Bjork, J. Rifkin, and J. C. Stover, "BRDF Error Analysis," *Proc. SPIE* **1165**:13 (1989).
33. SEMI ME1392-0305—*Guide for Angle Resolved Optical Scatter Measurements on Specular or Diffuse Surfaces.*
34. SEMI MF1048-1105—*Test Method for Measuring the Reflective Total Integrated Scatter.*
35. SEMI MF1811-0704—*Guide for Estimating the Power Spectral Density Function and Related Finish Parameters from Surface Profile Data.*
36. ASTM E2387-05 *Standard Practice for Goniometric Optical Scatter Measurements.*
37. SEMI M50-0307—*Test Method for Determining Capture Rate and False Count Rate for Surface Scanning Inspection Systems by the Overlay Method.*
38. SEMI M52-0307—*Guide for Specifying Scanning Surface Inspection Systems for Silicon Wafers for the 130 nm, 90 nm, 65 nm, and 45 nm Technology Generations.*
39. SEMI M53-0706—*Practice for Calibrating Scanning Surface Inspection Systems Using Certified Depositions of Monodisperse Polystyrene Latex Spheres on Unpatterned Semiconductor Wafer Surfaces.*
40. SEMI M58-0704—*Test Method for Evaluating DMA Based Particle Deposition Systems and Processes.*
41. T. A. Germer and C. C. Asmail, "Goniometric Optical Scatter Instrument for Out-of-Plane Ellipsometry Measurements," *Rev. Sci. Instrum.* **70**:3688–3695 (1999).
42. T. A. Germer, "Measuring Interfacial Roughness by Polarized Optical Scattering," in A. A. Maradudin (ed.), *Light Scattering and Nanoscale Surface Roughness*, Springer, New York, 2007, Chap. 10, pp. 259–284.
43. B. DeBoo, J. Sasian, and R. Chipman, "Depolarization of Diffusely Reflecting Manmade Objects," *Appl. Opt.* **44**(26):5434–5445 (2005).
44. B. DeBoo, J. Sasian, and R. Chipman, "Degree of Polarization Surfaces and Maps for Analysis of Depolarization," *Optics Express* **12**(20):4941–4958 (2004).

This page intentionally left blank

SPECTROSCOPIC MEASUREMENTS

Brian Henderson

*Department of Physics and Applied Physics
University of Strathclyde
Glasgow, United Kingdom*

2.1 GLOSSARY

A_{ba}	Einstein coefficient for spontaneous emission
a_o	Bohr radius
B_{if}	Einstein coefficient between initial state $ i\rangle$ and final state $ f\rangle$
e	charge on the electron
ED	electric dipole term
E_{DC}	Dirac Coulomb term
E_{hf}	hyperfine energy
E_n	eigenvalues of quantum state n
EQ	electric quadrupole term
$E(t)$	electric field at time t
$E(\omega)$	electric field at frequency ω
g_a	degeneracy of ground level
g_b	degeneracy of excited level
g_N	gyromagnetic ratio of nucleus
h	Planck's constant
H_{SO}	spin-orbit interaction Hamiltonian
I	nuclear spin
$I(t)$	emission intensity at time t
$\mathbf{j}$	total angular momentum vector given by $\mathbf{j} = \mathbf{l} \pm \frac{1}{2}$
l_i	orbital state
m	mass of the electron
MD	magnetic dipole term
M_N	mass of nucleus N
$n_\omega(T)$	equilibrium number of photons in a blackbody cavity radiator at angular frequency ω and temperature T

QED	quantum electrodynamics
$R_{nl}(r)$	radial wavefunction
R_∞	Rydberg constant for an infinitely heavy nucleus
s	spin quantum number with value $\frac{1}{2}$
s_i	electronic spin
T	absolute temperature
W_{ab}	transition rate in absorption transition between states a and b
W_{ba}	transition rate in emission transition from state b to state a
Z	charge on the nucleus
$\alpha = e^2/4\pi\epsilon_0\hbar c$	fine structure constant
$\Delta\omega$	natural linewidth of the transition
$\Delta\omega_D$	Doppler width of transition
ϵ_0	permittivity of free space
$\zeta(r)$	spin-orbit parameter
μ_B	Bohr magneton
$\rho(\omega)$	energy density at frequency ω
τ_R	radiative lifetime
ω	angular frequency
ω_k	mode k with angular frequency ω
$\langle f V i\rangle$	matrix element of perturbation V

2.2 INTRODUCTORY COMMENTS

The conceptual basis of optical spectroscopy and its relationship to the electronic structure of matter as presented in the chapter entitled “Optical Spectroscopy and Spectroscopic Lineshapes” in Vol. I, Chap. 10 of this *Handbook*. The chapter entitled “Optical Spectrometers” in Vol. I, Chap. 31 of this *Handbook* discusses the operating principles of optical spectrometers. This chapter illustrates the underlying themes of the earlier ones using the optical spectra of atoms, molecules, and solids as examples.

2.3 OPTICAL ABSORPTION MEASUREMENTS OF ENERGY LEVELS

Atomic Energy Levels

The interest in spectroscopic measurements of the energy levels of atoms is associated with tests of quantum theory. Generally, the optical absorption and luminescence spectra of atoms reveal large numbers of sharp lines corresponding to transitions between the stationary states. The hydrogen atom has played a central role in atomic physics because of the accuracy with which relativistic and quantum electrodynamic shifts to the one-electron energies can be calculated and measured. Tests of quantum electrodynamics usually involve transitions between low-lying energy states (i.e., states with small principal quantum number). For the atomic states $|a\rangle$ and $|b\rangle$, the absorption and luminescence lines occur at exactly the same wavelength and both spectra have the same gaussian lineshape. The $1s \rightarrow 2p$ transitions on atomic hydrogen have played a particularly prominent role, especially since the development of sub-Doppler laser spectroscopy.¹ Such techniques resulted in values of $R_\infty = 10973731.43 \text{ m}^{-1}$, 36.52 m^{-1} for the spin-orbit splitting in the $n = 2$ state and a Lamb shift of 3.53 m^{-1} in the $n = 1$ state. Accurate isotope shifts have been determined from hyperfine structure measurements on hydrogen, deuterium, and tritium.²

Helium is the simplest of the multielectron atoms, having the ground configuration ($1s^2$). The energy levels of helium are grouped into singlet and triplet systems. The observed spectra arise *within* these systems (i.e., singlet-to-singlet and triplet-to-triplet); normally transitions between singlet and triplet levels are not observed. The lowest-lying levels are 1^1S , 2^3S , 2^1S , 2^3P , and 2^1P in order of increasing energy. The $1^1S \rightarrow 2^1S$ splitting is of order 20.60 eV and transitions between these levels are not excited by photons. Transitions involving the 2^1S and 2^3S levels, respectively, and higher-lying spin singlet and spin triplet states occur at optical wavelengths. Experimental work on atomic helium has emphasized the lower-lying triplet levels, which have long excited-state lifetimes and large quantum electrodynamic (QED) shifts.

As with hydrogen, the spectra of He atoms are inhomogeneously broadened by the Doppler effect. Precision measurements have been made using two-photon laser spectroscopy (e.g., $2^3S \rightarrow n^3S$ ($n = 4 - 6$) and n^3D ($n = 3 - 6$), or laser saturation absorption spectroscopy ($2^3S \rightarrow 2^3P$ and $3^3P \rightarrow 3^3D$).³⁻⁶ The $2^1S \rightarrow 3^1P$ and two photon $2^1S \rightarrow n^1D$ ($n = 3 - 7$) spectra have been measured using dye lasers.^{7,8} The wide tune ranging of the Ti-sapphire laser and the capability for generating frequencies not easily accessible with dye lasers using frequency-generation techniques makes it an ideal laser to probe transitions starting on the 2S levels of He.⁹ Two examples are the two-photon transition $2^3S \rightarrow 3^3S$ at 855 nm and the $2^3S \rightarrow 3^3P$ transition of 389 nm. The power of Doppler-free spectroscopy is shown to advantage in measurements of the $2^3S \rightarrow 2^3P$ transition.¹⁰ Since both ^3S and ^3P are excited levels, the homogeneous width is determined by the sum of the reciprocal lifetimes of the two levels. Since both 2^3S and 2^3P levels are long-lived, the resulting homogeneous width is comparatively narrow. Figure 1a shows the Doppler-broadened profile of the $2^3S \rightarrow 2^3P$ transition of ^4He for which the FWHM is about 5.5 GHz. The inhomogeneously broadened line profile shown in Fig. 1a also shows three very weak “holes” corresponding to saturated absorption of the Ti-sapphire laser radiation used to carry out the experiment. These components correspond to the $2^3S_1 \rightarrow 3^3P_2$ and $2^3S_1 \rightarrow 3^3P_1$ transitions and their crossover resonance. The amplitude of the saturated signal is some 1–2 percent of the total absorption. The relativistic splittings including spin-orbit coupling of $^3P_0 - ^3P_1$ and $^3P_1 - ^3P_2$ are 8.1 GHz and 658.8 MHz, respectively. Frequency modulation of the laser (see Vol. I, Chap. 31 of this *Handbook*) causes the “hole” to be detected as a first derivative of the absorption line, Fig. 1b. The observed FWHM of the Doppler-free signals was only 20 MHz. The uncertainty in the measured $2^3S \rightarrow 3^3P$ interval was two parts in 10^9 (i.e., 1.5 MHz), an improvement by a factor of 60 on earlier measurements. A comparison of the experimental results with recent calculations of the non-QED terms¹¹ gives a value for the one-electron Lamb shift of -346.5 (2.8) MHz, where the uncertainty in the quoted magnitude is in parentheses. The theoretical value is -346.3 (13.9) MHz. Finally, the frequencies of the $2^3S_1 \rightarrow 3^3P_1$, 3^3P_2 transitions were determined to be 25708.60959 (5) cm^{-1} and 25708.58763 (5) cm^{-1} , respectively.

The H^- ion is another two-electron system, of some importance in astrophysics. It is the simplest quantum mechanical three-body species. Approximate quantum mechanical techniques give a wave function and energy eigenvalue which are exact, for all practical purposes. Experimentally, the optical absorption spectrum of H^- is continuous, a property of importance in understanding the opacity of the sun. H^- does not emit radiation in characteristic emission lines. Instead the system sheds its excess (absorbed) energy by ejecting one of the electrons. The radiant energy associated with the ejected electron consists of photons having a continuous energy distribution. Recent measurements with high-intensity pulsed lasers, counterpropagating through beams of 800-MeV H^- ions, have produced spectacular “doubly excited states” of the He^- ion. Such ions are traveling at 84 percent of the velocity of light and, in this situation, the visible laboratory photons are shifted into the vacuum ultraviolet region. At certain energies the H^- ion is briefly excited into a system in which both electrons are excited prior to one of the electrons being ejected. Families of new resonances up to the energy level $N = 8$ have been observed.¹² These resonances are observed as windows in the continuous absorption spectrum, at energies given by a remarkably simple relation reminiscent of the Bohr equation for the Balmer series in hydrogen.¹³ The resonance line with the lowest energy in each family corresponds to both electrons at comparable distances from the proton.

These experiments on H^- are but one facet of the increasingly sophisticated measurements designed to probe the interaction between radiation and matter driven by experiments in laser technology. “Quantum jump” experiments involving single ions in an electromagnetic trap have become almost commonplace. Chaos has also become a rapidly growing subfield of atomic spectroscopy.¹⁴

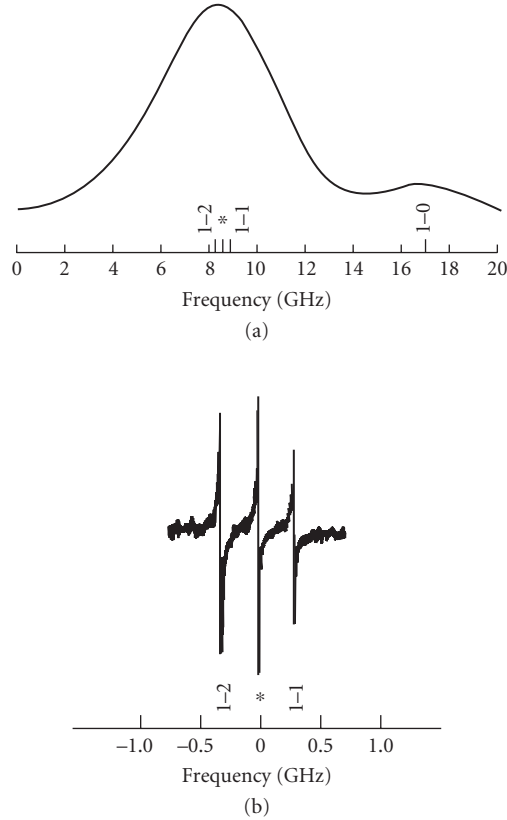


FIGURE 1 (a) Inhomogeneously broadened line profile of the $2^3S \rightarrow 2^3P$ absorption in ^4He , including the weak “holes” due to saturated absorption and the position of the $2^3S \rightarrow 2^3P_2$, $^3P_1$, and $^3P_0$ components. (b) Doppler-free spectra showing the $2^3S \rightarrow 2^3P_2$, $^3P_1$ transitions, and the associated crossover resonance. (After Adams, Riis, and Ferguson.¹⁰)

The particular conditions under which chaos may be observed in atomic physics include hydrogenic atoms in strong homogeneous magnetic fields such that the cyclotron radius of the electron approaches the dimensions of the atomic orbitals. A more easily realizable situation using magnetic field strengths of only a few tesla uses highly excited orbitals close to the ionization threshold. Iu et al.¹⁵ have reported the absorption spectrum of transitions from the $3s$ state of Li to bound and continuum states near the ionization limit in a magnetic field of approximately six tesla. There is a remarkable coincidence between calculations involving thousands of energy levels and experiments involving high-resolution laser spectroscopy.

Atomic processes play an important role in the energy balance of plasmas, whether they be created in the laboratory or in star systems. The analysis of atomic emission lines gives much information on the physical conditions operating in a plasma. In laser-produced plasmas, the densities of charged ions may be in the range $10^{20} - 10^{25} \text{ ions cm}^{-3}$, depending on the pulse duration of the laser. The spectra of many-electron ions are complex and may have the appearance of an unresolved transition

array between states belonging to specific initial and final configurations. Theoretical techniques have been developed to determine the average ionization state of the plasma from the observed optical spectrum. In many cases, the spectra are derived from ionic charge states in the nickel-like configuration containing 28 bound electrons. In normal nickel, the outershell configuration is $(3d^8)(4s^2)$, the 4s levels having filled before 3d because the electron-electron potentials are stronger than electron-nuclear potentials. However, in highly ionized systems, the additional electron-nuclear potential is sufficient to shift the configuration from $(3d^8)(4s^2)$ to the closed shell configuration $(3d^{10})$. The resulting spectrum is then much simpler than for atomic nickel. The Ni-like configuration has particular relevance in experiments to make x-ray lasers. For example, an analog series of collisionally pumped lasers using Ni-like ions has been developed, including a Ta^{45+} laser operating at 4.48 nm and a W^{46+} laser operating at 4.32 nm.¹⁶

Molecular Spectroscopy

The basic principles of gas-phase molecular spectroscopy were also discussed in “Optical Spectroscopy and Spectroscopic Lineshapes,” Vol. I, Chap. 10 of this *Handbook*. The spectra of even the simplest molecules are complicated by the effects of vibrations and of rotations about an axis. This complexity is illustrated elsewhere in this *Handbook* in Fig. 8 of this chapter, which depicts a photographically recorded spectrum of the $2\Pi \rightarrow 3\Sigma$ bands of the diatomic molecule NO, which was interpreted in terms of progressions and line sequences associated with the *P*-, *Q*-, and *R*-branches. The advent of Fourier transform spectroscopy led to great improvements in resolution and greater efficiency in revealing all the fine details that characterize molecular spectra. Figure 2a is a Fourier-transform infrared spectrum of nitrous oxide, N_2O , which shows the band center at 2462 cm^{-1} flanked by the *R*-branch and a portion of the *P*-branch; the density of lines in the *P*- and *R*-branch is evident. On an expanded scale, in Fig. 2b, there is a considerable simplification of the rotational-vibrational structure at the high-energy portion of the *P*-branch. The weaker lines are the so-called “hot bands.”

More precise determinations of the transition frequencies in molecular physics are measured using Lamb dip spectroscopy. The spectrum shown in Fig. 3a is a portion of the laser Stark spectrum of methyl fluoride measured using electric fields in the range 20 to 25 KV cm^{-1} with the $9\text{-}\mu\text{m } P(18)$ line of the CO_2 laser, which is close to the ν_3 band origin of CH_3F .¹⁷ The spectrum in Fig. 3a consists of a set of $\Delta M_J = \pm 1$ transitions, brought into resonance at different values of the static electric field. Results from the alternative high-resolution technique using a supersonic molecular beam and bolometric detector are shown in Fig. 3b; this spectrum was obtained using CH_3F in He mixture expanded through a $35\text{-}\mu\text{m}$ nozzle. The different M_J components of the $Q(1, 0)$, $Q(2, 1)$, and $Q(3, 3)$ components of the *Q*-branch are shown to have very different intensities relative to those in Fig. 3a on account of the lower measurement temperature.

There has been considerable interest in the interaction of intense laser beams with molecules. For example, when a diatomic molecule such as N_2 or CO is excited by an intense (10^{15} W cm^{-2}), ultrashort (0.6 ps) laser pulse, it multiply ionizes and then fragments as a consequence of Coulomb repulsion. The charge and kinetic energy of the resultant ions can be determined by time-of-flight (TOF) mass spectrometry. In this technique, the daughter ions of the “Coulomb explosion” drift to a cathode tube at different times, depending on their weight. In traditional methods, the TOF spectrum is usually averaged over many laser pulses to improve the signal-to-noise ratio. Such simple averaging procedures remove the possibility of correlations between particular charged fragments. This problem was overcome by the *covariance mapping* technique developed by Frasninski and Codling.¹⁸ Experimentally, a linearly polarized laser pulse with *E*-vector pointing toward the detector is used to excite the molecules, which line up with their internuclear axis parallel to the *E*-field. Under Coulomb explosion, one fragment heads toward the detector and the other away from the detector. The application of a dc electric field directs the “backward” fragment ion to the detector, arriving at some short time after the forward fragment. This temporal separation of two fragments arriving at the detector permits the correlation between molecular fragments to be retained. In essence, the TOF spectrum, which plots molecular weight versus counts, is arranged both horizontally (forward ions) and vertically

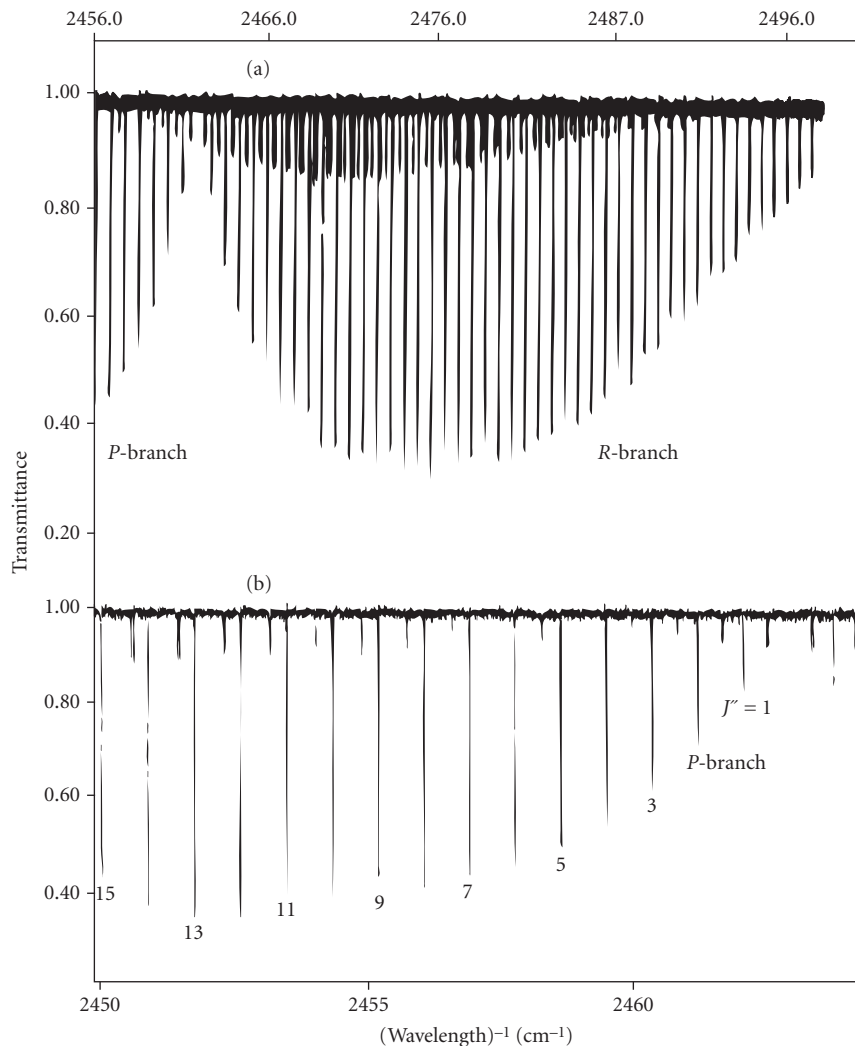


FIGURE 2 (a) *P*- and *R*-branches for the nitrous oxide molecule measured using Fourier-transform infrared spectroscopy. The gas pressure was 0.5 torr and system resolution 0.006 cm^{-1} , (b) on an expanded scale, the high-energy portion of the *P*-branch up to $J'' = 15$.

(backward ions) on a two-dimensional graph. A coordinate point on a preliminary graph consists of two ions along with their counts during a single pulse. Coordinates from 10^4 pulses or so are then assembled in a final map. Each feature on the final map relates to a specific fragmentation channel, i.e., the pair of fragments and their parent molecule. The strength of the method is that it gives the probability for the creation and fragmentation of the particular parent ion. Covariance mapping experiments on N_2 show that 610- and 305-nm pulses result in fragmentation processes that are predominantly charge-symmetric. In other words, the Coulomb explosion proceeds via the production of ions with the same charge.

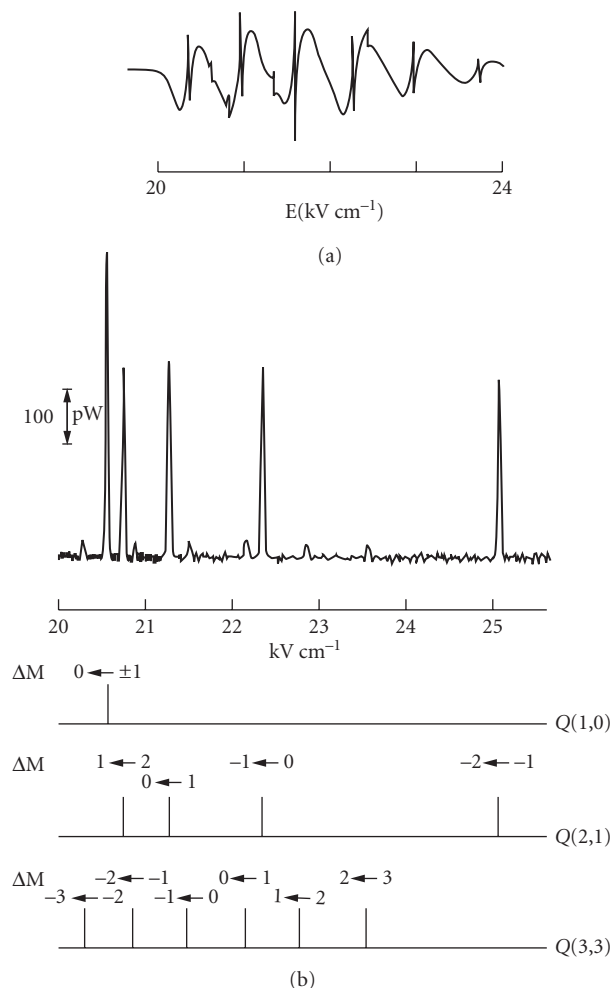


FIGURE 3 (a) Laser Stark absorption spectrum of methyl fluoride measured at 300 K using Lamb dip spectroscopy with a gas pressure of 5 mTorr and (b) the improved resolution obtained using molecular-beam techniques with low-temperature bolometric detection and CH_3F in He expanded through a 35- μm nozzle. (After Douketic and Gough.¹⁷)

Optical Spectroscopy of Solids

One of the more fascinating aspects of the spectroscopy of electronic centers in condensed matter is the variety of lineshapes displayed by the many different systems. Those discussed in Vol. I, Chap. 10 include Nd^{3+} in YAG (Fig. 6), O_2^- in KBr (Fig. 11), Cr^{3+} in YAG (Fig. 12), and F centers in KBr (Fig. 13). The very sharp Nd^{3+} lines (Fig. 6) are zero-phonon lines, inhomogeneously broadened by strain. The abundance of sharp lines is characteristic of the spectra of trivalent rare-earth ions in ionic crystals. Typical low-temperature linewidths for Nd^{3+} : YAG are 0.1–0.2 cm^{-1} . There is particular interest in the spectroscopy of Nd^{3+} because of the efficient laser transitions from the $^4\text{F}_{3,2}$ level into the $^4\text{I}_j$ -manifold. The low-temperature luminescence transitions between $^4\text{F}_{3,2} \rightarrow ^4\text{I}_{15/2}$, $^4\text{I}_{13/2}$, $^4\text{I}_{11/2}$, and $^4\text{I}_{9/2}$ levels are

shown in Fig. 6: all are split by the effects of the crystalline electric field. Given the relative sharpness of these lines, it is evident that the Slater integrals $F^{(k)}$, spin-orbit coupling parameters ζ , and crystal field parameters, B_r^k , may be measured with considerable accuracy. The measured values of the $F^{(k)}$ and ζ vary little from one crystal to another.¹⁹ However, the crystal field parameters, B_r^k , depend strongly on the rare-earth ion-ligand-ion separation. Most of the $4f^n$ ions have transitions which are the basis of solid-state lasers. Others such as Eu^{3+} and Tb^{3+} are important red-emitting and green-emitting phosphor ions, respectively.²⁰

Transition-metal ion spectra are quite different from those of the rare-earth ions. In both cases, the energy-level structure may be determined by solving the Hamiltonian

$$H = H_o + H' + H_{so} + H_c \quad (1)$$

in which H_o is a sum of one-electron Hamiltonians including the central field of each ion, H' is the interaction between electrons in the partially filled $3d^n$ or $4f^n$ orbitals, H_{so} is the spin-orbit interaction, and H_c is the interaction of the outer shell electrons with the crystal field. For rare-earth ions $H', H_{so} \gg H_c$ and the observed spectra very much reflect the free-ion electronic structure with small crystal field perturbations. The spectroscopy of the transition-metal ions is determined by the relative magnitudes of $H' \approx H_c \gg H_{so}$.^{19,21} The simplest of the transition-metal ions is Ti^{3+} : in this $3d^1$ configuration a single $3d$ electron resides outside the closed shells. In this situation, $H' = 0$ and only the effect of H_c needs be considered (Fig. 4a). The Ti^{3+} ion tends to form octahedral complexes, where the $3d$ configuration is split into ${}^2E$ and ${}^2T_2$ states with energy separation $10Dq$. In cation sites with weak, trigonally symmetric distortions, as in Al_2O_3 and $\text{Y}_3\text{Al}_5\text{O}_{12}$, the lowest-lying state, ${}^2T_2$, splits into ${}^2A_1$ and ${}^2E$ states (using the C_{3v} group symmetry labels). In oxides, the octahedral splitting is of order $10Dq \approx 20,000 \text{ cm}^{-1}$ and the trigonal field splitting $\nu \approx 700 - 1000^{-1}$. Further splittings of the levels

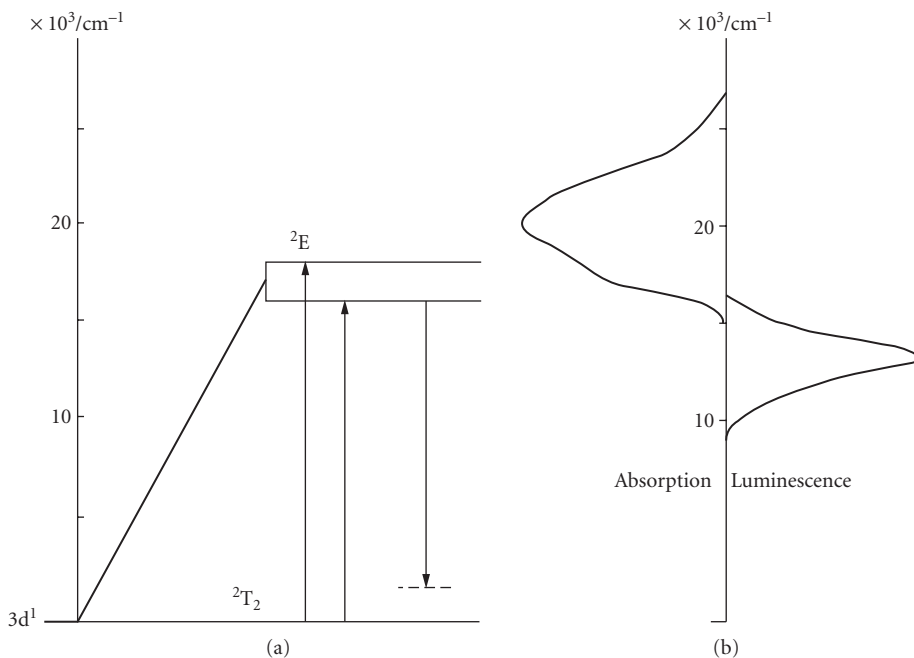


FIGURE 4 Absorption and emission spectra of Ti^{3+} ions [$3d^1$ configuration] in Al_2O_3 measured at 300 K.

occur because of spin-orbit coupling and the Jahn-Teller effect. The excited ${}^2\text{E}$ state splits into $2\bar{\text{A}}$ and $\bar{\text{E}}$ (from ${}^2\text{A}_1$), and $\bar{\text{E}}$ and $2\bar{\text{A}}$ (from ${}^2\text{E}$). The excited state splitting by a *static* Jahn-Teller effect is large, ~ 2000 to 2500 cm^{-1} , and may be measured from the optical absorption spectrum. In contrast, ground-state splittings are quite small: a *dynamic* Jahn-Teller effect has been shown to strongly quench the spin-orbit coupling ζ and trigonal field splitting ν parameters.²² In $\text{Ti}^{3+}:\text{Al}_2\text{O}_3$ the optical absorption transition, ${}^2\text{T}_2 \rightarrow {}^2\text{E}$, measured at 300 K, Fig. 4b, consists of two broad overlapping bands separated by the Jahn-Teller splitting, the composite band having a peak at approximately $20,000\text{ cm}^{-1}$. Luminescence occurs only from the lower-lying excited state $2\bar{\text{A}}$, the emission band peak occurring at approximately $14,000\text{ cm}^{-1}$. As Fig. 4 shows, both absorption and emission bands are broad because of strong electron-phonon coupling. At low temperatures the spectra are characterized by weak zero-phonon lines, one in absorption due to transitions from the $2\bar{\text{A}}$ ground state and three in emission corresponding to transitions in the $\bar{\text{E}}$, and $2\bar{\text{A}}$ levels of the electronic ground state.^{22,23} These transitions are strongly polarized.

For ions with $3d^n$ configuration it is usual to neglect H_c and H_{so} in Eq. (1), taking into account only the central ion terms and the Coulomb interaction between the $3d$ electrons. The resulting energies of the free-ion LS terms are expressed in terms of the Racah parameters A , B , and C . Because energy differences between states are measured in spectroscopy, only B and C are needed to categorize the free-ion levels. For pure d-functions, $C/B = 4.0$. The crystal field term H_c and H_{so} also are treated as perturbations. In many crystals, the transition-metal ions occupy octahedral or near-octahedral cation sites. The splittings of each free-ion level by an octahedral crystal field depend in a complex manner on B , C , and the crystal field strength Dq given by

$$Dq = \left(\frac{Ze^2}{24\pi\epsilon_0} \right) \frac{\langle r^4 \rangle 3d}{a^5} \quad (2)$$

The parameters D and q always occur as a product. The energy levels of the $3d^n$ transition-metal ions are usually represented on Tanabe-Sugano diagrams, which plot the energies $E(\Gamma)$ of the electronic states as a function of the octahedral crystal field.^{19,21} The crystal field levels are classified by irreducible representations Γ of the octahedral group, O_h . The Tanabe-Sugano diagram for the $3d^3$ configuration, shown in Fig. 5a, was constructed using a C/B ratio = 4.8: the vertical broken line drawn at $Dq/B = 2.8$ is appropriate for Cr^{3+} ions in ruby. If a particular value of C/B is assumed, only two variables, B and Dq , need to be considered: in the diagram $E(\Gamma)/B$ is plotted as a function of Dq/B . The case of ruby, where the ${}^2\text{E}$ level is below ${}^4\text{T}_2$, is referred to as the *strong field* case. Other materials where this situation exists include YAlO_3 , $\text{Y}_3\text{Al}_5\text{O}_{12}$ (YAG), and MgO . In many fluorides, Cr^{3+} ions occupy *weak field* sites, where $E({}^4\text{T}_2) < E({}^2\text{E})$ and Dq/B is less than 2.2. When the value of Dq/B is close to 2.3, the *intermediate* crystal field, the ${}^4\text{T}_2$ and ${}^2\text{E}$ states almost degenerate. The value of Dq/B at the level crossing between ${}^4\text{T}_2$ and ${}^2\text{E}$ depends slightly on the value of C .

The Tanabe-Sugano diagram represents the static lattice. In practice, electron-phonon coupling must be taken into account: the relative strengths of coupling to the states involved in transitions and the consequences may be inferred from Fig. 5a. Essentially ionic vibrations modulate the crystal field experienced by the central ion at the vibrational frequency. Large differences in slope of the E versus Dq graphs indicate large differences in coupling strengths and hence large homogeneous bandwidths. Hence, absorption and luminescence transitions from the ${}^4\text{A}_2$ ground state to the ${}^4\text{T}_2$ and ${}^4\text{T}_1$ states will be broadband due to the large differences in coupling of the electronic energy to the vibrational energy. For the ${}^4\text{A}_2 \rightarrow {}^2\text{E}$, ${}^2\text{T}_1$ transition, the homogeneous linewidth is hardly affected by lattice vibrations, and sharp line spectra are observed.

The Cr^{3+} ion occupies a central position in the folklore of transition-metal ion spectroscopy, having been studied by spectroscopists for over 150 years. An extensive survey of Cr^{3+} luminescence in many compounds was published as early as 1932.²⁴ The Cr^{3+} ions have the outer shell configuration, $3d^3$, and their absorption and luminescence spectra may be interpreted using Fig. 5. First, the effect of the octahedral crystal field is to remove the degeneracies of the free-ion states ${}^4\text{F}$ and ${}^2\text{G}$. The ground term of the free ion, ${}^4\text{F}_{3/2}$, is split by the crystal field into a ground-state orbital singlet, ${}^4\text{A}_{2g}$, and two orbital triplets ${}^4\text{T}_{2g}$ and ${}^4\text{T}_{1g}$, in order of increasing energy. Using the energy of the ${}^4\text{A}_2$ ground state as the zero, for all values of Dq/B , the energies of the ${}^4\text{T}_{2g}$ and ${}^4\text{T}_{1g}$ states are seen to vary strongly as a function of the octahedral crystal field. In a similar vein, the ${}^2\text{G}$ free-ion state splits into ${}^2\text{E}$, ${}^2\text{T}_1$,

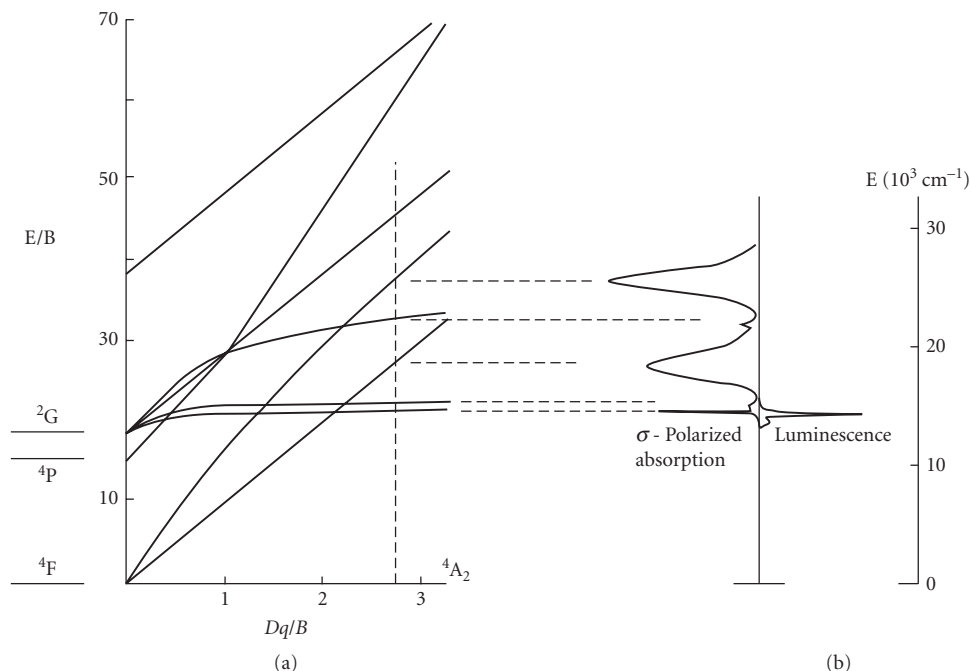


FIGURE 5 Tanabe-Sugano diagram for Cr³⁺ ions with $C/B = 4.8$, appropriate for ruby for which $Dq/B = 2.8$. On the right of the figure are shown the optical absorption and photoluminescence spectrum of ruby measured at 300 K.

${}^2T_2$, and ${}^2A_1$ states, the two lowest of which, ${}^2E$ and ${}^2T_1$, vary very little with Dq . The energies $E({}^2T_2)$ and $E({}^2A_1)$ are also only weakly dependent on Dq/B . The free-ion term, ${}^4P$, which transforms as the irreducible representation ${}^4T_1$ of the octahedral group is derived from (e^2t_2) configuration: this term is not split by the octahedral field although its energy is a rapidly increasing function of Dq/B . Low-symmetry distortions lead to strongly polarized absorption and emission spectra.²⁵

The σ -polarized optical absorption and luminescence spectra of ruby are shown in Fig. 5b. The expected energy levels predicted from the Tanabe-Sugano diagram are seen to coincide with appropriate features in the absorption spectrum. The most intense features are the vibronically broadened ${}^4A_2 \rightarrow {}^4T_1$, ${}^4T_2$ transitions. These transitions are broad and characterized by large values of the Huang-Rhys factor ($S \approx 6-7$). These absorptions occur in the blue and yellow-green regions, thereby accounting for the deep red color of ruby. Many other Cr³⁺-doped hosts have these bands in the blue and orange-red regions, by virtue of smaller values of Dq/B ; the colors of such materials (e.g., MgO, Gd₃Sc₂Ga₃O₁₂, LiSrAlF₆, etc.) are different shades of green. The absorption transitions, ${}^4A_2 \rightarrow {}^2E$, ${}^2T_1$, ${}^2T_2$ levels are spin-forbidden and weakly coupled to the phonon spectrum ($S < 0.5$). The spectra from these transitions are dominated by sharp zero-phonon lines. However, the low-temperature photoluminescence spectrum of ruby is in marked contrast to the optical absorption spectrum since only the sharp zero-phonon line (R -line) due to the ${}^2E \rightarrow {}^4A_2$ transition being observed. Given the small energy separations between adjacent states of Cr³⁺, the higher excited levels decay nonradiatively to the lowest level, ${}^2E$, from which photoluminescence occurs across a bandgap of approximately 15,000 cm⁻¹. Accurate values of the parameters Dq , B , and C may be determined from these absorption data. First, the peak energy of the ${}^4A_2 \rightarrow {}^4T_2$ absorption band is equal to $10Dq$. The energy shift between the ${}^4A_2 \rightarrow {}^4T_2$, ${}^4T_1$ bands is dependent on both Dq and B , and the energy separation between the two broad absorption bands is used to determine B . Finally, the position of the R -line varies with Dq , B , and C : in consequence, once Dq and B are known, the magnitude of C may be determined from the position of the ${}^4A_2 \rightarrow {}^2E$ zero-phonon line.

This discussion of the spectroscopy of the Cr^{3+} ion is easily extended to other multielectron configurations. The starting points are the Tanabe-Sugano diagrams collected in various texts.^{19,21} Analogous series of elements occur in the fifth and sixth periods of the periodic table, respectively, where the $4d^n$ (palladium) and $5d^n$ (platinum) groups are being filled. Compared with electrons in the $3d$ shell, the $4d$ and $5d$ shell electrons are less tightly bound to the parent ion. In consequence, charge transfer transitions, in which an electron is transferred from the cation to the ligand ion (or vice versa), occur quite readily. The charge transfer transitions arise from the movement of electronic charge over a typical interatomic distance, thereby producing a large dipole moment and a concomitant large oscillator strength for the absorption process. For the Fe-group ions ($3d^n$ configuration), such charge transfer effects result in the absorption of ultraviolet photons.

For example, the Fe^{3+} ion in MgO absorbs in a broad structureless band with peak at 220 nm and half-width of order 120 nm (i.e., 0.3 eV). The Cr^{2+} ion also absorbs by charge transfer process in this region. In contrast, the palladium and platinum groups have lower-lying charge transfer states. The resulting intense absorption bands in the visible spectrum may overlap spectra due to low-lying crystal field transitions. Rare-earth ions also give rise to intense charge transfer bands in the ultraviolet region.

Various metal cations have been used as broadband visible region phosphors. For example, transitions between the $4f^n$ and $4f^{(n-1)} 5d$ levels of *divalent* rare-earth ions give rise to intense broad transitions which overlap many of the sharp $4f^n$ transitions of the *trivalent* rare-earth ions. Of particular interest are Sm^{2+} , Dy^{2+} , Eu^{2+} , and Tm^{2+} . In Sm^{2+} ($4f^6$) broadband absorption transitions from the ground state $^7F_1$ to the $4f^5 5d$ level may result in either broadband emission (from the vibrationally relaxed $4f^5 5d$ level) or sharp line emission from $^5D_0$ ($4f^6$) depending upon the host crystal. The $4f^5 5d$ level, being strongly coupled to the lattice, accounts for this variability. There is a similar material-by-material variation in the absorption and emission properties of the Eu^{2+} ($4f^7$ configuration), which has the $^8S_{7/2}$ ground level. The next highest levels are derived from the $4f^6 5d$ state, which is also strongly coupled to the lattice. This state is responsible for the varying emission colors of Eu^{2+} in different crystals, e.g., violet in $\text{Sr}_2\text{P}_2\text{O}_7$, blue in $\text{BaAl}_{12}\text{O}_{19}$, green in SrAl_2O_4 , and yellow in Ba_2SiO_5 .

The heavy metal ions Tl^+ , In^+ , Ga^+ , Sn^{2+} , and Pb^{2+} may be used as visible-region phosphors. These ions all have two electrons in the ground configuration ns^2 and excited configurations $(ns)(np)$. The lowest-lying excited states, in the limit of Russell Saunders coupling, are then $^1S_0(ns^2)$, $^3P_{0,1,2}$, and $^1P_1$ from $(ns)(np)$. The spectroscopy of Tl^+ has been much studied especially in the alkali halides. Obviously $^1S_0 \rightarrow ^1P_1$ is the strongest absorption transition, occurring in the ultraviolet region. This is labeled as the C-band in Fig. 6. Next in order of observable intensity is the A-band, which is a spin-forbidden absorption transition $^1S_0 \rightarrow ^3P_1$, in which the relatively large oscillator strength is borrowed from the $^1P_1$ state by virtue of the strong spin-orbit interaction in these heavy metal ions. The B and D bands, respectively, are due to absorption transitions from $^1S_0$ to the $^3P_2$ and $^3P_0$ states induced by vibronic mixing.²⁶ A phenomenological theory^{26,27} quantitatively accounts for both absorption spectra and the triplet state emission spectra.^{28,29}

The examples discussed so far have all concerned the spectra of ions localized in levels associated with the central fields of the nucleus and closed-shells of electrons. There are other situations which warrant serious attention. These include electron-excess centers in which the positive potential of an anion vacancy in an ionic crystal will trap one or more electrons. The simplest theory treats such a *color center* as a particle in a finite potential well.^{19,27} The simplest such center is the F-center in the alkali halides, which consist of one electron trapped in an anion vacancy. As we have already seen (e.g., Fig. 13 in "Optical Spectroscopy and Spectroscopic Lineshapes", Vol. I, Chap. 10 of this *Handbook*), such centers give rise to broadbands in both absorption and emission, covering much of the visible and near-infrared regions for alkali halides. F-aggregate centers, consisting of multiple vacancies arranged in specific crystallographic relationships with respect to one another, have also been much studied. They may be positive, neutral, or negative in charge relative to the lattice depending upon the number of electrons trapped by the vacancy aggregate.³⁰

Multiquantum wells (MQWs) and strained-layer superlattices (SLSs) in semiconductors are yet another type of *finite-well* potential. In such structures, alternate layers of two different semiconductors are grown on top of each other so that the bandgap varies in one dimension with the periodicity

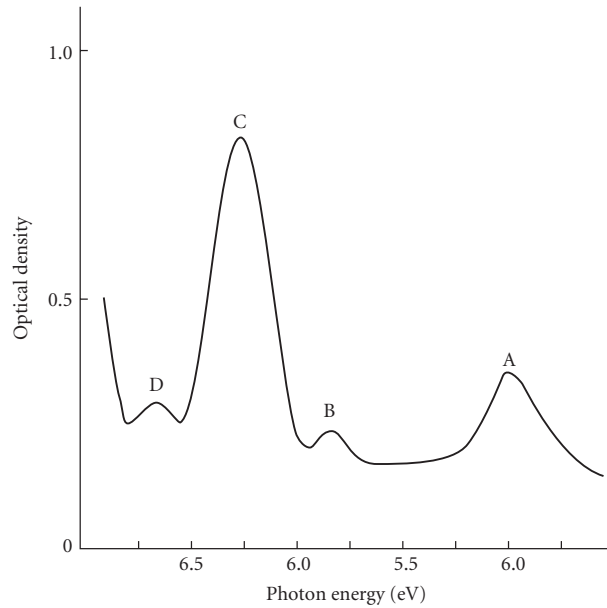


FIGURE 6 Ultraviolet absorption spectrum of Tl^+ ions in KCl measured at 77 K. (After Delbecq *et al.*²⁶)

of the epitaxial layers. A modified Kronig-Penney model is often used to determine the energy eigenvalues of electrons and holes in the conduction and valence bands, respectively, of the narrower gap material. Allowed optical transitions between valence band and conduction band are then subject to the selection rule $\Delta n = 0$, where $n = 0, 1, 2$, etc. The example given in Fig. 7 is for SLSs in the II-VI family of semiconductors ZnS/ZnSe.³¹ The samples were grown by metallo-organic vapor phase epitaxy³²

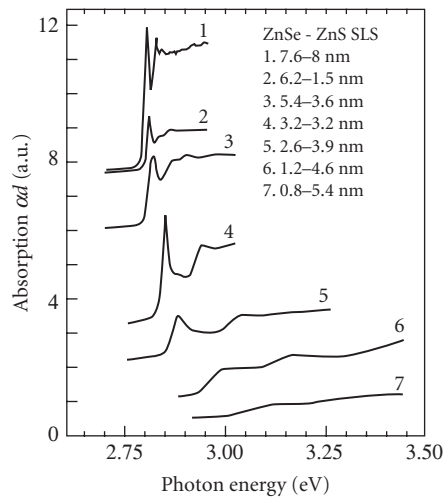


FIGURE 7 Optical absorption spectra of SLSs of ZnS/ZnSe measured at 14 K. (After Fang *et al.*³¹)

with a superlattice periodicity of 6 to 8 nm while varying the thickness of the narrow gap material (ZnSe) between 0.8 and 7.6 nm. The splitting between the two sharp features occurs because the valence band states are split into “light holes” (lh) and “heavy holes” (hh) by spin-orbit interaction. The absorption transitions then correspond to transitions from the $n = 1$ lh- and hh-levels in the valence band to the $n = 1$ electron states in the conduction band. Higher-energy absorption transitions are also observed. After absorption, electrons rapidly relax down to the $n = 1$ level from which emission takes place down to the $n = 1$, lh-level in the valence band, giving rise to a single emission line at low temperature.

2.4 THE HOMOGENEOUS LINESHAPE OF SPECTRA

Atomic Spectra

The homogeneous widths of atomic spectra are determined by the uncertainty principle, and hence by the radiative decaytime, τ_R (as discussed in Vol. I, Chap. 10, “Optical Spectroscopy and Spectroscopic Lineshapes”). Indeed, the so-called natural or homogeneous width of $\Delta\omega$, is given by the Einstein coefficient for spontaneous emission, $A_{ba} = (\tau_R)^{-1}$. The homogeneously broadened line has a Lorentzian lineshape with FWHM given by $(\tau_R)^{-1}$. In gas-phase spectroscopy, atomic spectra are also broadened by the Doppler effect: random motion of atoms broadens the lines in-homogeneously leading to a gaussian-shaped line with FWHM proportional to $(T/M)^{-1/2}$, where T is the absolute temperature and M the atomic mass. Saturated laser absorption or optical hole-burning techniques are among the methods which recover the true homogeneous width of an optical transition. Experimental aspects of these types of measurement were discussed in this *Handbook* in Vol. I, Chap. 31, “Optical Spectroscopy,” and examples of Doppler-free spectra (Figs. 1 and 3, Vol. I, Chap. 10) were discussed in terms of the fundamental tests of the quantum and relativistic structure of the energy levels of atomic hydrogen. Similar measurements were also discussed for the case of He (Fig. 1) and in molecular spectroscopy (Fig. 3). In such examples, the observed lineshape is very close to a true Lorentzian, typical of a lifetime-broadened optical transition.

Zero-Phonon Lines in Solids

Optical hole burning (OHB) reduces the effects of inhomogeneous broadening in solid-state spectra. For rare-earth ions, the homogeneous width amounts to some 0.1–1.0 MHz, the inhomogeneous widths being determined mainly by strain in the crystal. Similarly, improved resolution is afforded by fluorescence line narrowing (FLN) in the R -line of Cr^{3+} (Vol. I, Chap. 10, Fig. 12). However, although the half-width measured using OHB is the true homogeneous width, the observed FLN half-width, at least in resonant FLN, is a convolution of the laser width and twice the homogeneous width of the transition.³³ In solid-state spectroscopy, the underlying philosophy of OHB and FLN experiments may be somewhat different from that in atomic and molecular physics. In the latter cases, there is an intention to relate theory to experiment at a rather sophisticated level. In solids, such high-resolution techniques are used to probe a range of other dynamic processes than the natural decay rate. For example, hole-burning may be induced by photochemical processes as well as by intrinsic lifetime processes.^{34,35} Such photochemical hole-burning processes have potential in optical information storage systems. OHB and FLN may also be used to study distortions in the neighborhood of defects. Figure 8 is an example of Stark spectroscopy and OHB on a zero-phonon line at 607 nm in irradiated NaF .³⁴ This line had been attributed to an aggregate of four F -centers in nearest-neighbor anion sites in the rocksalt-structured lattice on the basis of the polarized absorption/emission measurements. The homogeneous width in zero electric field was only 21 MHz in comparison with the inhomogeneous width of 3 GHz. Interpretation of these results is inconsistent with the four-defect model.

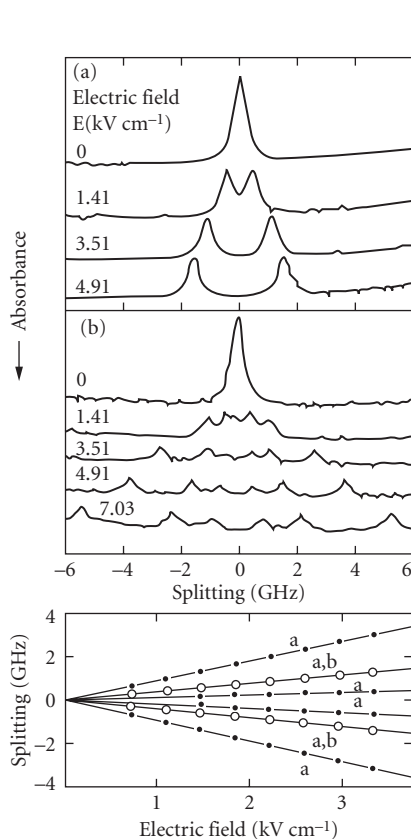


FIGURE 8 Effects of an applied electric field in a hole burned in the 607-nm zero-phonon line observed in irradiated NaF. (After Macfarlane et al.²²)

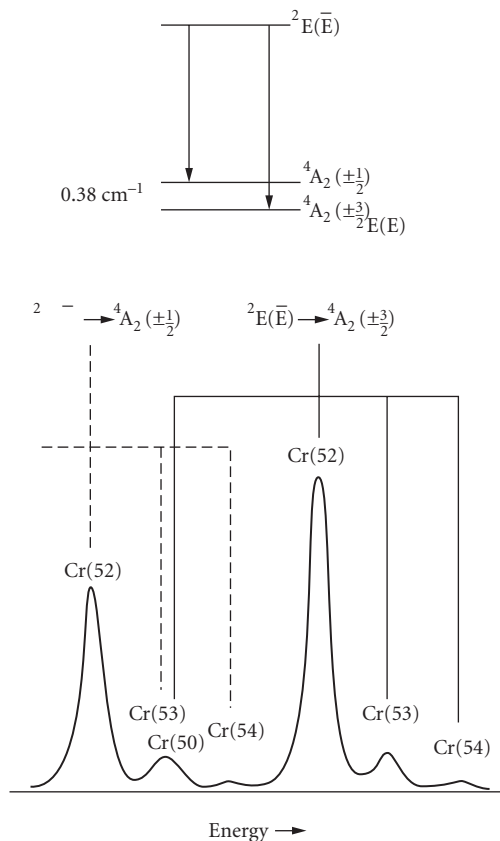


FIGURE 9 Fine structure splitting in the ${}^4A_2$ ground state and the isotope shifts of Cr^{3+} in the R_1 -line of ruby measured using FLN spectroscopy. (After Jessop and Szabo.³⁶)

The FLN technique may also be used to measure the effects of phonon-induced relaxation processes and isotope shifts. Isotope and thermal shifts have been reported for $Cr^{3+}:Al_2O_3$ ³⁶ and $Nd^{3+}:LaCl_3$.³⁷ The example given in Fig. 9 shows both the splitting in the ground ${}^4A_2$ state of Cr^{3+} in ruby and the shift between lines due to the Cr(50), Cr(52), Cr(53), and Cr(54) isotopes. The measured differential isotope shift of 0.12 cm^{-1} is very close to the theoretical estimate.¹⁹ Superhyperfine effects by the 100 percent abundant Al isotope with $I = 5/2$ also contribute to the homogeneous width of the FLN spectrum of Cr^{3+} in Al_2O_3 (Fig. 12 in Vol. I, Chap. 10).³⁶ Furthermore, in antiferromagnetic oxides such as $GdAlO_3$, $Gd_3Ga_5O_{12}$, and $Gd_3Sc_2Ga_3O_{12}$, spin-spin coupling between the Cr^{3+} ions ($S = 3/2$) and nearest-neighbor Gd^{3+} ions ($S = 7/2$) contributes as much to the zero-phonon R -linewidth as inhomogeneous broadening by strain.³⁸

Configurational Relaxation in Solids

In the case of the broadband ${}^4T_2 \rightarrow {}^4A_2$ transition of Cr^{3+} in YAG (Fig. 12 in Vol. I, Chap. 10) and MgO (Fig. 6), the application of OHB and FLN techniques produce no such narrowing because the

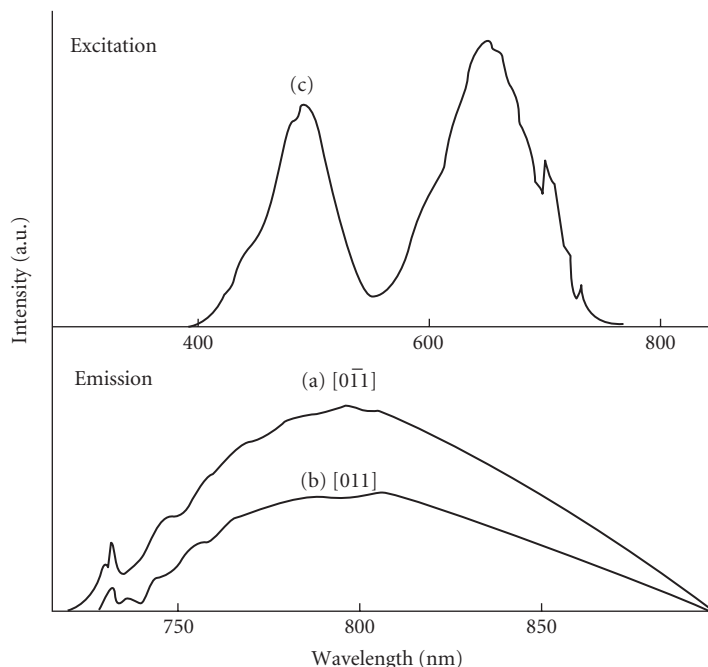


FIGURE 10 Polarized emission of the ${}^4T_2 \rightarrow {}^4A_2$ band from Cr^{3+} ions in orthorhombic sites in MgO. Shown also, (c) is the excitation spectrum appropriate to (a). (After Yamaga *et al.*⁴⁸)

vibronic sideband is the homogeneously broadened shape determined by the phonon lifetime rather than the radiative lifetime. It is noteworthy that the vibronic sideband emission of Cr^{3+} ions in orthorhombic sites in MgO, Fig. 10, shows very little structure. In this case, the Huang-Rhys factor $S \approx 6$, i.e., the strong coupling case, where the multiphonon sidebands tend to lose their separate identities to give a smooth bands shape on the lower-energy side of the peak. By way of contrast, the emission sideband of the *R*-line transition of Cr^{3+} ions in octahedral sites in MgO is very similar in shape to the known density of one-phonon vibrational modes of MgO³⁹ (Fig. 11), although there is a difference in the precise positions of the peaks, because the Cr^{3+} ion modifies the lattice vibrations in its neighborhood relative to those of the perfect crystal. Furthermore, there is little evidence in Fig. 11 of higher-order sidebands which justifies treating the MgO *R*-line process in the weak coupling limit. The absence of such sidebands suggests that $S < 1$, as the discussion in “Optical Spectroscopy and Spectroscopic Lineshapes” (Vol. I, Chap. 10 of this *Handbook*) showed. That the relative intensities of the zero-phonon line and broadband, which should be about e^{-S} , is in the ratio 1:4 shows that the sideband is induced by odd parity phonons. In this case it is partially electric-dipole in character, whereas the zero-phonon line is magnetic-dipole in character.¹⁹

There has been much research on bands shapes of Cr^{3+} -doped spectra in many solids. This is also the situation for *F*-centers and related defects in the alkali halides. Here, conventional optical spectroscopy has sometimes been supplemented by laser Raman and sub-picosecond relaxation spectroscopies to give deep insights into the dynamics of the optical pumping cycle. The *F*-center in the alkali halides is a halide vacancy that traps an electron. The states of such a center are reminiscent of a particle in a finite potential well¹⁹ and strong electron-phonon coupling. Huang-Rhys factors in the range $S = 15 - 40$ lead to broad, structureless absorption/luminescence bands with large Stokes

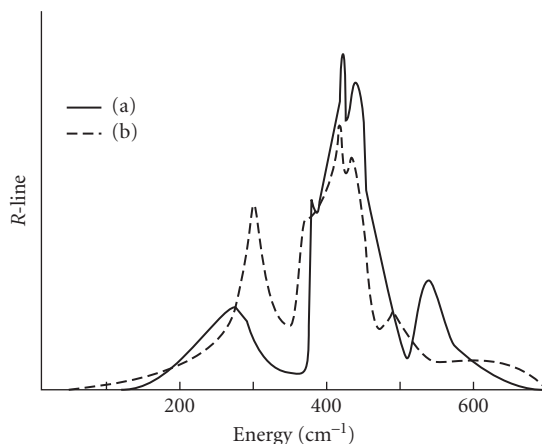


FIGURE 11 A comparison of (a) the vibrational sideband accompanying the ${}^2E \rightarrow {}^4A_2$ R-line of Cr^{3+} : MgO with (b) the density of phonon modes in MgO as measured by Peckham et al.,³⁹ using neutron scattering. (After Henderson and Imbusch.¹⁹)

shifts (see Fig. 13 in Vol. I, Chap. 10). Raman-scattering measurements on F -centers in NaCl and KCl (Fig. 23 in Vol. I, Chap. 31), showed that the first-order scattering is predominantly due to defect-induced local modes.⁴⁰

The F_A -center is a simple variant on the F -center in which one of the six nearest cation neighbors of the F -center is replaced by an alkali impurity.⁴¹ In KCl the K^+ may be replaced by Na^+ or Li^+ . For the case of the Na^+ substituent, the $F_A(\text{Na})$ center has tetragonal symmetry about a $\langle 100 \rangle$ crystal axis, whereas in the case of Li^+ an off-axis relaxation in the excited state leads to interesting polarized absorption/emission characteristics.^{19,41} The most dramatic effect is the enormous Stokes shift between absorption and emission bands, of order $13,000 \text{ cm}^{-1}$, which has been used to advantage in color center lasers.^{42,43} For $F_A(\text{Li})$ centers, configurational relaxation has been probed using picosecond relaxation and Raman-scattering measurements. Mollenauer et al.⁴⁴ used the experimental system shown in Fig. 10 in Vol. I, Chap. 31 to carry out measurements of the configurational relaxation time of $F_A(\text{Li})$ centers in KCl. During deexcitation many phonons are excited in the localized modes coupled to the electronic states which must be dissipated into the continuum of lattice modes. Measurement of the relaxation time constitutes a probe of possible phonon damping. A mode-locked dye laser producing pulses of 0.7-ps duration at 612 nm was used both to pump the center in the F_{A2} -absorption band and to provide the timing beam. Such pumping leads to optical gain in the luminescence band and prepares the centers in their relaxed state. The probe beam, collinear with the pump beam, is generated by a CW $F_A(\text{Li})$ -center laser operating at 2.62 μm . The probe beam and gated pulses from the dye laser are mixed in a non-linear optical crystal (lithium iodate). A filter allows only the sum frequency at 496 nm to be detected. The photomultiplier tube then measures the rise in intensity of the probe beam which signals the appearance of gain where the $F_A(\text{Li})$ -centers have reached the relaxed excited state. The pump beam is chopped at low frequency to permit phase-sensitive detection. The temporal evolution of $F_A(\text{Li})$ -center gains (Fig. 12a and b) was measured by varying the time delay between pump and gating pulses. In this figure, the solid line is the instantaneous response of the system, whereas in b the dashed line is the instantaneous response convolved with a 1.0-ps rise time.

Measurements of the temperature dependence of the relaxation times of $F_A(\text{Li})$ -centers in potassium chloride (Fig. 12c) show that the process is very fast, typically of order 10 ps at 4 K. Furthermore, configurational relaxation is a multiphonon process which involves mainly the creation of some 20 low-energy phonons of energy $E_p/hc \approx 47 \text{ cm}^{-1}$. That only about $(20 \times 47/8066) \text{ eV} = 0.1 \text{ eV}$ deposited into the 47 cm^{-1} mode, whereas 1.6 eV of optical energy is lost to the overall relaxation process, indicates

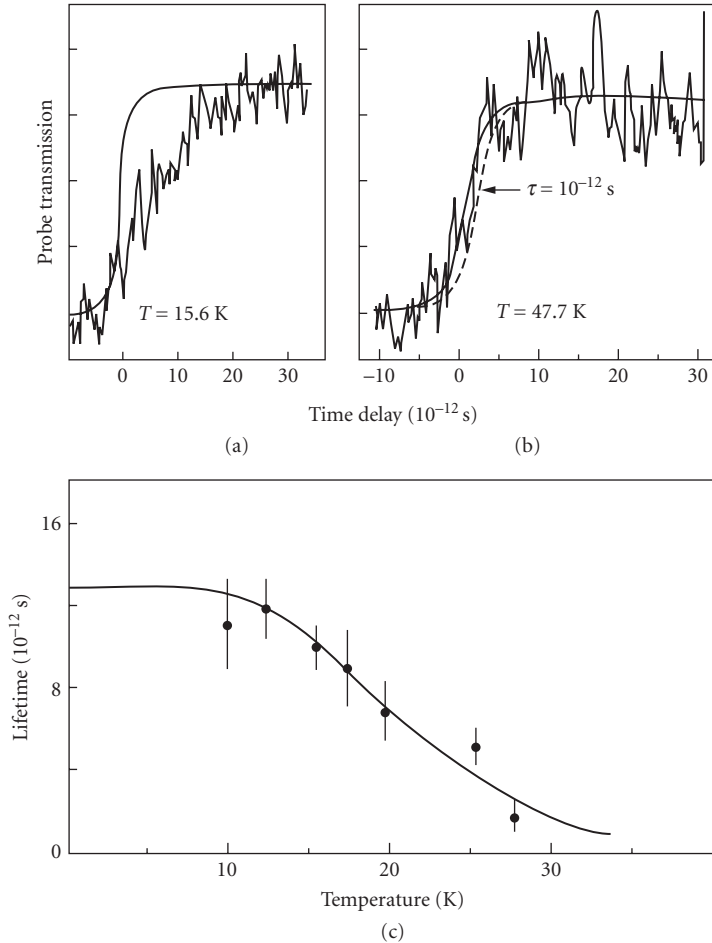


FIGURE 12 (a) Temporal evolution of gain in the $F_A(\text{Li})$ center emission in picosecond pulse probe measurements; (b) the temperature-dependence of the gain process; and (c) temperature dependence of the relaxation time. (After Mollenauer *et al.*⁴⁴)

that other higher-energy modes of vibrations must be involved.⁴⁴ This problem is resolved by Raman-scattering experiments. For $F_A(\text{Li})$ -centers in potassium chloride, three sharp Raman-active local modes were observed with energies of 47 cm^{-1} , 216 cm^{-1} , and 266 cm^{-1} , for the ^7Li isotope.⁴⁵ These results and later polarized absorption/luminescence studies indicated that the Li^+ ion lies in an off-center position in a $\langle 110 \rangle$ crystal direction relative to the z axis of the F_A center. Detailed polarized Raman spectroscopy resonant and nonresonant with the F_A -center absorption bands are shown in Fig. 13.⁴⁶ These spectra show that under resonant excitation in the F_{A1} absorption band, each of the three lines due to the sharp localized modes is present in the spectrum. The polarization dependence confirms that the 266 cm^{-1} mode is due to Li^+ ion motion in the mirror plane and parallel to the defect axis. The 216 cm^{-1} mode is stronger under nonresonant excitation, reflecting the off-axis vibrations of the Li^+ ion vibrating in the mirror plane perpendicular to the z axis. On the other hand, the low-frequency mode is an amplified band mode of the center which hardly involves the motion of the Li^+ ion.

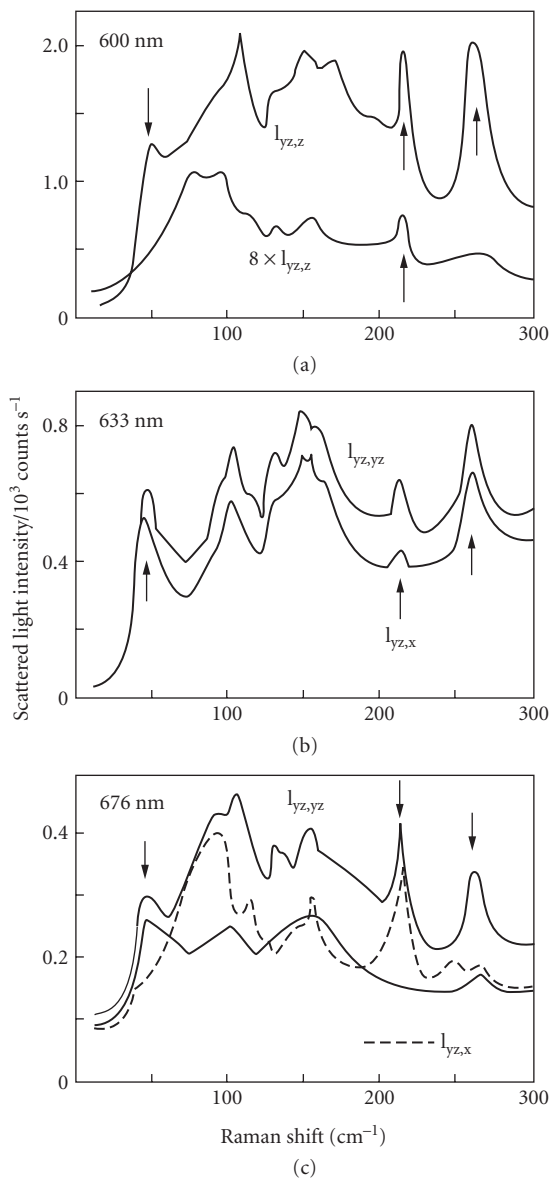


FIGURE 13 Raman spectra of F_A (Li) centers in potassium chloride measured at 10 K for different senses of polarization. In (a) the excitation wavelength $\lambda = 600$ nm is midway between the peaks of the F_{A1} bands; (b) $\lambda = 632.8$ nm is resonant with the F_{A1} band; and (c) $\lambda = 676.4$ nm is nonresonant. (After Joosen *et al.*⁴⁴)

2.5 ABSORPTION, PHOTOLUMINESCENCE, AND RADIATIVE DECAY MEASUREMENTS

The philosophy of solid-state spectroscopy is subtly different from that of atomic and molecular spectroscopies. It is often required not only to determine the nature of the absorbing/emitting species but also the symmetry and structure of the local environment. Also involved is the interaction of the electronic center with other neighboring ions, which leads to lineshape effects as well as time-dependent phenomena. The consequence is that a combination of optical spectroscopic techniques may be used in concert. This general approach to optical spectroscopy of condensed matter phenomena is illustrated by reference to the case of Al_2O_3 and MgO doped with Cr^{3+} .

Absorption and Photoluminescence of Cr^{3+} in Al_2O_3 and MgO

The absorption and luminescence spectra may be interpreted using the Tanabe-Sugano diagram shown in Fig. 5, as discussed previously. Generally, the optical absorption spectrum of Cr^{3+} : Al_2O_3 (Fig. 14) is dominated by broadband transitions from the ${}^4\text{A}_2 \rightarrow {}^4\text{T}_2$ and ${}^4\text{A}_2 \rightarrow {}^4\text{T}_1$. The crystal used in this measurement contained some $10^{18} \text{ Cr}^{3+} \text{ ions cm}^{-3}$. Since the absorption coefficient at the peak of the ${}^4\text{A}_2 \rightarrow {}^4\text{T}_2$ band is only 2 cm^{-1} , it is evident from Eq. (6) in Chap. 31, "Optical Spectrometers," in Vol. I, that the cross section at the band peak is $\sigma_o \cong 5 \times 10^{-19} \text{ cm}^2$. The spin-forbidden absorption transitions ${}^4\text{A}_2 \rightarrow {}^2\text{E}$, ${}^2\text{T}_1$ are just distinguished as weak absorptions ($\sigma_o \sim 10^{-21} \text{ cm}^2$) on the long-wavelength side of the ${}^4\text{A}_2 \rightarrow {}^4\text{T}_2$ band. This analysis strictly applies to the case of octahedral symmetry. Since the cation site in ruby is distorted from perfect octahedral symmetry, there are additional electrostatic energy terms associated with this reduced symmetry. One result of this distortion, as illustrated in Fig. 14, is that the absorption and emission spectra are no longer optically isotropic. By measuring the peak shifts of the ${}^4\text{A}_2 \rightarrow {}^4\text{T}_2$ and ${}^4\text{A}_2 \rightarrow {}^4\text{T}_1$ absorption transitions between π and σ senses of polarization, the trigonal field splittings of the ${}^4\text{T}_2$ and ${}^4\text{T}_1$ levels may be determined.²⁵

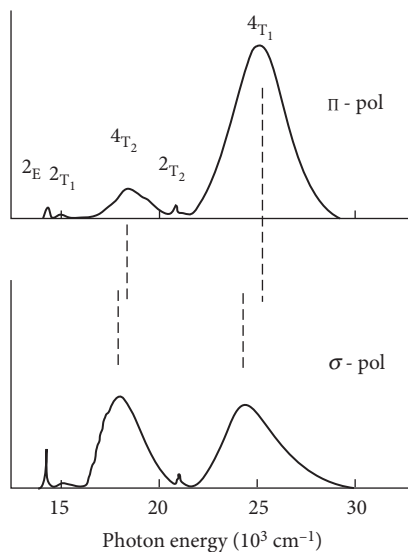


FIGURE 14 Polarized optical absorption spectrum of a ruby Cr^{3+} : Al_2O_3 crystal containing $2 \times 10^{18} \text{ Cr}^{3+} \text{ ions cm}^{-3}$ measured at 77 K.

The Cr^{3+} ion enters the MgO substitutionally for the Mg^{2+} ion. The charge imbalance requires that for every two impurity ions there must be one cation vacancy. At low-impurity concentrations, charge-compensating vacancies are mostly remote from the Cr^{3+} ions. However, some 10 to 20 percent of the vacancies occupy sites close to individual Cr^{3+} ions, thereby reducing the local symmetry from octahedral to tetragonal or orthorhombic.¹⁹ The optical absorption spectrum of $\text{Cr}^{3+}:\text{MgO}$ is also dominated by broadband ${}^4\text{A}_2 \rightarrow {}^4\text{T}_2$, ${}^4\text{T}_1$ transitions; in this case, there are overlapping contributions from Cr^{3+} ions in three different sites. There are substantial differences between the luminescence spectra of Cr^{3+} in the three different sites in MgO (Fig. 15), these overlapping spectra being determined by the ordering of the ${}^4\text{T}_2$ and ${}^2\text{E}$ excited states. For strong crystal fields, $Dq/B > 2.5$, ${}^2\text{E}$ lies lowest and nonradiative decay from ${}^4\text{T}_1$ and ${}^4\text{T}_2$ levels to ${}^2\text{E}$ results in very strong emission in the sharp *R*-lines, with rather weaker vibronic sidebands. This is the situation from Cr^{3+} ions in octahedral and tetragonal sites in MgO.¹⁹ The ${}^2\text{E} \rightarrow {}^4\text{A}_2$ luminescence transition is both spin- and parity-forbidden (see Vol. I, Chap. 10) and this is signaled by relatively long radiative lifetimes—11.3 ms for octahedral sites and 8.5 ms for tetragonal sites at 77 K. This behavior is in contrast to that of Cr^{3+} in orthorhombic sites, for which the ${}^4\text{T}_2$ level lies below the ${}^2\text{E}$ level. The stronger electron-phonon coupling for the ${}^4\text{T}_2 \rightarrow {}^4\text{A}_2$ transition at

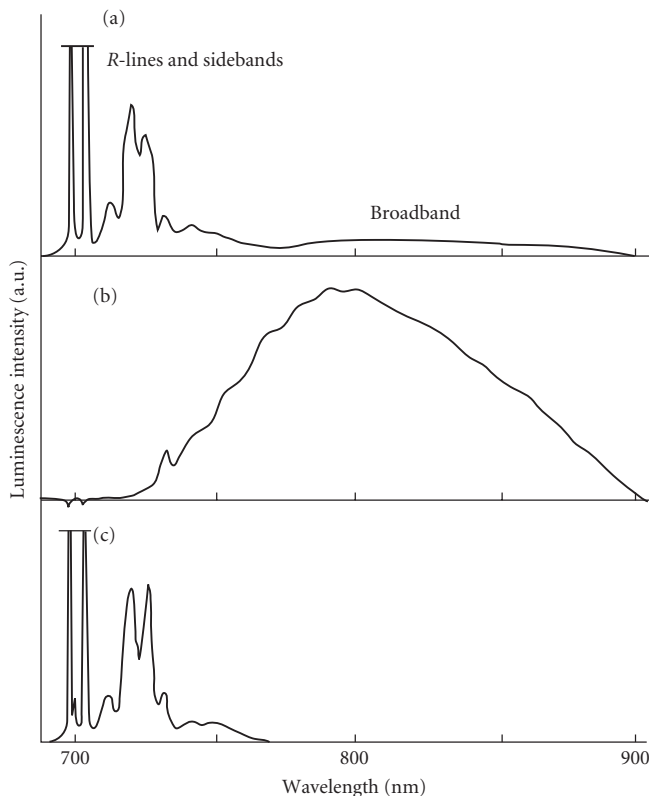


FIGURE 15 Photoluminescence spectra of $\text{Cr}^{3+}:\text{MgO}$ using techniques of phase-sensitive detection. In (a) the most intense features are sharp *R*-lines near 698 to 705 nm due to Cr^{3+} ions at sites with octahedral and tetragonal symmetry; a weak broadband with peak at 740 nm is due to Cr^{3+} ions in sites of orthorhombic symmetry. By adjusting the phase-shift control on the lock-in amplifier (Fig. 4 in Vol. I, Chap. 31), the relative intensities of the three components may be adjusted as in plots (b) and (c).

orthorhombic sites leads to a broadband luminescence with peak at 790 nm. Since this is a spin-allowed transition, the radiative lifetime is much shorter—only 35 μs .⁴⁷

As noted previously the decay time of the luminescence signals from Cr^{3+} ions in octahedral and tetragonal symmetry are quite similar, and good separation of the associated R -lines using the phase-nulling technique are then difficult. However, as Fig. 15 shows, good separation of these signals from the ${}^4\text{T}_2 \rightarrow {}^4\text{A}_2$ broadband is very good. This follows from the applications of Eqs. (8) through (12) in Vol. I, Chap. 31. For Cr^{3+} ions in cubic sites, the long lifetime corresponds to a signal phase angle of 2° : the R -line intensity can be suppressed by adjusting the detector phase angle to $(90^\circ + 2^\circ)$. In contrast, the Cr^{3+} ions in orthorhombic sites give rise to a phase angle of 85° : this signal is reduced to zero when $\phi_D = (90^\circ + 85^\circ)$.

Excitation Spectroscopy

The precise positions of the ${}^4\text{A}_2 \rightarrow {}^4\text{T}_2$, ${}^4\text{T}_1$ absorption peaks corresponding to the sharp lines and broadbands in Fig. 15 may be determined by excitation spectroscopy (see Vol. I, Chap. 31). An example of the application of this technique is given in Fig. 10, which shows the emission band of the ${}^4\text{T}_2 \rightarrow {}^4\text{A}_2$ transition at centers with orthorhombic symmetry, Figs. 10a and b, and its excitation spectrum, Fig. 10c.⁴⁷ The latter was measured by setting the wavelength of the detection spectrometer at $\lambda = 790$ nm, i.e., the emission band peak, and scattering the excitation monochromator over the wavelength range 350 to 750 nm of the Xe lamp. Figure 10 gives an indication of the power of excitation spectroscopy in uncovering absorption bands not normally detectable under the much stronger absorptions from cubic and tetragonal centers. Another example is given in Fig. 16—in this case, of recombining excitons in the smaller gap material (GaAs) in GaAs/AlGaAs quantum wells. In this case, the exciton luminescence peak energy, $h\nu_x$, is given by

$$h\nu_x = E_G + E_{1e} + E_{1h} + E_b \quad (3)$$

where E_G is the bandgap of GaAs, E_{1e} and E_{1h} are the $n = 1$ state energies of electrons (e) and holes (h) in conduction and valence bands, respectively, and E_b is the electron-hole binding energy. Optical transitions involving electrons and holes in these structures are subject to the $\Delta n = 0$ selection rule. In consequence, there is a range of different absorption transitions at energies above the bandgap. Due to the rapid relaxation of energy in levels with $n > 1$, the recombination luminescence occurs between the $n = 1$ electron and hole levels only, in this case at 782 nm. The excitation spectrum in which this luminescence is detected and excitation wavelength varied at wavelengths shorter than 782 nm reveals the presence of absorption transitions above the bandgap. The first absorption transition shown is the $1\text{lh} \rightarrow 1e$ transition, which occurs at slightly longer wavelength than the $1\text{hh} \rightarrow 1e$ transition. The light hole (lh)-heavy hole (hh) splitting is caused by spin-orbit splitting and strain in these epilayer structures. Other, weaker transitions are also discernible at higher photon energies.

Polarization Spectroscopy

The discussions on optical selection rules, in Vol. 1, Chaps. 10 and 31, showed that when a well-defined axis is presented, the strength of optical transitions may depend strongly on polarization. In atomic physics the physical axis is provided by an applied magnetic field (Zeeman effect) or an applied electric field (Stark effect). Polarization effects in solid-state spectroscopy may be used to give information about the site symmetry of optically active centers. The optical properties of octahedral crystals are normally isotropic. In this situation, the local symmetry of the center must be lower than octahedral so that advantage may be taken of the polarization-sensitivity of the selection rules. Several possibilities exist in noncubic crystals. If the local symmetry of all centers in the crystal point in the same direction, then the crystal as a whole displays an axis of symmetry. Sapphire (Al_2O_3) is an example, in which the Al^{3+} ions occupy trigonally distorted octahedral sites. In consequence, the optical absorption and luminescence spectra of ions in this crystal are naturally polarized. The observed π - and σ -polarized absorption spectra of ruby shown in Fig. 14 are in general agreement with the calculated selection rules,²⁵

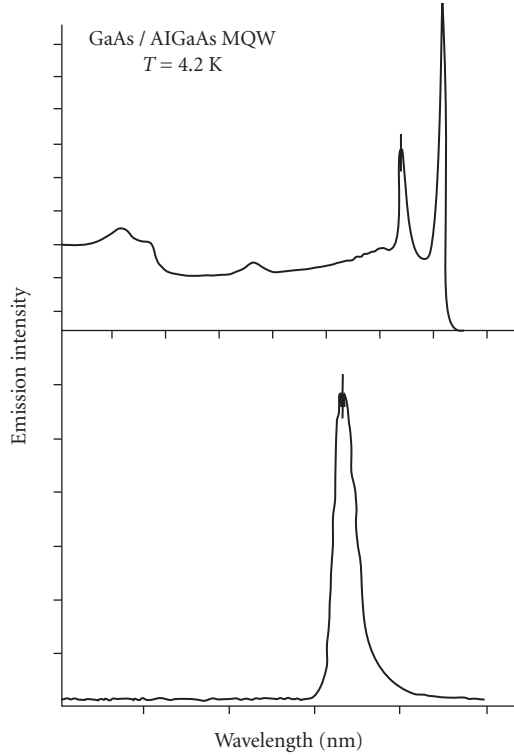


FIGURE 16 Luminescence spectrum and excitation spectrum of multiple quantum wells in GaAs/AlGaAs samples measured at 6 K. (*P. Dawson, 1986 private communication to the author.*)

although there are undoubtedly vibronic processes contributing to these broadband intensities.⁴⁷ The other important ingredient in the spectroscopy of the Cr^{3+} ions in orthorhombic symmetry sites in MgO is that the absorption and luminescence spectra are strongly polarized. It is then quite instructive to indicate how the techniques of polarized absorption/luminescence help to determine the symmetry axes of the dipole transitions. The polarization of the ${}^4\text{T}_2 \rightarrow {}^4\text{A}_2$ emission transition in Fig. 10 is clear. In measurements employing the “straight-through” geometry, Henry et al.⁴⁷ reported the orientation intensity patterns shown in Fig. 17 for the broadband spectrum. A formal calculation of the selection rules and the orientation dependence of the intensities shows that the intensity at angle θ is given by

$$I(\theta) = (A_\pi - A_\sigma)(E_\pi - E_\sigma) \sin^2\left(\theta + \frac{\pi}{4}\right) + \text{constant} \quad (4)$$

where A and E refer to the absorbed and emitted intensities for π - and σ -polarizations.⁴⁸ The results in Fig. 17 are consistent with the dipoles being aligned along $\langle 110 \rangle$ directions of the octahedral MgO lattice. This is in accord with the model of the structure of the Cr^{3+} ions in orthorhombic symmetry, which locates the vacancy in the nearest neighbor cation site relative to the Cr^{3+} ion along a $\langle 110 \rangle$ direction.

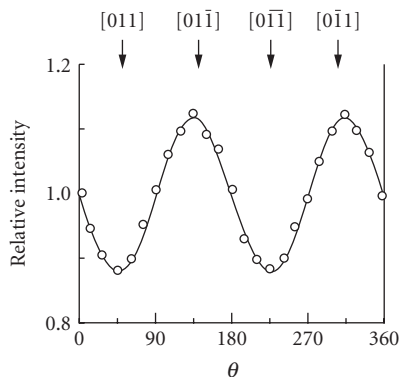


FIGURE 17 The polarization characteristics of the luminescence spectrum of Cr^{3+} ions in orthorhombic sites in $\text{Cr}^{3+}:\text{MgO}$. (After Henry et al.⁴⁷)

Zeeman Spectroscopy

The Zeeman effect is the splitting of optical lines by a static magnetic field due to the removal of the spin degeneracy of levels involved in the optical transitions. In many situations the splittings are not much larger than the optical linewidth of zero-phonon lines and much less than the width of vibronically broadened bands. The technique of optically detected magnetic resonance (ODMR) is then used to measure the Zeeman splittings. As we have already shown, ODMR also has the combined ability to link inextricably, an excited-state ESR spectrum with an absorption band and a luminescence band. The spectrum shown in Fig. 16 in Vol. I, Chap. 31 is an example of this unique power, which has been used in such diverse situations as color centers, transition-metal ions, rare-earth ions, phosphor- and laser-active ions (e.g., Ga^+ , Tl^0), as well as donor-acceptor and exciton recombination in semiconductors.¹⁹ We now illustrate the relationship of the selection rules and polarization properties of the triplet-singlet transitions.

The F -center in calcium oxide consists of two electrons trapped in the Coulomb field of a negative-ion vacancy. The ground state is a spin singlet, $^1A_{1g}$, from which electric dipole absorption transitions are allowed into a $^1T_{1u}$ state derived from the $(1s2p)$ configuration. Such $^1A_{1g} \rightarrow ^1T_{1u}$ transitions are signified by a strong optical absorption band centered at a wavelength $\lambda \approx 400$ nm (Vol. I, Chap. 31, Fig. 16). Dexcitation of this $^1T_{1u}$ state does not proceed via $^1T_{1u} \rightarrow ^1A_{1g}$ luminescence. Instead, there is an efficient nonradiative decay from $^1T_{1u}$ into the triplet $^3T_{1u}$ state also derived from the $(1s2p)$ configuration.⁴⁹ The spin-forbidden $^3T_{1u} \rightarrow ^1A_{1g}$ transition gives rise to a striking orange fluorescence, which occurs with a radiative lifetime $\tau_R = 3.4$ ms at 4.2 K. The ODMR spectrum of the F -center and its absorption and emission spectral dependences are depicted in Fig. 16 in Vol. I, Chap. 31, other details are shown in Fig. 18. With the magnetic field at some general orientation in the (100) plane there are six lines. From the variation of the resonant fields with the orientation of the magnetic field in the crystal, Edel et al. (1972)⁵⁰ identified the spectrum with the $S = 1$ state of tetragonally distorted F -center. The measured orientation dependence gives $g_{\parallel} \approx g_{\perp} = 1.999$ $D = 60.5$ mT.

Figure 18 shows the selection rules for emission of circularly polarized light by $S = 1$ states in axial crystal fields. We denote the populations of the $M_s = 0, \pm 1$ levels as N_0 and $N_{\pm 1}$. The low-field ESR line, corresponding to the $M_s = 0 \rightarrow M_s = +1$ transition, should be observed as an increase in σ_+ -light because $N_0 > N_{+1}$ and ESR transitions enhance the $M_s = \pm 1$ level. However, the high-field line is observed as a change in intensity of σ_- -light. If spin-lattice relaxation is efficient (i.e., $T_1 < \tau_R$), then the spin states are in thermal equilibrium, $N_0 < N_{-1}$, ESR transitions depopulate the $|M_s = -1\rangle$ level. Thus, the high-field ODMR line is seen as a decrease in the F -center in these crystals (viz., that

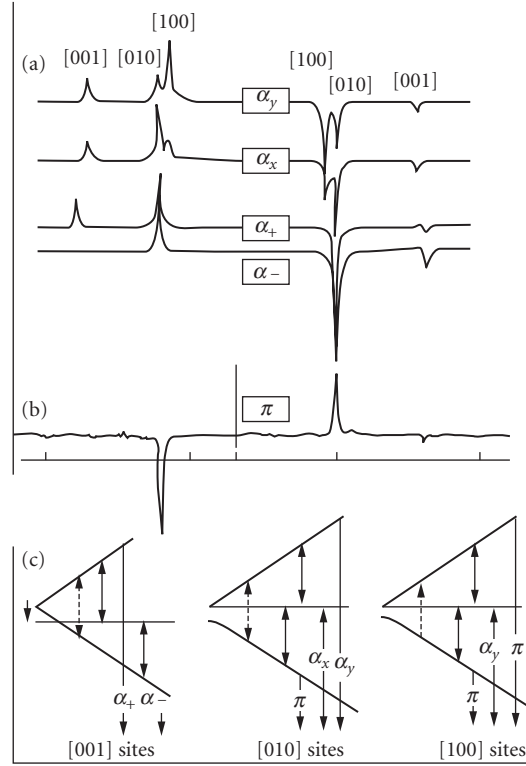


FIGURE 18 Polarization selection rules and the appropriately detected ODMR spectra of the ${}^3T_{1u} \rightarrow {}^1A_{1g}$ transition of F -centers in CaO. (After Edel *et al.*⁵⁰)

for the lowest ${}^3T_1$ state, D is positive and the spin states are in thermal equilibrium). It is worth noting that since the $|M_s = 0\rangle \rightarrow |M_s = \pm 1\rangle$ ESR transitions occur at different values of the magnetic field, ODMR may be detected simply as a change in the emission intensity at resonance; it is not necessary to measure specifically the sense of polarization of the emitted light. The experimental data clearly establish the tetragonal symmetry of the F -center in calcium oxide: the tetragonal distortion occurs in the excited ${}^3T_{1u}$ state due to vibronic coupling to modes of E_g symmetry resulting in a static Jahn-Teller effect.⁵⁰

2.6 REFERENCES

1. T. W. Hänsch, I. S. Shakin, and A. L. Shawlow, *Nature* (London), **225**:63 (1972).
2. D. N. Stacey, private communication to A. I. Ferguson.
3. E. Giacobino and F. Birabem, *J. Phys.* B**15**:L385 (1982).
4. L. Housek, S. A. Lee, and W. M. Fairbank Jr., *Phys. Rev. Lett.* **50**:328 (1983).
5. P. Zhao, J. R. Lawall, A. W. Kam, M. D. Lindsay, F. M. Pipkin, and W. Lichten, *Phys. Rev. Lett.* **63**:1593 (1989).

6. T. J. Sears, S. C. Foster, and A. R. W. McKellar, *J. Opt. Soc. Am.* **B3**:1037 (1986).
7. C. J. Sansonetti, J. D. Gillaspy, and C. L. Cromer, *Phys. Rev. Lett.* **65**:2539 (1990).
8. W. Lichten, D. Shinen, and Zhi-Xiang Zhou, *Phys. Rev.* **A43**:1663 (1991).
9. C. Adams and A. I. Ferguson, *Opt. Commun.* **75**:419 (1990) and **79**:219 (1990).
10. C. Adams, E. Riis, and A. I. Ferguson, *Phys. Rev. A* (1992) and **A45**:2667 (1992).
11. G. W. F. Drake and A. J. Makowski, *J. Opt. Soc. Amer.* **B5**:2207 (1988).
12. P. G. Harris, H. C. Bryant, A. H. Mohagheghi, R. A. Reeder, H. Sharifian, H. Tootoonchi, C. Y. Tang, J. B. Donahue, C. R. Quick, D. C. Rislove, and W. W. Smith, *Phys. Rev. Lett.* **65**:309 (1990).
13. H. R. Sadeghpour and C. H. Greene, *Phys. Rev. Lett.* **65**:313 (1990).
14. See, for example, H. Friedrich, *Physics World* **5**:32 (1992).
15. Iu et al., *Phys. Rev. Lett.* **66**:145 (1991).
16. B. MacGowan et al., *Phys. Rev. Lett.* **65**:420 (1991).
17. C. Douketis and T. E. Gough, *J. Mol. Spectrosc.* **101**:325 (1983).
18. See, for example, K. Codling et al., *J. Phys.* **B24**:L593 (1991).
19. B. Henderson and G. F. Imbusch, *Optical Spectroscopy of Inorganic Solids*, Clarendon Press, Oxford, 1989.
20. G. Blasse, in B. Di Bartolo (ed.), *Energy Transfer Processes in Condensed Matter*, Plenum Press, New York, 1984.
21. Y. Tanabe and S. Sugano, *J. Phys. Soc. Jap.* **9**:753 (1954).
22. R. M. MacFarlane, J. Y. Wong, and M. D. Sturge, *Phys. Rev.* **166**:250 (1968).
23. B. F. Gachter and J. A. Königstein, *J. Chem. Phys.* **66**:2003 (1974).
24. See O. Deutschbein, *Ann. Phys.* **20**:828 (1932).
25. D. S. McClure, *J. Chem. Phys.* **36**:2757 (1962).
26. After C. J. Delbecq, W. Hayes, M. C. M. O'Brien, and P. H. Yuster, *Proc. Roy. Soc.* **A271**:243 (1963).
27. W. B. Fowler, in Fowler (ed.), *Physics of Color Centers*, Academic Press, New York, 1968. See also G. Boulon, in B. Di Bartolo (ed.), *Spectroscopy of Solid State Laser-Type Materials*, Plenum Press, New York, 1988.
28. Le Si Dang, Y. Merle d'Aubigné, R. Romestain, and A. Fukuda, *Phys. Rev. Lett.* **38**:1539 (1977).
29. A. Ranfagni, D. Mugna, M. Bacci, G. Villiani, and M. P. Fontana, *Adv. in Phys.* **32**:823 (1983).
30. E. Sonder and W. A. Sibley, in J. H. Crawford and L. F. Slifkin (eds.), *Point Defects in Solids*, Plenum Press, New York, vol. 1, 1972.
31. Y. Fang, P. J. Parbrook, B. Henderson, and K. P. O'Donnell, *Appl. Phys. Letts.* **59**:2142 (1991).
32. P. J. Parbrook, B. Cockayne, P. J. Wright, B. Henderson, and K. P. O'Donnell, *Semicond. Sci. Technol.* **6**:812 (1991).
33. T. Kushida and E. Takushi, *Phys. Rev.* **B12**:824 (1975).
34. R. M. Macfarlane, R. T. Harley, and R. M. Shelby, *Radn. Effects* **72**:1 (183).
35. W. Yen and P. M. Selzer, in W. Yen and P. M. Selzer (eds.), *Laser Spectroscopy of Solids*, Springer-Verlag, Berlin, 1981.
36. P. E. Jessop and A. Szabo, *Optics Comm.* **33**:301 (1980).
37. N. Pelletier-Allard and R. Pelletier, *J. Phys.* **C17**:2129 (1984).
38. See Y. Gao, M. Yamaga, B. Henderson, and K. P. O'Donnell, *J. Phys. (Cond. Matter)* (1992) in press (and references therein).
39. G. E. Peckham, *Proc. Phys. Soc. (Lond.)* **90**:657 (1967).
40. J. M. Worlock and S. P. S. Porto, *Phys. Rev. Lett.* **15**:697 (1965).
41. F. Luty, in W. B. Fowler (eds.), *The Physics of Color Centers*, Academic Press, New York, 1968.
42. L. F. Mollenauer and D. H. Olson, *J. App. Phys.* **24**:386 (1974).
43. F. Luty and W. Gellerman, in C. B. Collins (ed.), *Lasers '81*, STS Press, McClean, 1982.
44. L. F. Mollenauer, J. M. Wiesenfeld, and E. P. Ippen, *Radiation Effects* **72**:73 (1983); see also J. M. Wiesenfeld, L. F. Mollenauer, and E. P. Ippen, *Phys. Rev. Lett.* **47**:1668 (1981).

45. B. Fritz, J. Gerlach, and U. Gross, in R. F. Wallis (ed.), *Localised Excitations in Solids*, Plenum Press, New York, 1968, p. 496.
46. W. Joosen, M. Leblans, M. Vahimbeek, M. de Raedt, E. Goovaert, and D. Schoemaker, *J. Cryst. Def. Amorph. Solids* **16**:341 (1988).
47. M. O. Henry, J. P. Larkin, and G. F. Imbusch, *Phys. Rev.* **B13**:1893 (1976).
48. M. Yamaga, B. Henderson, and K. P. O'Donnell, *J. Luminescence* **43**:139 (1989); see also *ibid.* **46**:397 (1990).
49. B. Henderson, S. E. Stokowski, and T. C. Ensign, *Phys. Rev.* **183**:826 (1969).
50. P. Edel, C. Hennies, Y. Merle d'Aubigné, R. Romestain, and Y. Twarowski, *Phys. Rev. Lett.* **28**:1268 (1972).

PART

2

ATMOSPHERIC
OPTICS

This page intentionally left blank

ATMOSPHERIC OPTICS

Dennis K. Killinger

*Department of Physics
Center for Laser Atmospheric Sensing
University of South Florida
Tampa, Florida*

James H. Churnside

*National Oceanic and Atmospheric Administration
Earth System Research Laboratory
Boulder, Colorado*

Laurence S. Rothman

*Harvard-Smithsonian Center for Astrophysics
Atomic and Molecular Physics Division
Cambridge, Massachusetts*

3.1 GLOSSARY

c	speed of light
C_n^2	atmospheric turbulence strength parameter
D	beam diameter
F	hypergeometric function
$g(\nu)$	optical absorption lineshape function
H	height above sea level
h	Planck's constant
I	Irradiance (intensity) of optical beam (W/m^2)
k	optical wave number
K	turbulent wave number
L	propagation path length
L_0	outer scale size of atmospheric turbulence
l_0	inner scale size of atmospheric turbulence
N	density or concentration of molecules
$p(I)$	probability density function of irradiance fluctuations
P_ν	Planck radiation function
R	gas constant

S	molecular absorption line intensity
T	temperature
v	wind speed
β	backscatter coefficient of the atmosphere
γ_p	pressure-broadened half-width of absorption line
κ	optical attenuation
λ	wavelength
ν	optical frequency (wave numbers)
ρ_0	phase coherence length
σ_l^2	variance of irradiance fluctuations
σ_R	Rayleigh scattering cross section

3.2 INTRODUCTION

Atmospheric optics involves the transmission, absorption, emission, refraction, and reflection of light by the atmosphere and is probably one of the most widely observed of all optical phenomena.¹⁻⁵ The atmosphere interacts with light due to the composition of the atmosphere, which under normal conditions, consists of a variety of different molecular species and small particles like aerosols, water droplets, and ice particles. This interaction of the atmosphere with light is observed to produce a wide variety of optical phenomena including the blue color of the sky, the red sunset, the optical absorption of specific wavelengths due to atmospheric molecules, the twinkling of stars at night, the greenish tint sometimes observed during a severe storm due to the high density of particles in the atmosphere, and is critical in determining the balance between incoming sunlight and outgoing infrared (IR) radiation and thus influencing the earth's climate.

One of the most basic optical phenomena of the atmosphere is the absorption of light. This absorption process can be depicted as in Fig. 1 which shows the transmission spectrum of the atmosphere as

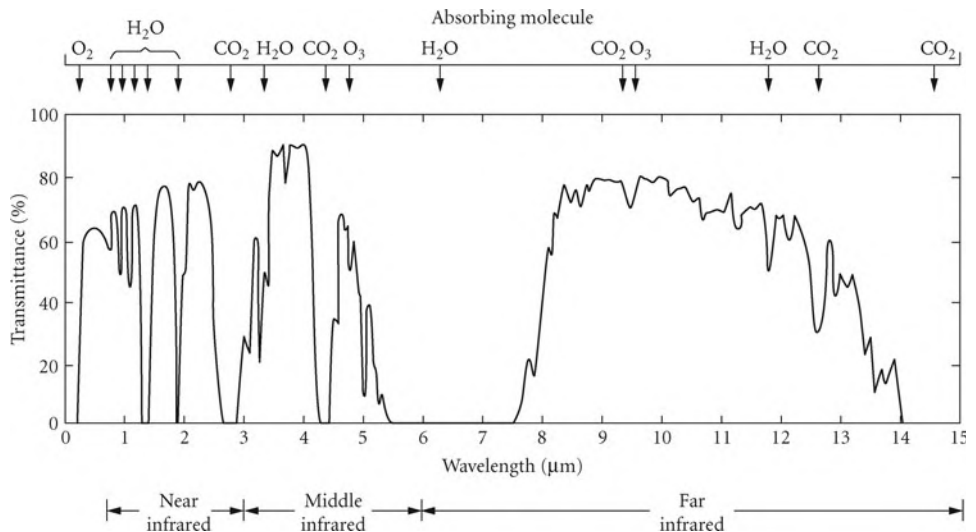


FIGURE 1 Transmittance through the earth's atmosphere as a function of wavelength taken with low spectral resolution (path length 1800 m). (From Measures, Ref. 5.)

a function of wavelength.⁵ The transmission of the atmosphere is highly dependent upon the wavelength of the spectral radiation, and, as will be covered later in this chapter, upon the composition and specific optical properties of the constituents in the atmosphere. The prominent spectral features in the transmission spectrum in Fig. 1 are primarily due to absorption bands and individual absorption lines of the molecular gases in the atmosphere, while a portion of the slowly varying background transmission is due to aerosol extinction and continuum absorption.

This chapter presents a tutorial overview of some of the basic optical properties of the atmosphere, with an emphasis on those properties associated with optical propagation and transmission of light through the earth's atmosphere. The physical phenomena of optical absorption, scattering, emission, and refractive properties of the atmosphere will be covered for optical wavelengths from the ultraviolet (UV) to the far-infrared. The primary focus of this chapter is on *linear* optical properties associated with the transmission of light through the atmosphere. Historically, the study of atmospheric optics has centered on the radiance transfer function of the atmosphere, and the linear transmission spectrum and blackbody emission spectrum of the atmosphere. This emphasis was due to the large body of research associated with passive, electro-optical sensors which primarily use the transmission of ambient optical light or light from selected emission sources. During the past few decades, however, the use of lasers has added a new dimension to the study of atmospheric optics. In this case, not only is one interested in the transmission of light through the atmosphere, but also information regarding the optical properties of the backscattered optical radiation.

In this chapter, the standard linear optical interactions of an optical or laser beam with the atmosphere will be covered, with an emphasis placed on linear absorption and scattering interactions. It should be mentioned that the first edition of the OSA *Handbook of Optics* chapter on "Atmospheric Optics" had considerable nomographs and computational charts to aid the user in numerically calculating the transmission of the atmosphere.² Because of the present availability of a wide range of spectral databases and computer programs (such as the HITRAN Spectroscopy Database, LOWTRAN, MODTRAN, and FASCODE atmospheric transmission computer programs) that model and calculate the transmission of light through the atmosphere, these nomographs, while still useful, are not as vital. As a result, the emphasis on this edition of the "Atmospheric Optics" chapter is on the basic theory of the optical interactions, how this theory is used to model the optics of the atmosphere, the use of available computer programs and databases to calculate the optical properties of the atmosphere, and examples of instruments and meteorological phenomena related to optical or visual remote sensing of the atmosphere.

The overall organization of this chapter begins with a description of the natural, homogeneous atmosphere and the representation of its physical and chemical composition as a function of altitude. A brief survey is then made of the major linear optical interactions that can occur between a propagating optical beam and the naturally occurring constituents in the atmosphere. The next section covers several major computational programs (HITRAN, LOWTRAN, MODTRAN, and FASCODE) and U.S. Standard Atmospheric Models which are used to compute the optical transmission, scattering, and absorption properties of the atmosphere. The next major technical section presents an overview of the influence of atmospheric refractive turbulence on the statistical propagation of an optical beam or wavefront through the atmosphere. Finally, the last few sections of the chapter include a brief introduction to some optical and laser remote sensing experiments of the atmosphere, a brief introduction to the visually important field of meteorological optics, and references to the critical influence of atmospheric optics on global climate change.

It should be noted that the material contained within this chapter has been compiled from several recent overview/summary publications on the optical transmission and atmospheric composition of the atmosphere, as well as from a large number of technical reports and journal publications. These major overview references are (1) *Atmospheric Radiation*, (2) the previous edition of the OSA *Handbook of Optics* (chapter on "Optical Properties of the Atmosphere"), (3) *Handbook of Geophysics and the Space Environment* (chapter on "Optical and Infrared Properties of the Atmosphere"), (4) *The Infrared Handbook*, and (5) *Laser Remote Sensing*.¹⁻⁵ The interested reader is directed toward these comprehensive treatments as well as to the listed references therein for detailed information concerning the topics covered in this brief overview of atmospheric optics.

3.3 PHYSICAL AND CHEMICAL COMPOSITION OF THE STANDARD ATMOSPHERE

The atmosphere is a fluid composed of gases and particles whose physical and chemical properties vary as a function of time, altitude, and geographical location. Although these properties can be highly dependent upon local and regional conditions, many of the optical properties of the atmosphere can be described to an adequate level by looking at the composition of what one normally calls a *standard atmosphere*. This section will describe the background, homogeneous standard composition of the atmosphere. This will serve as a basis for the determination of the quantitative interaction of the molecular gases and particles in the atmosphere with a propagating optical wavefront.

Molecular Gas Concentration, Pressure, and Temperature

The majority of the atmosphere is composed of lightweight molecular gases. Table 1 lists the major gases and trace species of the terrestrial atmosphere, and their approximate concentration (volume fraction) at standard room temperature (296 K), altitude at sea level, and total pressure of 1 atm.⁶ The major optically active molecular constituents of the atmosphere are N_2 , O_2 , H_2O , and CO_2 , with a secondary grouping of CH_4 , N_2O , CO , and O_3 . The other species in the table are present in the atmosphere at trace-level concentrations (ppb, down to less than ppt by volume); however, the concentration may be increased by many orders of magnitude due to local emission sources of these gases.

The temperature of the atmosphere varies both with seasonal changes and altitude. Figure 2 shows the average temperature profile of the atmosphere as a function of altitude presented for the U.S. Standard Atmosphere.⁷⁻⁹ The temperature decreases significantly with altitude until the level of the stratosphere is reached where the temperature profile has an inflection point. The U.S. Standard Atmosphere is one of six basic atmospheric models developed by the U.S. government; these different models furnish a good representation of the different atmospheric conditions which are often encountered. Figure 3 shows the temperature profile for the six atmospheric models.⁷⁻⁹

The pressure of the atmosphere decreases with altitude due to the gravitational pull of the earth and the hydrostatic equilibrium pressure of the atmospheric fluid. This is indicated in Fig. 4 which shows the total pressure of the atmosphere in millibars (1013 mb = 1 atm = 760 torr) as a function of altitude for the different atmospheric models.⁷⁻⁹ The fractional or partial pressure of most of the major gases (N_2 , O_2 , CO_2 , N_2O , CO , and CH_4) follows this profile and these gases are considered uniformly mixed. However, the concentration of water vapor is very temperature-dependent due to freezing and is not uniformly mixed in the atmosphere. Figure 5a shows the density of water vapor as a function of altitude; the units of density are in molecules/cm³ and are related to 1 atm by the appropriate value of Loschmidts number (the number of molecules in 1 cm³ of air) at a temperature of 296 K, which is 2.479×10^{19} molecules/cm³.⁷⁻⁹

The partial pressure of ozone (O_3) also varies significantly with altitude because it is generated in the upper altitudes and near ground level by solar radiation, and is in chemical equilibrium with other gases in the atmosphere which themselves vary with altitude and time of day. Figure 5b shows the typical concentration of ozone as a function of altitude.⁷⁻⁹ The ozone concentration peaks at an altitude of approximately 20 km and is one of the principle molecular optical absorbers in the atmosphere at that altitude. Further details of these atmospheric models under different atmospheric conditions are contained within the listed references and the reader is encouraged to consult these references for more detailed information.^{3,7-9}

Aerosols, Water Droplets, and Ice Particles

The atmospheric propagation of optical radiation is influenced by particulate matter suspended in the air such as aerosols (e.g., dust, haze) and water (e.g., ice or liquid cloud droplets, precipitation). Figure 6 shows the basic characteristics of particles in the atmosphere as a function of altitude,³ and Fig. 7 indicates the approximate size of common atmospheric particles.⁵

TABLE 1 List of Molecular Gases and Their Typical Concentration (Volume Fraction) for the Ambient U.S. Standard Atmosphere

Molecule	Concentration (Volume Fraction)
N ₂	0.781
O ₂	0.209
H ₂ O	0.0775 (variable)
CO ₂	3.3 × 10 ⁻⁴ (higher now: 3.9 × 10 ⁻⁴ , or 390 ppm)
A (argon)	0.0093
CH ₄	1.7 × 10 ⁻⁶
N ₂ O	3.2 × 10 ⁻⁷
CO	1.5 × 10 ⁻⁷
O ₃	2.66 × 10 ⁻⁸ (variable)
H ₂ CO	2.4 × 10 ⁻⁹
C ₂ H ₆	2 × 10 ⁻⁹
HCl	1 × 10 ⁻⁹
CH ₃ Cl	7 × 10 ⁻¹⁰
OCS	6 × 10 ⁻¹⁰
C ₂ H ₂	3 × 10 ⁻¹⁰
SO ₂	3 × 10 ⁻¹⁰
NO	3 × 10 ⁻¹⁰
H ₂ O ₂	2 × 10 ⁻¹⁰
HCN	1.7 × 10 ⁻¹⁰
HNO ₃	5 × 10 ⁻¹¹
NH ₃	5 × 10 ⁻¹¹
NO ₂	2.3 × 10 ⁻¹¹
HOCl	7.7 × 10 ⁻¹²
HI	3 × 10 ⁻¹²
HBr	1.7 × 10 ⁻¹²
OH	4.4 × 10 ⁻¹⁴
HF	1 × 10 ⁻¹⁴
ClO	1 × 10 ⁻¹⁴
HCOOH	1 × 10 ⁻¹⁴
COF ₂	1 × 10 ⁻¹⁴
SF ₆	1 × 10 ⁻¹⁴
H ₂ S	1 × 10 ⁻¹⁴
PH ₃	1 × 10 ⁻²⁰
HO ₂	Trace
O (atom)	Trace
ClONO ₂	Trace
NO ⁺	Trace
HOBr	Trace
C ₂ H ₄	Trace
CH ₃ OH	Trace
CH ₃ Br	Trace
CH ₃ CN	Trace
CF ₄	Trace

Note: The trace species have concentrations less than 1×10^{-9} , with a value that is variable and often dependent upon local emission sources.

Aerosols in the boundary layer (surface to 1 to 2 km altitude) are locally emitted, wind-driven particulates, and have the greatest variability in composition and concentration. Over land, the aerosols are mostly soil particles, dust, and organic particles from vegetation. Over the oceans, they are mostly sea salt particles. At times, however, long-range global winds are capable of transporting land particulates vast distances across the oceans or continents, especially those particulates associated with dust storms or large biomass fires, so that substantial mixing of the different particulate types may occur.

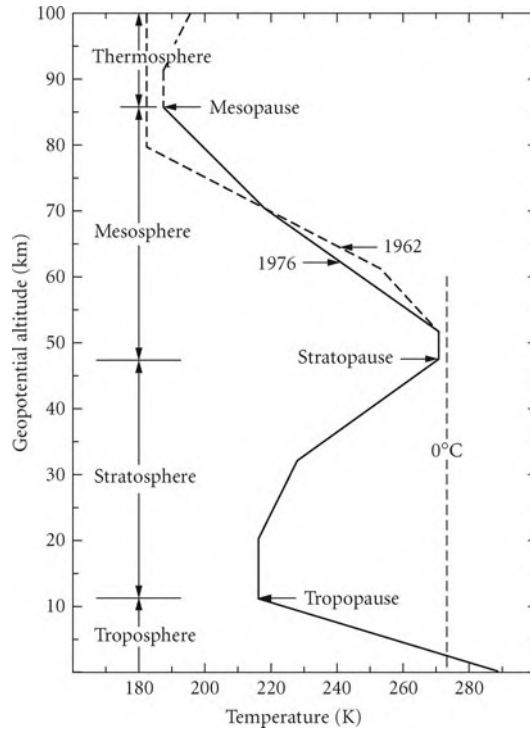


FIGURE 2 Temperature-height profile for U.S. Standard Atmosphere (0 to 86 km).

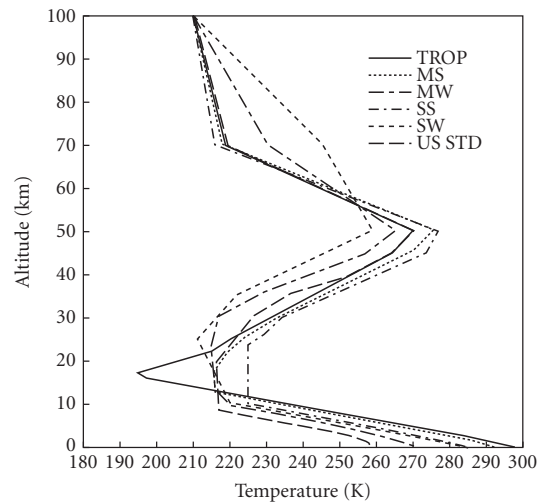


FIGURE 3 Temperature vs. altitude for the six model atmospheres: tropical (TROP), midlatitude summer (MS), midlatitude winter (MW), subarctic summer (SS), subarctic winter (SW), and U.S. standard (US STD).

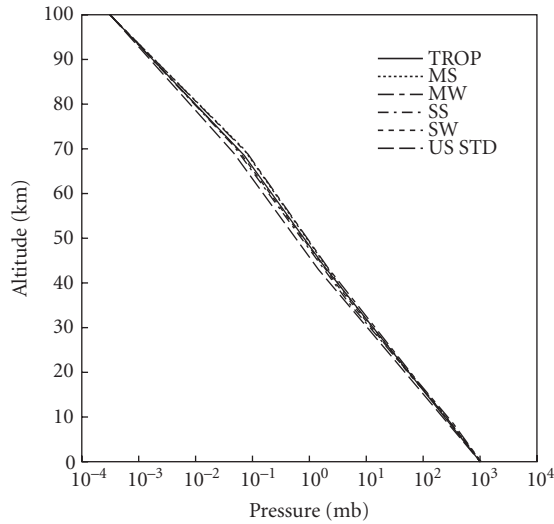


FIGURE 4 Pressure vs. altitude for the six model atmospheres.

In the troposphere above the boundary layer, the composition is less dependent upon local surface conditions and a more uniform, global distribution is observed. The aerosols observed in the troposphere are mostly due to the coagulation of gaseous compounds and fine dust. Above the troposphere, in the region of the stratosphere from 10 to 30 km, the background aerosols are mostly sulfate particles and are uniformly mixed globally. However, the concentration can be perturbed by several orders of magnitude due to the injection of dust and SO_2 by volcanic activity, such as the

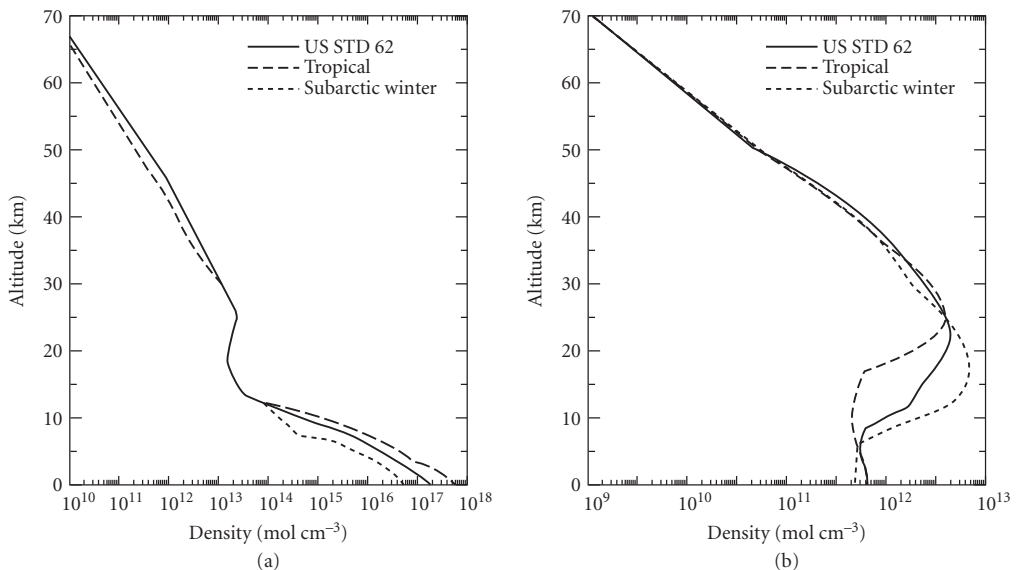


FIGURE 5 (a) Water vapor profile of several models and (b) ozone profile for several models; the U.S. standard model shown is the 1962 model.

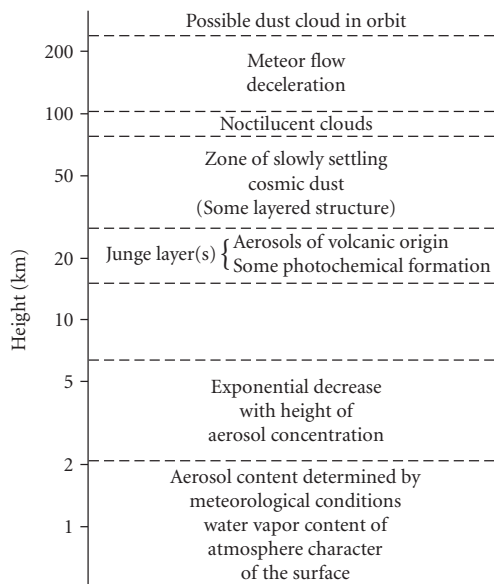


FIGURE 6 Physical characteristics of atmospheric aerosols.

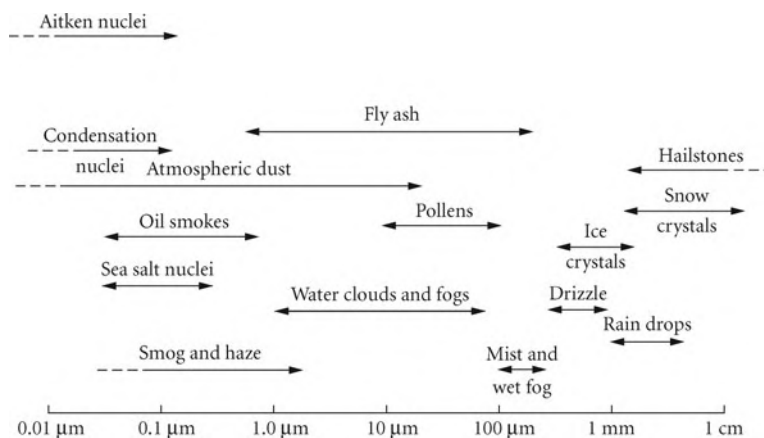


FIGURE 7 Representative diameters of common atmospheric particles. (From *Measures*, Ref. 5.)

recent eruption of Mt. Pinatubo.¹⁰ Such increases in the aerosol concentration may persist for several years and significantly impact the global temperature of the earth.

Several models have been developed for the number density and size distribution of aerosols in the atmosphere.⁷⁻⁹ Figures 8 and 9 show two aerosol distribution models appropriate for the rural environment and maritime environment, as a function of relative humidity;⁷⁻⁹ the humidity influences the size distribution of the aerosol particles and their growth characteristics. The greatest number density (particles/cm³) occurs near a size of 0.01 μm but a significant number of aerosols are still present even at the larger sizes near 1 to 2 μm . Finally, the optical characteristics of the aerosols can also be dependent upon water vapor concentration, with changes in surface, size, and growth

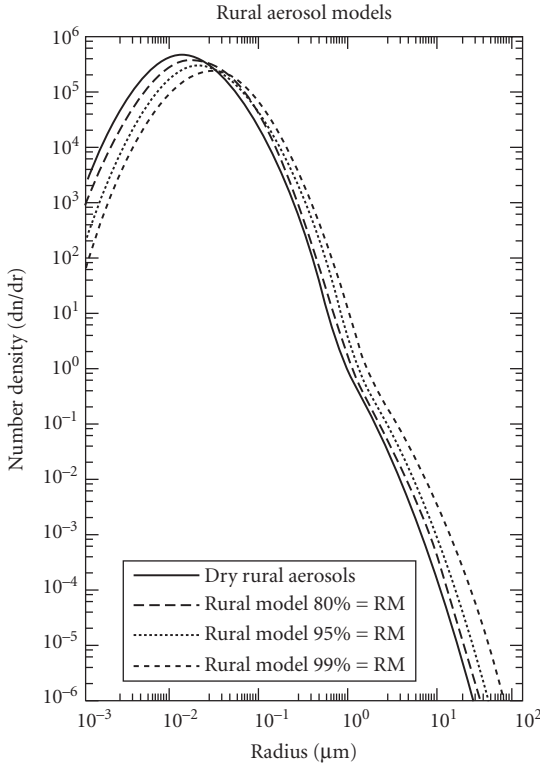


FIGURE 8 Aerosol number density distribution ($\text{cm}^{-3} \mu\text{m}^{-1}$) for the rural model at different relative humidities with total particle concentrations fixed at $15,000 \text{ cm}^{-3}$.

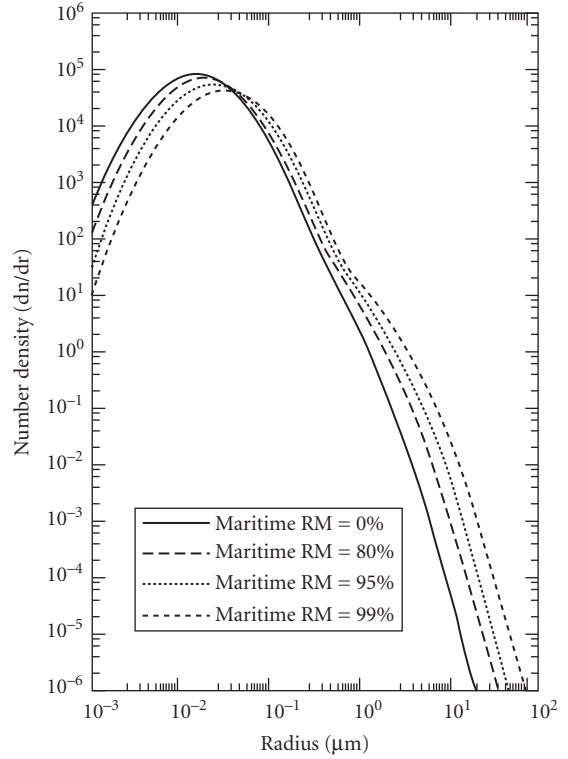


FIGURE 9 Aerosol number density distribution ($\text{cm}^{-3} \mu\text{m}^{-1}$) for the maritime model at different relative humidities with total particle concentrations fixed at 4000 cm^{-3} .

characteristics of the aerosols sometimes observed to be dependent upon the relative humidity. Such humidity changes can also influence the concentration of some pollutant gases (if these gases have been absorbed onto the surface of the aerosol particles).⁷⁻⁹

3.4 FUNDAMENTAL THEORY OF INTERACTION OF LIGHT WITH THE ATMOSPHERE

The propagation of light through the atmosphere depends upon several optical interaction phenomena and the physical composition of the atmosphere. In this section, we consider some of the basic interactions involved in the transmission, absorption, emission, and scattering of light as it passes through the atmosphere. Although all of these interactions can be described as part of an overall radiative transfer process, it is common to separate the interactions into distinct optical phenomena of molecular absorption, Rayleigh scattering, Mie or aerosol scattering, and molecular emission. Each of these basic phenomena is discussed in this section following a brief outline of the fundamental equations for the transmission of light in the atmosphere centered on the Beer-Lambert law.^{1,2}

The linear transmission (or absorption) of monochromatic light by species in the atmosphere may be expressed approximately by the Beer-Lambert law as

$$I(\lambda, t', x) = I(\lambda, t, 0) e^{-\int_0^x k(\lambda) N(x', t) dx'} \quad (1)$$

where $I(\lambda, t', x)$ is the intensity of the optical beam after passing through a path length of x , $\kappa(\lambda)$ is the optical attenuation or extinction coefficient of the species per unit of species density and length, and $N(x, t)$ is the spatial and temporal distribution of the species density that is producing the absorption; λ is the wavelength of the monochromatic light, and the parameter time t' is inserted to remind one of the potential propagation delay. Equation (1) contains the term $N(x, t)$ which explicitly indicates the spatial and temporal variability of the concentration of the attenuating species since in many experimental cases such variability may be a dominant feature.

It is common to write the attenuation coefficient in terms of coefficients that can describe the different phenomena that can cause the extinction of the optical beam. The most dominant interactions in the natural atmosphere are those due to Rayleigh (elastic) scattering, linear absorption, and Mie (aerosol/particulate) scattering; elastic means that the scattered light does not change in wavelength from that which was transmitted while inelastic infers a shift in the wavelength. In this case, one can write $\kappa(\lambda)$ as

$$\kappa(\lambda) = \kappa_a(\lambda) + \kappa_R(\lambda) + \kappa_M(\lambda) \quad (2)$$

where these terms represent the individual contributions due to absorption, Rayleigh scattering, and Mie scattering, respectively. The values for each of these extinction coefficients are described in the following sections along with the appropriate species density term $N(x, t)$. In some of these cases, the reemission of the optical radiation, possibly at a different wavelength, is also of importance. Rayleigh extinction will lead to Rayleigh backscatter, Raman extinction leads to spontaneous Raman scattering, absorption can lead to fluorescence emission or thermal heating of the molecule, and Mie extinction is defined primarily in terms of the scattering coefficient. Under idealized conditions, the scattering processes can be related directly to the value of the attenuation processes. However, if several complex optical processes occur simultaneously, such as in atmospheric propagation, the attenuation and scattering processes are not directly linked via a simple analytical equation. In this case, independent measurements of the scattering coefficient and the extinction coefficient have to be made, or approximation formulas are used to relate the two coefficients.^{4,5}

Molecular Absorption

The absorption of optical radiation by molecules in the atmosphere is primarily associated with individual optical absorption transitions between the allowed quantized energy levels of the molecule. The energy levels of a molecule can usually be separated into those associated with rotational, vibrational, or electronic energy states. Absorption transitions between the rotational levels occur in the far-IR and microwave spectral region, transitions between vibrational levels occur in the near-IR (2 to 20 μm wavelength), and electronic transitions generally occur in the UV-visible region (0.3 to 0.7 μm). Transitions can occur which combine several of these categories, such as rotational-vibrational transitions or electronic-vibrational-rotational transitions.

Some of the most distinctive and identifiable absorption lines of many atmospheric molecules are the rotational-vibrational optical absorption lines in the infrared spectral region. These lines are often clustered together into vibrational bands according to the allowed quantum transitions of the molecule. In many cases, the individual lines are distinct and can be resolved if the spectral resolution of the measuring instrument is fine enough (i.e., $<0.1 \text{ cm}^{-1}$). An example of such a region is the absorption feature near 2.04 μm in Fig. 1 which is actually composed of individual absorption lines if viewed under higher spectral resolution. Figure 10 is a computed high-resolution expansion of the atmospheric spectrum of Fig. 1 near 2.04 μm over a path length of 1800 m which shows these individual lines. In this case, the individual lines are well-separated and appear like a “picket fence” spectrum showing gaps between the absorption lines. Many of the atmospheric gaseous molecules listed earlier in Table 1 have similar spectral structure. These gases are relatively lightweight and have few (less than 5 or 6) atoms per molecule so that their moments of inertia are relatively small. The resulting energy spacing between the allowed rotational-vibrational absorption transitions is large and well-separated in wavelength.

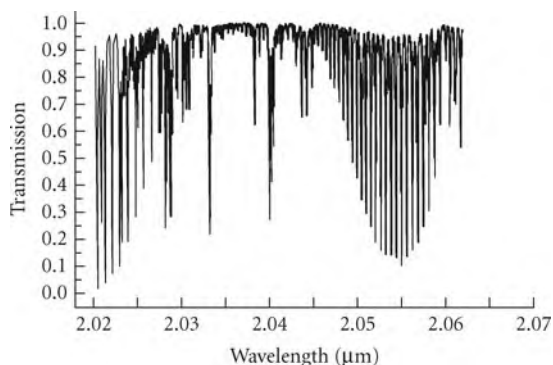


FIGURE 10 High-resolution transmission spectrum of the atmosphere for a horizontal path of 1800 m (similar to that in Fig. 1) for the spectral region near 2.04 μm . The individual rotational absorption lines due to CO_2 and H_2O are easily observed.

In other spectral regions, however, the individual lines overlap or are so strong or saturated that the transmission spectrum displays only broad spectral features; an example of such a spectral region is the strong absorption seen near 2.7 μm , 5 to 8 μm , and beyond 13 μm in Fig. 1. Finally, the molecular absorption observed in the UV is often due to optical transitions to an electronic energy level, molecular energy continuum or a predissociation energy level. In some cases, such as that for O_3 or SO_2 , this results in broad absorption bands that extend throughout the UV region (250 to 350 nm).

More complex, heavier molecules, such as benzene or chlorofluorocarbons, have absorption spectra which are blended or merged together into band spectra due to the complexity and overlap of the rotational-vibrational transitions of these molecules. Figure 11 shows a transmission spectrum of Freon-12 (CCl_2F_2) which has a complex spectrum near 11 μm . As seen, the individual rotational lines are merged into a band spectrum. The band spectrum is unique for each gas and can be used to identify the chemical composition of the gas. Heavy molecules in the atmosphere are not normally part of the natural atmosphere and are usually the result of pollution or gaseous plumes injected into the atmosphere.

The overall transmission or absorption of the atmosphere due to an individual molecular absorption line can be given quantitatively as^{1,2}

$$\kappa_a(\lambda)N(x, t) = Sg(\nu - \nu_0)NP_a \quad (3)$$

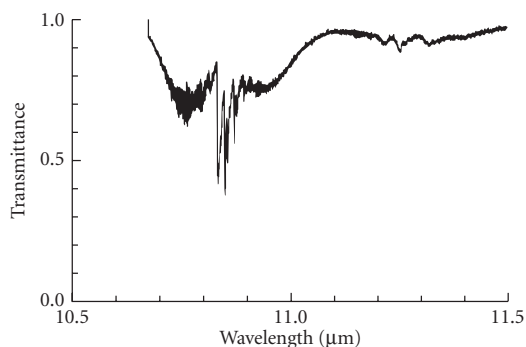


FIGURE 11 High-resolution absorption spectrum of Freon-12 gas showing complex band structure typical of heavy, complex molecules in the atmosphere (path of 100 m with 1 ppm concentration).

where S is the molecular transition line intensity (units of $\text{cm}/\text{molecule}$), $g(\nu - \nu_0)$ is the normalized lineshape function (units of cm or $1/\text{cm}^{-1}$), N is the number of molecules of absorbing species per cm^3 per atm, and P_a is the partial pressure of the absorbing gas in atm. The value of N is equal to the value of Loschmidt's number, which is 2.479×10^{19} molecules cm^{-3} atm $^{-1}$ at a temperature of 296 K; the value of N is inversely proportional to temperature due to the change in gas concentration as a function of temperature for 1 atm of pressure. As will be seen later, the definition of S as given in Eq. (3) is that used in the HITRAN database.^{11–13} In this case, S contains the Boltzmann population factor and isotope fraction (natural abundance) as well as the stimulated emission term due to finite population in the upper energy states of the molecule.^{11,12} In Eq. (3), $Sg(\nu - \nu_0)$ is the absorption cross section per molecule ($\text{cm}^2/\text{molecule}$) and NP_a is the number of absorbing molecules in units of molecules/ cm^3 .

The lineshape function can be described by several different models. The two most prevalent are the Lorentzian lineshape associated with pressure broadening and the Gaussian lineshape associated with Doppler broadening which becomes important at elevated temperatures or low pressures.

The Lorentzian/pressure-broadened profile is given by

$$g_L(\nu - \nu_0) = (\gamma_p/\pi) \left[(\nu - \nu_0)^2 + \gamma_p^2 \right]^{-1} \quad (4)$$

where γ_p is the pressure-broadened half-width at half-maximum (HWHM) in wave numbers (cm^{-1}). The pressure-broadened half-width is obtained from the air-broadened half-width parameter g as $\gamma_p = gP_t$, where P_t is the total background atmospheric pressure. Under ambient atmospheric conditions, g is approximately 0.05 cm^{-1} (i.e., 1.5 GHz) for many molecules in the atmosphere.

It should be noted that under very low pressure conditions, where the time between collisions with other molecules is relatively long, the intrinsic radiative lifetime of the molecule will determine the lineshape profile. Under these conditions, the linewidth is called the natural linewidth. The natural linewidth of many molecules is on the order of a few MHz (i.e., approximately 0.0001 cm^{-1}) or less.

The Gaussian or Doppler line profile is expressed as

$$g_D(\nu - \nu_0) = (1/\gamma_D)(\ln 2/\pi)^{1/2} \left[-\ln 2(\nu - \nu_0)^2/\gamma_D^2 \right] \quad (5)$$

where γ_D is the Doppler linewidth (HWHM in cm^{-1}) given by

$$\gamma_D = (\nu_0/c) [2RT \ln 2/M]^{1/2} \quad (6)$$

where R is the gas constant, T is the temperature in Kelvin, and M is the molecular weight of the molecule.

The value for the lineshape at the peak (line center) is equal to $1/(\pi\gamma_p) = 0.318/\gamma_p$ for the pressure-broadened case. For the Doppler peak, the maximum value is $(\ln 2/\pi)^{1/2}/\gamma_D = 0.469/\gamma_D$. Under ambient atmospheric conditions, the Doppler linewidth is usually much smaller than the pressure-broadened linewidth.

For those cases where both Lorentzian and Doppler broadening are present in approximately equal amounts, a convolution of the Doppler and Lorentzian profile must be used. This convolution of a Doppler and Lorentzian is called a Voigt profile and involves a double integral for an exact calculation. Fortunately, several numerical approximations are available for the computation of the Voigt profile and lineshape parameters.^{14–16} The Voigt profile is important in the spectroscopy of molecules in the upper atmosphere where the ambient pressure is low and the Doppler and pressure-broadened linewidths are of the same order of magnitude. Recent advances in remote-sensing experiments have suggested that further refinements to the description of the line shape in the atmosphere are required. These phenomena include line coupling, speed-dependent corrections, collision-induced narrowing, and other effects on the lineshape.¹⁷

Finally, the large number of transition lines of water vapor and other gases in the atmosphere can produce a significant level of background “quasi-continuum” absorption in the atmosphere. This phenomenon is primarily due to the additive contribution from the wings of the absorption lines even at wavelengths far removed ($>25 \text{ cm}^{-1}$) from the line centers. Such an effect has been studied by Burch and by Clough et al. for water vapor due to strong self-broadening interactions.¹⁸ Figure 12

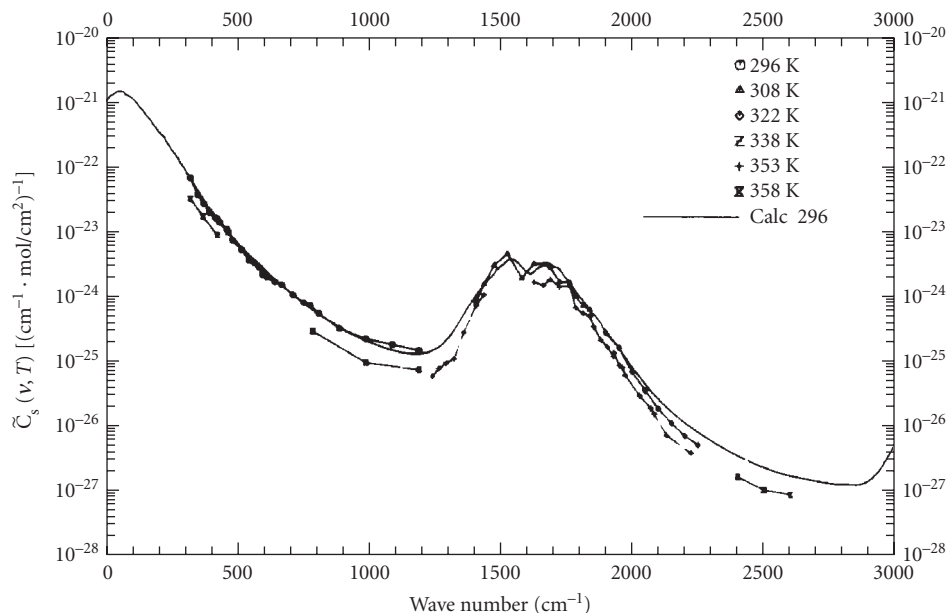


FIGURE 12 Self-density absorption continuum values C_s for water vapor as a function of wave number. The experimental values were measured by Burch. (From Ref. 3.)

shows a plot of the relative continuum coefficient for water vapor as a function of wave number. Good agreement with the experimental data and model calculations is shown. Models for water vapor and nitrogen continuum absorption are contained within many of the major atmospheric transmission programs (such as FASCODE). The typical value for the continuum absorption is negligible in the visible to the near-IR, but can be significant at wavelengths in the range of 5 to 20 μm .

Molecular Rayleigh Scattering

Rayleigh scattering is elastic scattering of the optical radiation due to the displacement of the weakly bound electronic cloud surrounding the gaseous molecule which is perturbed by the incoming electromagnetic (optical) field. This phenomenon is associated with optical scattering where the wavelength of light is much larger than the physical size of the scatterers (i.e., atmospheric molecules). Rayleigh scattering, which makes the color of the sky blue and the setting or rising sun red, was first described by Lord Rayleigh in 1871. The Rayleigh differential scattering cross section for polarized, monochromatic light is given by⁵

$$d\sigma_R/d\Omega = [\pi^2(n^2 - 1^2)/N^2\lambda^4][\cos^2\phi\cos^2\theta + \sin^2\phi] \quad (7)$$

where n is the index of refraction of the atmosphere, N is the density of molecules, λ is the wavelength of the optical radiation, and ϕ and θ are the spherical coordinate angles of the scattered polarized light referenced to the direction of the incident light. As seen from Eq. (7), shorter-wavelength light (i.e., blue) is more strongly scattered out from a propagating beam than the longer wavelengths (i.e., red), which is consistent with the preceding comments regarding the color of the sky or the sunset. A typical value for $d\sigma_R/d\Omega$, at a wavelength of 700 nm in the atmosphere (STP) is approximately $2 \times 10^{-28} \text{ cm}^2 \text{ sr}^{-1}$.³ This value depends upon the molecule and has been tabulated for many of the major gases in the atmosphere.¹⁴⁻¹⁹

The total Rayleigh scattering cross section can be determined from Eq. (7) by integrating over 4π steradians to yield

$$\sigma_R(\text{total}) = [8/3][\pi^2(n^2 - 1)^2/N^2\lambda^4] \quad (8)$$

At sea level (and room temperature, $T = 296$ K) where $N = 2.5 \times 10^{19}$ molecules/cm³, Eq. (8) can be multiplied by N to yield the total Rayleigh scattering extinction coefficient as

$$\kappa_R(\lambda)N(x, t) = N\sigma_R(\text{total}) = 1.18 \times 10^{-8} [550 \text{ nm}/\lambda(\text{nm})]^4 \text{ cm}^{-1} \quad (9)$$

The neglect of the effect of dispersion of the atmosphere (variation of the index of refraction n with wavelength) results in an error of less than 3 percent in Eq. (9) in the visible wavelength range.⁵

The molecular Rayleigh backscatter ($\theta = \pi$) cross section for the atmosphere has been given by Collins and Russell for polarized incident light (and received scattered light of the same polarization) as¹⁹

$$\sigma_R = 5.45 \times 10^{-28} [550 \text{ nm}/\lambda(\text{nm})]^4 \text{ cm}^2 \text{ sr}^{-1} \quad (10)$$

At sea level where $N = 2.47 \times 10^{19}$ molecules/cm³, the atmospheric volume backscatter coefficient, β_R , is thus given by

$$\beta_R = N\sigma_R = 1.39 \times 10^{-8} [550 \text{ nm}/\lambda(\text{nm})]^4 \text{ cm}^{-1} \text{ sr}^{-1} \quad (11)$$

The backscatter coefficient for the reflectivity of a laser beam due to Rayleigh backscatter is determined by multiplying β_R by the range resolution or length of the optical interaction being considered.

For unpolarized incident light, the Rayleigh scattered light has a depolarization factor δ which is the ratio of the two orthogonal polarized backscatter intensities. δ is usually defined as the ratio of the perpendicular and parallel polarization components measured relative to the direction of the incident polarization. Values of δ depend upon the anisotropy of the molecules or scatters, and typical values range from 0.02 to 0.11.²⁰ Depolarization also occurs for multiple scattering and is of considerable interest in laser or optical transmission through dense aerosols or clouds.²¹ The depolarization factor can sometimes be used to determine the physical and chemical composition of the cloud constituents, such as the relative ratio of water vapor or ice crystals in a cloud.

Mie Scattering: Aerosols, Water Droplets, and Ice Particles

Mie scattering is similar to Rayleigh scattering, except the size of the scattering sites is on the same order of magnitude as the wavelength of the incident light, and is, thus, due to aerosols and fine particulates in the atmosphere. The scattered radiation is the same wavelength as the incident light but experiences a more complex functional dependence upon the interplay of the optical wavelength and particle size distribution than that seen for Rayleigh scattering.

In 1908, Mie investigated the scattering of light by dielectric spheres of size comparable to the wavelength of the incident light.²² His analysis indicated the clear asymmetry between the forward and backward directions, where for large particle sizes the forward-directed scattering dominates. Complete treatments of Mie scattering can be found in several excellent works by Deirmendjian and others, which take into account the complex index of refraction and size distribution of the particles.^{23,24} These calculations are also influenced by the asymmetry of the aerosols or particulates which may not be spherical in shape.

The effect of Mie scattering in the atmosphere can be described as in the following figures. Figure 13 shows the aerosol Mie extinction coefficient as a function of wavelength for several atmospheric models, along with a typical Rayleigh scattering curve for comparison.²⁵ Figure 14 shows similar values for the volume Mie backscatter coefficient as a function of wavelength.²⁵ Extinction and backscatter coefficient values are highly dependent upon the wavelength and particulate composition.

Figures 15 and 16 show the calculated extinction coefficient for the rural and maritime aerosol models described in Sec. 3.3 as a function of relative humidity and wavelength.³ Significant changes in the backscatter can be produced by relatively small changes in the humidity.

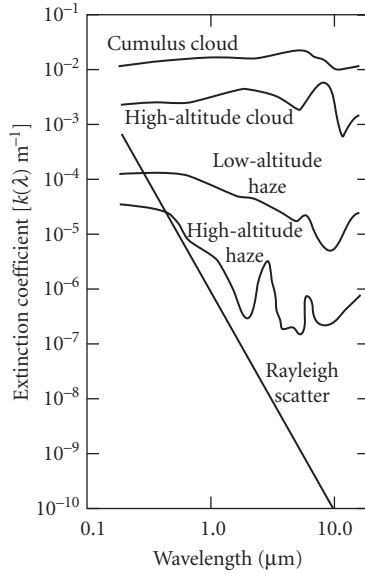


FIGURE 13 Aerosol extinction coefficient as a function of wavelength. (From Measures, Ref. 5.)

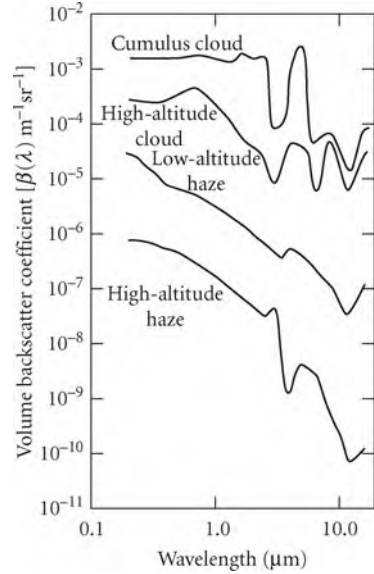


FIGURE 14 Aerosol volume backscattering coefficient as a function of wavelength. (From Measures, Ref. 5.)

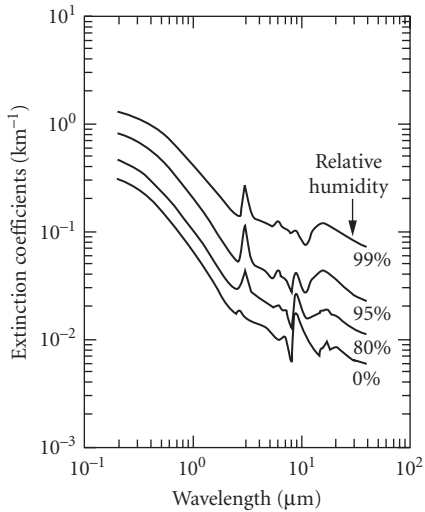


FIGURE 15 Extinction coefficients vs. wavelength for the rural aerosol model for different relative humidities and constant number density of particles.

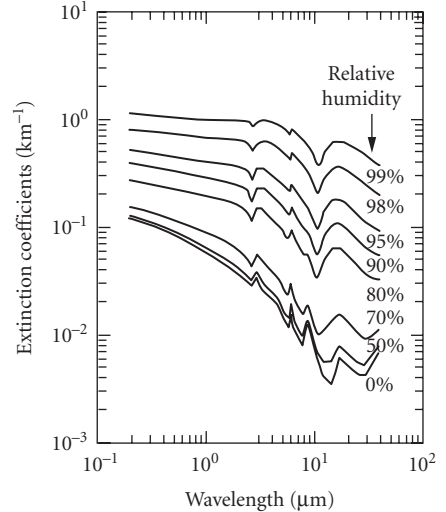


FIGURE 16 Extinction coefficients vs. wavelength for the maritime aerosol model for different relative humidities and constant number density of particles.

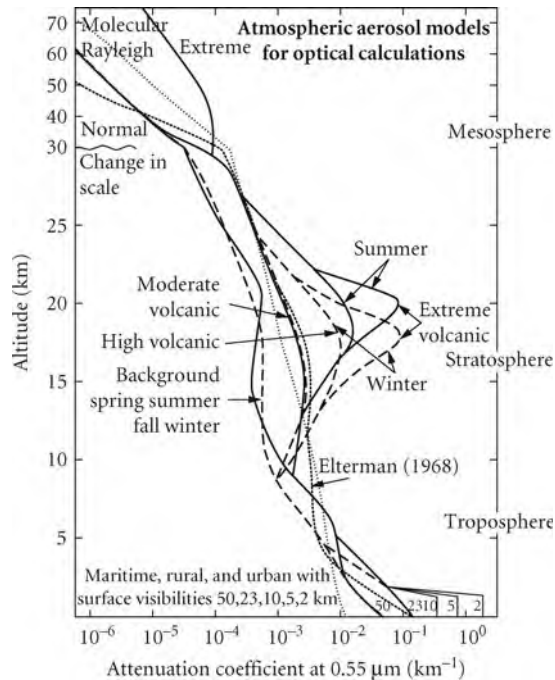


FIGURE 17 The vertical distribution of the aerosol extinction coefficient (at 0.55- μm wavelength) for the different atmospheric models. Also shown for comparison are the Rayleigh profile (dotted line). Between 2 and 30 km, where the distinction on a seasonal basis is made, the spring-summer conditions are indicated with a solid line and fall-winter conditions are indicated by a dashed line. (From Ref. 3.)

The extinction coefficient is also a function of altitude, following the dependence of the composition of the aerosols. Figure 17 shows an atmospheric aerosol extinction model as a function of altitude for a wavelength of 0.55 μm .^{3,10} The influence of the visibility (in km) at ground level dominates the extinction value at the lower altitudes and the composition and density of volcanic particulate dominates the upper altitude regions. The dependence of the extinction on the volcanic composition at the upper altitudes is shown in Fig. 18 which shows these values as a function of wavelength and of composition.^{3,10}

The variation of the backscatter coefficient as a function of altitude is shown in Fig. 19 which displays atmospheric backscatter data obtained by McCormick using a 1.06- μm Nd:YAG Lidar.²⁶ The boundary layer aerosols dominate at the lower levels and the decrease in the atmospheric particulate density determines the overall slope with altitude. Of interest is the increased value near 20 km due to the presence of volcanic aerosols in the atmosphere due to the eruption of Mt. Pinatubo in 1991.

Molecular Emission and Thermal Spectral Radiance

The same optical molecular transitions that cause absorption also emit light when they are thermally excited. Since the molecules have a finite temperature T , they will act as blackbody radiators with optical emission given by the Planck radiation law. The allowed transitions of the molecules

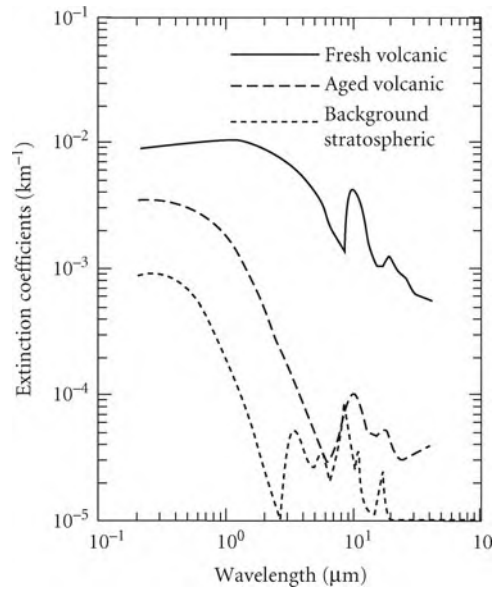


FIGURE 18 Extinction coefficients for the different stratospheric aerosol models (background, volcanic, and fresh volcanic). The extinction coefficients have been normalized to values around peak levels for these models.

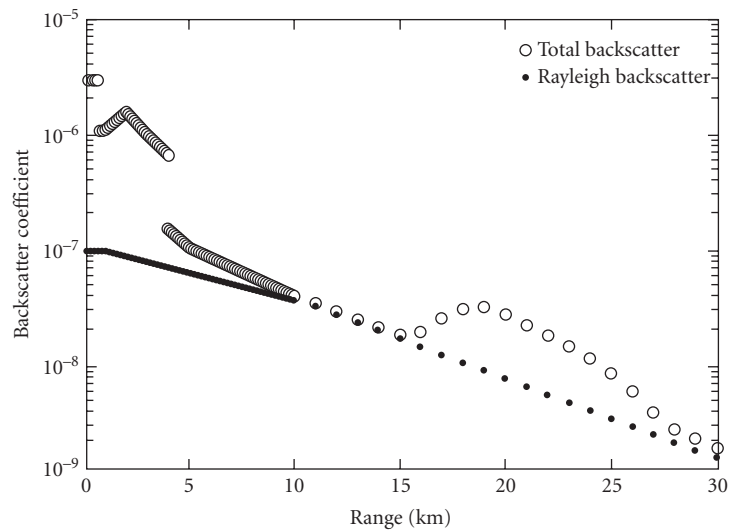


FIGURE 19 1.06- μm lidar backscatter coefficient measurements as a function of vertical altitude. (From McCormick and Winker, Ref. 26.)

will modify the radiance distribution of the radiation due to emission of the radiation according to the thermal distribution of the population within the energy levels of the molecule; it should be noted that the Boltzmann thermal population distribution is essentially the same as that which is described by the Planck radiation law for local thermodynamic equilibrium conditions. As such, the molecular emission spectrum of the radiation is similar to that for absorption. The thermal radiance from the clear atmosphere involves the calculation of the blackbody radiation emitted by each elemental volume of air multiplied by the absorption spectral distribution of the molecular absorption lines, $\kappa_a(s)$ and then this emission spectrum is attenuated by the rest of the atmosphere as the emission propagates toward the viewer. This may be expressed as

$$I_\nu = \int_0^s \kappa_a(s) P_\nu(s) \exp\left[-\int_0^s \kappa_a(s') ds'\right] ds \quad (12)$$

where the exponential term is Beer's law, and $P_\nu(s)$ is the Planck function given by

$$P_\nu(s) = 2h\nu^3/[c^2 \exp([h\nu/kT(s)] - 1)] \quad (13)$$

In these equations, s is the distance from the receiver along the optical propagation path, ν is the optical frequency, h is Planck's constant, c is the speed of light, k is Boltzmann's constant, and $T(s)$ is the temperature at position s along the path. As seen in Eq. (12), each volume element emits thermal radiation of $\kappa_a(s)P_\nu(s)$, which is then attenuated by Beer's law. The total emission spectral density is obtained by summing or integrating over all the emission volume elements and calculating the appropriate absorption along the optical path for each element.

As an example, Fig. 20 shows a plot of the spectral radiance measured on a clear day with 1 cm^{-1} spectral resolution. Note that the regions of strong absorption produce more radiance as the foregoing equation suggests, and that regions of little absorption correspond to little radiance. In the 800- to 1200-wave number spectral region (i.e., 8.3- to 12.5- μm wavelength region), the radiance is relatively low. This is consistent with the fact that the spectral region from 8 to 12 μm is a transmission window of the atmosphere with relatively little absorption of radiation.

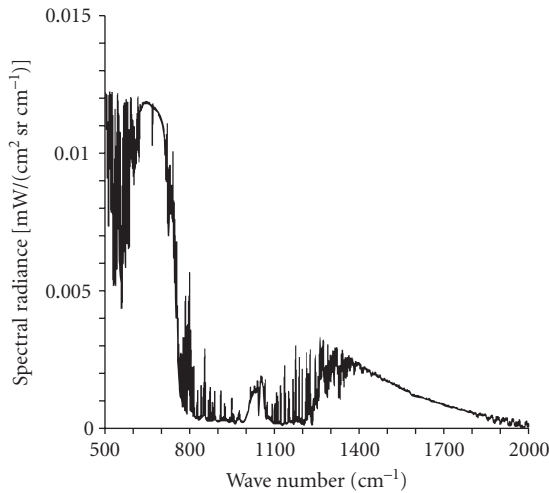


FIGURE 20 Spectral radiance (molecular thermal emission) measured on a clear day showing the relatively low value of radiance near 1000 cm^{-1} (i.e., 10- μm wavelength). (Provided by Churnside.)

Surface Reflectivity and Multiple Scattering

The spectral intensity of naturally occurring light at the earth's surface is primarily due to the incident intensity from the sun in the visible to mid-IR wavelength range, and due to thermal emission from the atmosphere and background radiance in the mid-IR. In both cases, the optical radiation is affected by the reflectance characteristics of the clouds and surface layers. For instance, the fraction of light that falls on the earth's surface and is reflected back into the atmosphere is dependent upon the reflectivity of the surface, the incident solar radiation (polarization and spectral density), and the absorption of the atmosphere.

The reflectivity of a surface, such as the earth's surface, is often characterized using the bidirectional reflectance function (BRDF). This function accounts for the nonspecular reflection of light from common rough surfaces and describes the changes in the reflectivity of a surface as a function of the angle which the incident beam makes with the surface. In addition, the reflectivity of a surface is usually a function of wavelength. This latter effect can be seen in Fig. 21 which shows the reflectance of several common substances for normal incident radiation.² As seen in Fig. 21, the reflectivity of these surfaces is a strong function of wavelength.

The effect of multiple scattering sometimes must be considered when the scattered light undergoes more than one scatter event, and is rescattered on other particles or molecules. These multiple scattering events increase with increasing optical thickness and produce deviations from the Beer-Lambert law. Extensive analyses of the scattering processes for multiple scattering have been conducted and have shown some success in predicting the overall penetration of light through a thick dense cloud. Different computational techniques have been used including the Gauss-Seidel Iterative Method, Layer Adding Method, and Monte-Carlo Techniques.^{3,5}

Additional Optical Interactions

In some optical experiments on the atmosphere, a laser beam is used to excite the molecules in the atmosphere to emit inelastic radiation. Two important inelastic optical processes for atmospheric remote sensing are fluorescence and Raman scattering.^{5,27}

For the case of laser-induced fluorescence, the molecules are excited to an upper energy state and the reemitted photons are detected. In these experiments, the inelastic fluorescence emission is red-shifted in wavelength and can be distinguished in wavelength from the elastic scattered Rayleigh or Mie backscatter. Laser-induced fluorescence is mostly used in the UV to visible spectral region; collisional quenching is quite high in the infrared so that the fluorescence efficiency is higher in the UV-visible than in the IR. Laser-induced fluorescence is sometimes reduced by saturation effects due to stimulated emission from the upper energy levels. However, in those cases where laser-induced fluorescence can be successfully used, it is one of the most sensitive optical techniques for the detection of atomic or molecular species in the atmosphere.

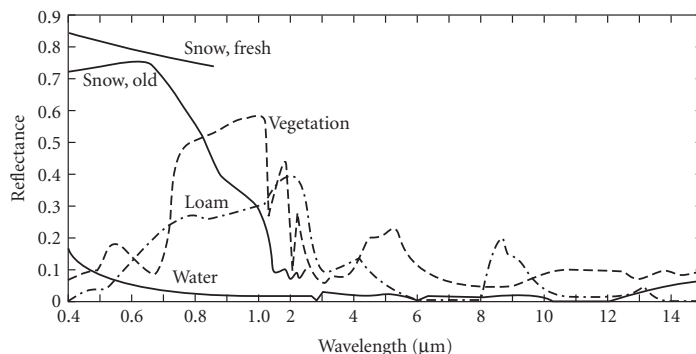


FIGURE 21 Typical reflectance of water surface, snow, dry soil, and vegetation. (From Ref. 2.)

Laser-induced Raman scattering of the atmosphere is a useful probe of the composition and temperature of concentrated species in the atmosphere. The Raman-shifted emitted light is often weak due to the relatively small cross section for Raman scattering. However, for those cases where the distance is short from the laser to the measurement cloud, or where the concentration of the species is high, it offers significant information concerning the composition of the gaseous atmosphere.

The use of an intense laser beam can also bring about nonlinear optical interactions as the laser beam propagates through the atmosphere. The most important of these are stimulated Raman scattering, thermal blooming, dielectric breakdown, and harmonic conversion. Each of these processes requires a tightly focused laser beam to initiate the nonlinear optical process.^{28,29}

3.5 PREDICTION OF ATMOSPHERIC OPTICAL TRANSMISSION: COMPUTER PROGRAMS AND DATABASES

During the past three decades, several computer programs and databases have been developed which are very useful for the determination of the optical properties of the atmosphere. Many of these are based upon programs originally developed at the U.S. Air Force Cambridge Research Laboratories. The latest versions of these programs and databases are the HITRAN database,¹³ FASCODE computer program,^{30–32} and the LOWTRAN or MODTRAN computer code.^{33,34} In addition, several PC (personal computer) versions of these database/computer programs have recently become available so that the user can easily use these computational aids.

Molecular Absorption Line Database: HITRAN

The HITRAN database contains optical spectral data on most of the major molecules contributing to absorption or radiance in the atmosphere; details of HITRAN are covered in several recent journal articles.^{11–13} The 40 molecules contained in HITRAN are given in Table 1, and cover over a million individual absorption lines in the spectral range from 0.000001 cm^{-1} to $25,233\text{ cm}^{-1}$ (i.e., 0.3963 to $10^{10}\text{ }\mu\text{m}$). A free copy of this database can be obtained by filling out a request form in the HITRAN Web site.³⁵ Each line in the database contains 19 molecular data items that consist of the molecule formula code, isotopologue type, transition frequency (cm^{-1}), line intensity S in $\text{cm}/\text{molecule}$, Einstein A -coefficient (s^{-1}), air-broadened half-width ($\text{cm}^{-1}/\text{atm}$), self-broadened half-width ($\text{cm}^{-1}/\text{atm}$), lower state energy (cm^{-1}), temperature coefficient for air-broadened linewidth, air-pressure induced line shift ($\text{cm}^{-1}\text{ atm}^{-1}$), upper-state global quanta index, lower-state global quanta index, upper- and lower-state quanta, uncertainty codes, reference numbers, a flag for line coupling if necessary, and upper- and lower-level statistical weights. The density of the lines of the 2004 HITRAN database is shown in Fig. 22. The recently released 2008 HITRAN database has been expanded to 42 molecules.

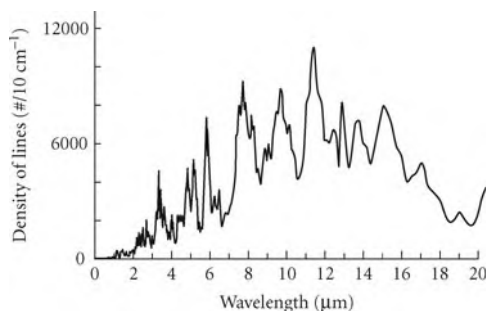


FIGURE 22 Density of absorption lines in HITRAN 2004 spectral database.

M	I	wn (cm ⁻¹)	S (T)	R	g (T)	E''	Q'	Q''
NH ₃	1	5000.04530	5.161E-23	2.33E-5	.0599	244.09880		
CO ₂	1	5000.05653	3.823E-27	5.99E-6	.0680	2783.34814		P 33
CO ₂	3	5000.06953	1.766E-26	1.46E-6	.0657	975.59039		P 51
NH ₃	1	5000.20320	3.096E-23	1.54E-6	.1059	16.20000	2 0-1	s 1 1 0 s
CO ₂	1	5000.21444	1.358E-25	8.01E-6	.0627	2240.23755		R 63
H ₂ O	1	5000.22500	2.160E-24	1.86E-4	.0088	2358.30396	14 014	15 015
NH ₃	1	5000.32760	2.854E-23	1.42E-6	.1059	17.00000	2 0-1	a 1 1 0 a
CO ₂	2	5000.33471	1.524E-26	1.00E-5	.0800	1398.12122		R 8
CO ₂	1	5000.34883	3.939E-25	8.02E-6	.0640	2004.66040		R 58
NH ₃	1	5000.39930	2.100E-23	1.18E-5	.0599	289.09781		
N ₂ O	1	5000.42800	2.990E-23	6.82E-7	.0719	316.67001		P 27
CO ₂	2	5000.47621	2.938E-25	4.38E-6	.0739	754.88672		P 16
CO ₂	1	5000.48083	4.762E-22	9.21E-6	.0677	464.17169		R 34
CO ₂	2	5000.48726	5.469E-27	4.31E-6	.0681	1710.65881		P 32
H ₂ O	1	5000.54070	6.959E-27	1.10E-8	.0556	1411.61206	9 5 4	8 6 3
NH ₃	1	5000.65640	2.221E-23	1.54E-5	.0599	333.09299		
H ₂ O	4	5000.75000	4.629E-27	3.10E-6	.0856	809.39502	8 3 5	8 4 4
CO ₂	1	5000.81774	8.595E-26	8.01E-6	.0623	2340.73218		R 65
H ₂ O	4	5000.83600	4.469E-26	3.43E-6	.0922	308.61700	5 0 5	6 1 6
CO ₂	2	5000.85619	7.150E-24	4.82E-6	.0762	60.87380		R 12
NH ₃	1	5000.88490	4.193E-23	1.4E0				

FIGURE 23 Example of data contained within the HITRAN database showing individual absorption lines, frequency, line intensity, and other spectroscopic parameters.

Figure 23 shows an output from a computer program that was used to search the HITRAN database and display some of the pertinent information.³⁶ The data in HITRAN are in sequential order by transition frequency in wave numbers, and list the molecular name, isotope, absorption line strength S , transition probability R , air-pressure-broadened linewidth γ , lower energy state E'' , and upper/lower quanta for the different molecules and isotopic species in the atmosphere.

Line-by-Line Transmission Program: FASCODE

FASCODE is a large, sophisticated computer program that uses molecular absorption equations (similar to those under "Molecular Absorption") and the HITRAN database to calculate the high-resolution spectra of the atmosphere. It uses efficient algorithms to speed the computations of the spectral transmission, emission, and radiance of the atmosphere at a spectral resolution that can be set to better than the natural linewidth,³² and includes the effects of Rayleigh and aerosol scattering and the continuum molecular extinction. FASCODE also calculates the radiance and transmittance of atmospheric slant paths, and can calculate the integrated transmittance through the atmosphere from the ground up to higher altitudes. Voigt lineshape profiles are also used to handle the transition from pressure-broadened lineshapes near ground level to the Doppler-dominated lineshapes at very high altitudes. Several representative models of the atmosphere are contained within FASCODE, so that the user can specify different seasonal and geographical models. Figure 24 shows a sample output generated from data produced from the FASCODE program and a comparison with experimental data obtained by J. Dowling at NRL.^{30,31} As can be seen, the agreement is very good. There are also several other line-by-line codes available for specialized applications; examples include GENLN2 developed by D.P. Edwards at NCAR (National Center for Atmospheric Research/Boulder), and LBLRTM (Atmospheric and Environmental Research, Inc.).

Broadband Transmission: LOWTRAN and MODTRAN

The LOWTRAN computer program does not use the HITRAN database directly, but uses absorption band models based on degrading spectral calculations based on HITRAN to calculate the moderate resolution (20 cm⁻¹) transmission spectrum of the atmosphere. LOWTRAN uses extensive

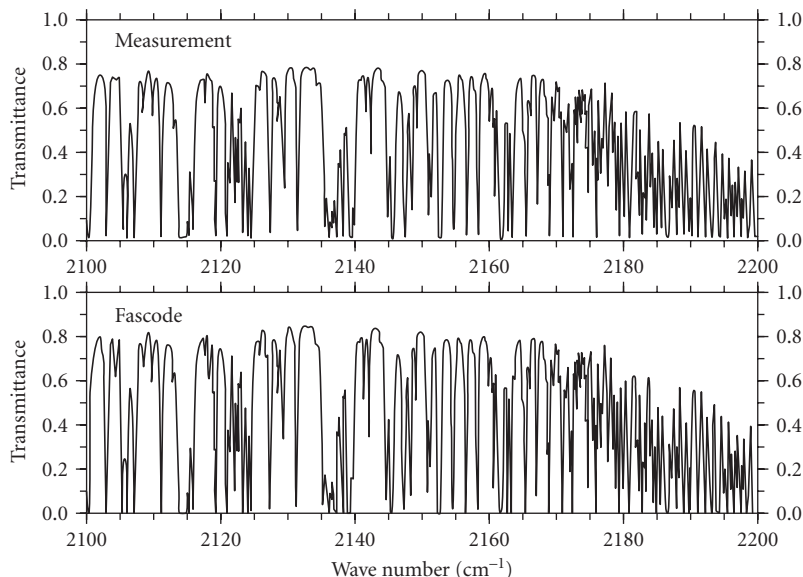


FIGURE 24 Comparison of an FASCOD2 transmittance calculation with an experimental atmospheric measurement (from NRL) over a 6.4-km path at the ground. (*Courtesy of Clough, Ref. 30.*)

band-model calculations to speed up the computations, and provides an accurate and rapid means of estimating the transmittance and background radiance of the earth's atmosphere over the spectral interval of 350 cm^{-1} to $40,000\text{ cm}^{-1}$ (i.e., 250-nm–28- μm wavelength). The spectral range of the LOWTRAN program extends into the UV. In the LOWTRAN program, the total transmittance at a given wavelength is given as the product of the transmittances due to molecular band absorption, molecular scattering, aerosol extinction, and molecular continuum absorption. The molecular band absorption is composed of four components of water vapor, ozone, nitric acid, and the uniformly mixed gases (CO_2 , N_2O , CH_4 , CO , O_2 , and N_2).

The latest version of LOWTRAN(7) contains models treating solar and lunar scattered radiation, spherical refractive geometry, slant-path geometry, wind-dependent maritime aerosols, vertical structure aerosols, standard seasonal and geographic atmospheric models (e.g., mid-latitude summer), cirrus cloud model, and a rain model.³⁴ As an example, Fig. 25 shows a ground-level solar radiance model used by LOWTRAN, and Fig. 26 shows an example of a rain-rate model and its effect upon the transmission of the atmosphere as a function of rain rate in mm of water per hour.³⁴

Extensive experimental measurements have been made to verify LOWTRAN calculations. Figure 27 shows a composite plot of the LOWTRAN-predicted transmittance and experimental data for a path length of 1.3 km at sea level.³ As can be seen, the agreement is quite good. It is estimated that the LOWTRAN calculations are good to about 10 percent.³ It should be added that the molecular absorption portion of the preceding LOWTRAN (moderate-resolution) spectra can also be generated using the high-resolution FASCOD/HITRAN program and then spectrally smoothing (i.e., degrading) the spectra to match that of the LOWTRAN spectra.

The most recent extension of the LOWTRAN program is the MODTRAN program. MODTRAN is similar to LOWTRAN but has increased spectral resolution. At present, the resolution for the latest version of MODTRAN, called MODTRAN(5), can be specified by the user between 0.1 and 20 cm^{-1} .

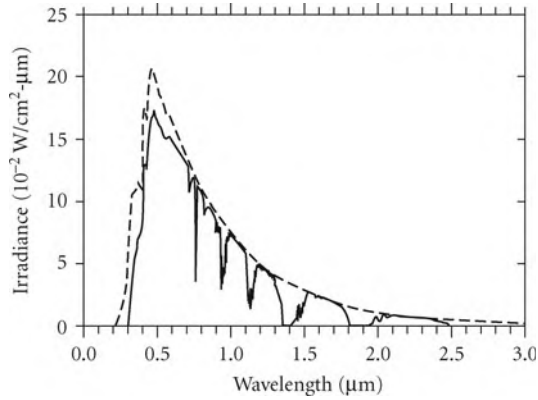


FIGURE 25 Solar radiance model (dashed line) and directly transmitted solar irradiance (solid line) for a vertical path, from the ground (U.S. standard 1962 model, no aerosol extinction) as used by the LOWTRAN program.

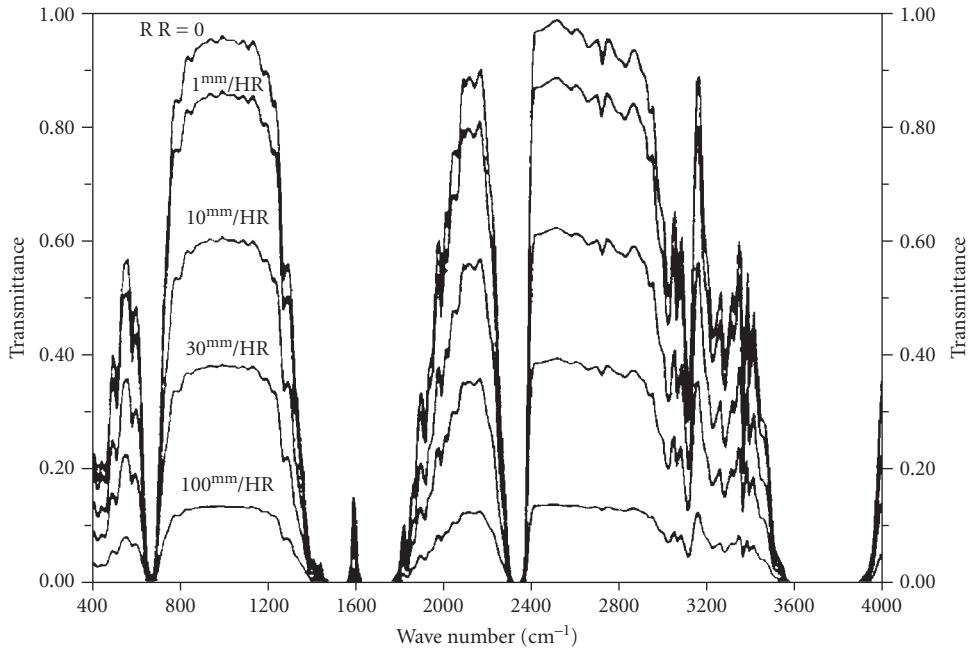


FIGURE 26 Atmospheric transmittance for different rain rates and for spectral frequencies from 400 to 4000 cm^{-1} . The measurement path is 300 m at the surface with $T = T_{\text{dew}} = 10^\circ\text{C}$, with a meteorological range of 23 km in the absence of rain.

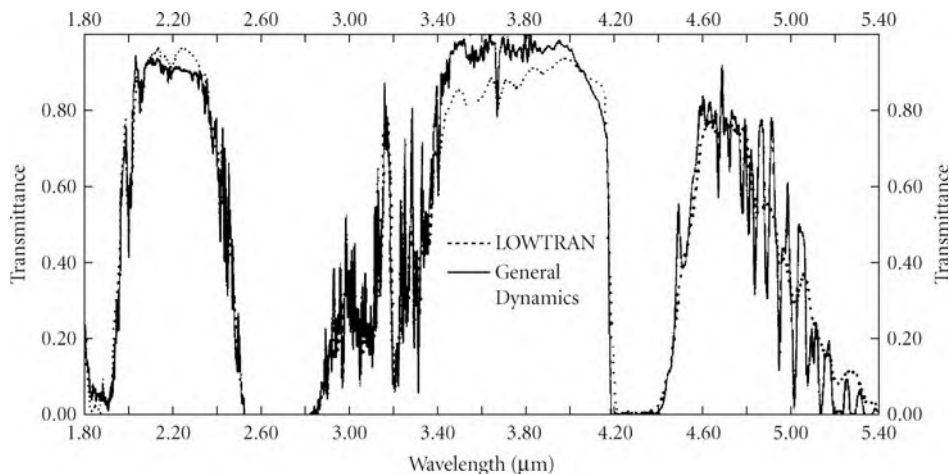


FIGURE 27 Comparison between LOWTRAN predicted spectrum and General Dynamics atmospheric measurements; range = 1.3 km at sea levels. (From Ref. 3.)

Programs and Databases for Use on Personal Computers

The preceding databases and computer programs have been converted or modified to run on different kinds of personal computers.^{35,36} The HITRAN database has been available on CD-ROMs for the past decade, but is now available via the internet. Several related programs are available, ranging from a complete copy of the FASCODE and LOWTRAN programs³⁵ to a simpler molecular transmission program of the atmosphere.³⁶ These programs calculate the transmission spectrum of the atmosphere and some show the overlay spectra of known laser lines. As an example, Fig. 28 shows the transmission spectrum produced by the HITRAN-PC program³⁶ for a horizontal path of 300 m (U.S. Standard Atmosphere) over the wavelength range of 250 nm ($40,000\text{ cm}^{-1}$) to $20\text{ }\mu\text{m}$ (500 cm^{-1}); the transmission spectrum includes water, nitrogen, and CO_2 continuum and urban aerosol attenuation, and was smoothed to a spectral resolution of 1 cm^{-1} to better display the overall transmission features of the atmosphere.

While these PC versions of the HITRAN database and transmission programs have become available only recently, they have already made a significant impact in the fields of atmospheric optics and optical remote sensing. They allow quick and easy access to atmospheric spectral data which was previously only available on a mainframe computer. It should be added that other computer programs are available which allow one to add or subtract different spectra generated by these HITRAN-based programs, from spectroscopic instrumentation such as FT-IR spectrometers or from other IR gas spectra databases. In the latter case, for example, the U.S. National Institute of Standards and Technology (NIST) has compiled a computer database of the qualitative IR absorption spectra of over 5200 different gases (toxic and other hydrocarbon compounds) with a spectral resolution of 4 cm^{-1} .³⁷ The Pacific Northwest National Laboratory also offers absorption cross-section files of numerous gases.³⁸ In addition, higher-resolution quantitative spectra for a limited group of gases can be obtained from several commercial companies.³⁹

3.6 ATMOSPHERIC OPTICAL TURBULENCE

The most familiar effects of refractive turbulence in the atmosphere are the twinkling of stars and the shimmering of the horizon on a hot day. The first of these is a random fluctuation of the amplitude of light also known as scintillation. The second is a random fluctuation of the phase front that

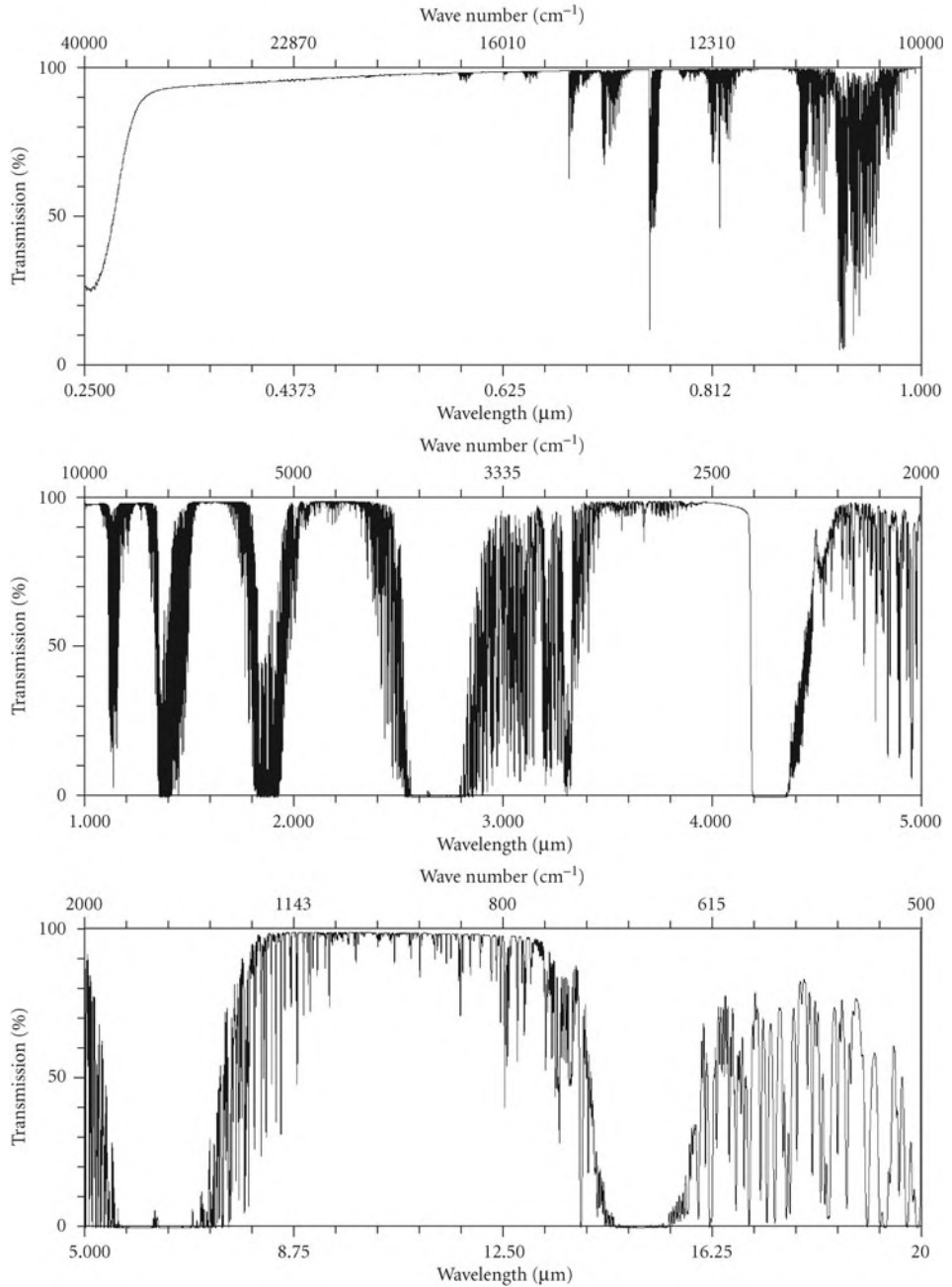


FIGURE 28 Example of generated atmospheric transmission spectrum of the atmosphere for a horizontal path of 300 m for the wavelength range from UV (250 nm or 40,000 cm^{-1}) to the IR (20 μm or 500 cm^{-1}); the spectrum includes water, nitrogen, and CO_2 continuum, urban ozone and NO_2 , and urban aerosol attenuation, and has been smoothed to a resolution of 0.5 cm^{-1} to better show the absorption and transmission windows of the atmosphere. (From Ref. 36 and D. Pluittau.)

leads to a reduction in the resolution of an image. Other effects include the wander and break-up of an optical beam. A detailed discussion of all of these effects and the implications for various applications can be found in Ref. 40.

In the visible and near-IR region of the spectrum, the fluctuations of the refractive index in the atmosphere are determined by fluctuations of the temperature. These temperature fluctuations are caused by turbulent mixing of air of different temperatures. In the far-IR region, humidity fluctuations also contribute.

Turbulence Characteristics

Refractive turbulence in the atmosphere can be characterized by three parameters. The outer scale L_0 is the length of the largest scales of turbulent eddies. The inner scale l_0 is the length of the smallest scales. For eddies in the inertial subrange (sizes between the inner and outer scale), the refractive index fluctuations are best described by the structure function. This function is defined by

$$D_n(r_1, r_2) = \langle [n(r_1) - n(r_2)]^2 \rangle \quad (14)$$

where $n(r_i)$ is the index of refraction at point r_i and the angle brackets denote an ensemble average. For homogeneous and isotropic turbulence it depends only on the distance between the two points r and is given by

$$D_n(r) = C_n^2 r^{2/3} \quad (15)$$

where C_n^2 is a measure of the strength of turbulence and is defined by this equation.

The power spectrum of turbulence is the Fourier transform of the correlation function, which is contained in the cross term of the structure function. For scales within the inertial subrange, it is given by the Kolmogorov spectrum:

$$\Phi_n(K) = 0.033 C_n^2 K^{-11/3} \quad (16)$$

For scales larger than the outer scale, the spectrum actually approaches a constant value, and the result can be approximated by the von Kármán spectrum:

$$\Phi_n(K) = 0.033 C_n^2 (K^2 + K_0^2)^{-11/6} \quad (17)$$

where K_0 is the wave number corresponding to the outer scale. For scales near the inner scale, there is a small increase over the Kolmogorov spectrum, with a large decrease at smaller scales.⁴¹ The resulting spectrum can be approximated by a rather simple function.^{42,43}

In the boundary layer (the lowest few hundred meters of the atmosphere), turbulence is generated by radiative heating and cooling of the ground. During the day, solar heating of the ground drives convective plumes. Refractive turbulence is generated by the mixing of these warm plumes with the cooler air surrounding them. At night, the ground is cooled by radiation and the cooler air near the ground is mixed with warmer air higher up by winds. A period of extremely low turbulence exists at dawn and at dusk when there is no temperature gradient in the lower atmosphere. Turbulence levels are also very low when the sky is overcast and solar heating and radiative cooling rates are low. Measured values of turbulence strength near the ground vary from less than 10^{-17} to greater than $10^{-12} \text{ m}^{-2/3}$ at heights of 2 to 2.5 m.^{44,45}

Figure 29 illustrates typical summertime values near Boulder, Colorado. This is a 24-hour plot of 15-minute averages of C_n^2 measured at a height of about 1.5 m on August 22, 1991. At night, the sky was clear, and C_n^2 was a few parts times 10^{-13} . The dawn minimum is seen as a very short period of low turbulence just after 6:00. After sunrise, C_n^2 increases rapidly to over 10^{-12} . Just before noon, cumulus clouds developed, and C_n^2 became lower with large fluctuations. At about 18:00, the clouds dissipated,

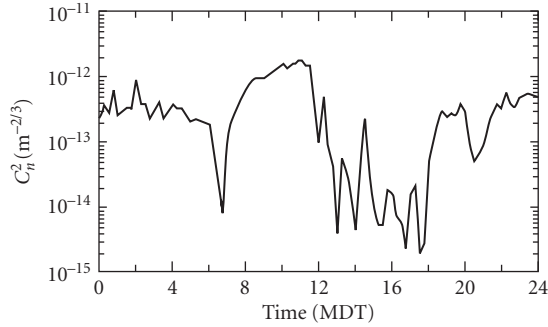


FIGURE 29 Plot of refractive-turbulence structure parameter C_n^2 for a typical summer day near Boulder, Colorado. (Courtesy G. R. Ochs, NOAA WPL.)

and turbulence levels increased. The dusk minimum is evident just after 20:00, and then turbulence strength returns to typical nighttime levels.

From a theory introduced by Monin and Obukhov,⁴⁶ a theoretical dependence of turbulence strength on height in the boundary layer above flat ground can be derived.^{47,48} During periods of convection C_n^2 decreases as the $-4/3$ power of height. At other times (night or overcast days), the power is more nearly $-2/3$. These height dependencies have been verified by a number of experiments over relatively flat terrain.^{49–53} However, values measured in mountainous regions are closer to the $-1/3$ power of height day or night.⁵⁴ Under certain conditions, the turbulence strength can be predicted from meteorological parameters and characteristics of the underlying surface.^{55–57}

Farther from the ground, no theory for the turbulence profile exists. Measurements have been made from aircraft^{44,49} and with balloons.^{58–60} Profiles of C_n^2 have also been measured remotely from the ground using acoustic sounders,^{61–63} radar,^{49,50,64–69} and optical techniques.^{59,70–73} The measurements show large variations in refractive turbulence strength. They all exhibit a sharply layered structure in which the turbulence appears in layers of the order of 100 m thick with relatively calm air in between. In some cases these layers can be associated with orographic features; that is, the turbulence can be attributed to mountain lee waves. Generally, as height increases, the turbulence decreases to a minimum value that occurs at a height of about 3 to 5 km. The turbulence level then increases to a maximum at about the tropopause (10 km). Turbulence levels decrease rapidly above the tropopause.

Model turbulence profiles have evolved from this type of measurement. Perhaps the best available model for altitudes of 3 to 20 km is the Hufnagel model:^{74,75}

$$C_n^2 = \left\{ \left[(2.2 \times 10^{-53}) H^{10} \left(\frac{W}{27} \right)^2 \right] \exp \left(-\frac{H}{1000} \right) + 10^{-16} \exp \left(-\frac{H}{1500} \right) \right\} \exp[u(H, t)] \quad (18)$$

where H is the height above sea level in meters, W is the vertical average of the square of the wind speed, and u is a random variable that allows the random nature of the profiles to be modeled. W is defined by

$$W^2 = \frac{1}{1500} \int_{5000}^{20,000} v^2(H) dH \quad (19)$$

where $v(H)$ is the wind speed at height H . In data taken over Maryland, W was normally distributed with a mean value of 27 m/s and a standard deviation of 9 m/s. The random variable u is assumed to be a zero-mean, Gaussian variable with a covariance function given by

$$\langle u(H, t) u(H + \delta H, t + \delta t) \rangle = A(\delta H / 100) \exp(-\delta t / 5) + A(\delta H / 2000) \exp(-\delta t / 80) \quad (20)$$

where

$$A(\delta H/L) = 1 - |\delta H/L| \quad \text{for } |H| < L \quad (21)$$

and equals 0 otherwise.

The time interval δt is measured in minutes. The average C_n^2 profile can be found by recognizing that $\langle \exp(u) \rangle = \exp(1)$. To extend the model to local ground level, one should add the surface layer dependence (e.g., $H^{-4/3}$ for daytime).

Another attempt to extend the model to ground level is the Hufnagel-Valley model.⁷⁶ This is given by

$$C_n^2 = 0.00594 \left(\frac{W}{27} \right)^2 (H \times 10^{-5})^{10} \exp\left(-\frac{H}{1000}\right) + 2.7 \times 10^{-16} \exp\left(-\frac{H}{1500}\right) + A \exp\left(-\frac{H}{100}\right) \quad (22)$$

where W is commonly set to 21 and A to 1.7×10^{-14} . This specific model is referred to as the $HV_{5/7}$ model because it produces a coherence diameter r_0 of about 5 cm and an isoplanatic angle of about 7 μrad for a wavelength of 0.5 μm . Although this is not as accurate for modeling turbulence near the ground, it has the advantage that the moments of the turbulence profile important to propagation can be evaluated analytically.⁷⁶

The $HV_{5/7}$ model is plotted as a function of height in the dashed line in Fig. 30. The solid line in the figure is a balloon measurement taken in College Station, Pennsylvania. The data were reported with 20-m vertical resolution and smoothed with a Gaussian filter with a 100-m $\exp(-1)$ full-width. This particular data set was chosen because it has a coherence diameter of about 5 cm and an isoplanatic angle of about 7 μrad . The layered structure of the real atmosphere is clear in the data. Note also the difference between the model atmosphere and the real atmosphere even when the coherence diameter and the isoplanatic angle are similar.

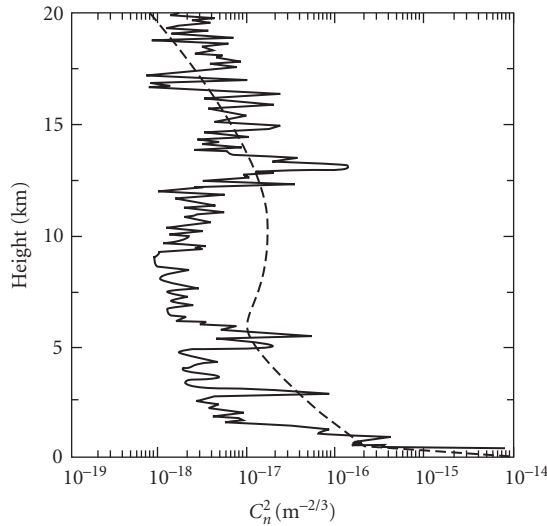


FIGURE 30 Turbulence strength C_n^2 as a function of height. The solid line is a balloon measurement made in College Station, Pennsylvania, and the dashed line is the $HV_{5/7}$ model. (Courtesy R. R. Beland, Geophysics Directorate, Phillips Laboratory, U.S. Air Force.)

Less is known about the vertical profiles of inner and outer scales. Near the ground (1 to 2 m) we typically observe inner scales of 5 to 10 mm over flat grassland in Colorado. Calculations of inner scale from measured values of Kolmogorov microscale range from 0.5 to 9 mm at similar heights.⁷⁷ Aircraft measurements of dissipation rate were used along with a viscosity profile calculated from typical profiles of temperature and pressure to estimate a profile of microscale.⁷⁸ Values increase monotonically to about 4 cm at a height of 10 km and to about 8 cm at 20 km.

Near the ground, the outer scale can be estimated using Monin-Obukhov similarity theory.⁴⁶ The outer scale can be defined as that separation at which the structure function of temperature fluctuations is equal to twice the variance. Using typical surface layer scaling relationships⁷⁹ we see that

$$L_0 = \begin{cases} 7.04H(1-7S_{MO})(1-16S_{MO})^{-32} & \text{for } -2 < S_{MO} \\ 7.04H(1+S_{MO})^{-3}(1+2.75S_{MO}^{2/3})^{-32} & \text{for } 0 < S_{MO} < \end{cases} \quad (23)$$

where S_{MO} is the Monin-Obukhov stability parameter. For typical daytime conditions ($S_{MO} < 0$), L_0 is generally between $H/2$ and H .

Above the boundary layer, the situation is more complex. Barat and Bertin⁸⁰ measured outer scale values of 10 to 100 m in a turbulent layer using a balloonborne instrument. Some recent optical data^{81,82} suggest that the outer scale is generally between 10 and 40 m.

Beam Wander

The first effect of refractive turbulence to consider is the wander of an optical beam in the atmosphere. The deviations of the centroid of a beam in each of the two orthogonal transverse axes will be independent of Gaussian random variables. This wander is generally characterized statistically by the variance of the angular displacement. In isotropic turbulence, the variances in the two axes are equal and the magnitude of the displacement is a Rayleigh random variable with a variance that is twice that of the displacement in a single axis. Both the single-axis and the magnitude variances are reported in the literature.

Three main approaches to the calculation of beam-wander variance have been used: (1) If diffraction effects are negligible and if the path-integrated turbulence is small enough, the geometric optics approximation^{83–86} can be used. Diffraction effects are negligible when the aperture diameter is greater than the Fresnel zone size.⁸⁶ The other condition requires that the product of the transverse coherence of the field and the aperture diameter is greater than the square of the Fresnel zone size.⁸⁶ (2) If diffraction must be considered, the Huygens-Fresnel approximation^{87–90} can be used. (3) The most complete theory uses the Markov random process approximation to the moment equation.^{91–93} In the two types of calculation that include diffraction, beams with a Gaussian irradiance profile are generally assumed.

In the geometric optics approximation, the variance of the angular displacement in a single axis is given by⁸⁶

$$\sigma_d^2 = 2.92D^{-1/3} \int_0^L dz C_n^2(z) \frac{\left(1 - \frac{z}{L}\right)^2}{\left|1 - \frac{z}{F}\right|^{1/3}} \quad (24)$$

where D is the diameter of the initial beam, z is position along the path, L is the propagation path length, and F is the geometric focal range of the initial beam (negative for a diverging beam). For homogeneous turbulence and a beam that does not come to a focus between the transmitter and observation plane, Eq. (24) reduces to

$$\sigma_d^2 = 0.97C_n^2D^{-1/3}L_2F_1\left(\frac{1}{3}, 1; 4; \frac{L}{F}\right) \quad (25)$$

where ${}_2F_1$ is the hypergeometric function. The hypergeometric function is 1 for a collimated beam and 1.125 for a focused beam.

In the Markov approximation, the single-axis variance is given by⁹¹

$$\sigma_d^2 = 4\pi^2 \int_0^L dz \left(1 - \frac{z}{L}\right)^2 \int_0^\infty dK K^3 \Phi_n(K, z) \times \exp \left\{ -\frac{K^2 D^2}{8} \left[\left(1 - \frac{z}{F}\right)^2 + \frac{16z^2}{K^2 D^2} \right] - \pi D_\Psi \left(\frac{Kz}{k} \right) \right\} \quad (26)$$

where $\Phi_n(K, z)$ is the path-dependent refractive index spectrum, K is the wave number of turbulence, D is the exp (-1) irradiance diameter of the initial beam, k is the optical wave number, and $D_\Psi(r)$ is the wave structure function for separation r of a spherical wave. The structure function is given by

$$D_\Psi(r) = 8\pi^2 k^2 \int_0^z dz' \int_0^\infty dK' K' \Phi_n(K', z') [1 - J_0(K' r z'/z)] \quad (27)$$

where J_0 is the zero-order Bessel function of the first kind. The last term in the exponential of Eq. (26) is a correction term for strong turbulence. The middle term includes the effects of diffraction.

Beam Spreading

The next effect of refractive turbulence to consider is the spread of an optical beam as it propagates through the atmosphere. There are two types of beam spread denoted as long term and short term. The long-term beam spread is defined as the turbulence-induced beam spread observed over a long time average. It includes the effects of the slow wander of the entire beam. The short-term beam spread is defined as the beam spread observed at an instant of time. It does not include the effects of beam wander and is often approximated by the long-term beam spread with the effects of wander removed, although the two are not identical.

We can consider beam wander to be caused by turbulent eddies that are larger than the beam. Short-term beam spread is caused by turbulent eddies that are smaller than the beam. There are more small eddies than large in the beam at any time which implies that the beam spread at any instant is averaged over more eddies. As a result, the fluctuations of short-term beam spread are much smaller than those of beam wander and are typically neglected. The primary effect of short-term beam spread is to spread the average energy over a larger area. Thus, the average value of the on-axis irradiance is reduced, and the average value of the irradiance at large angles is increased.

The extended Huygens-Fresnel principle can be used to calculate the long-term spread of a Gaussian beam in refractive turbulence.⁹⁴⁻⁹⁸ The exp (-1) irradiance radius of a Gaussian beam is

$$p_1 = \left[\frac{4}{k^2 D^2} + \frac{D^2}{4} \left(1 - \frac{L}{F}\right)^2 + \frac{4}{k^2 \rho_0^2} \right]^{1/2} \quad (28)$$

where D is the exp (-1) irradiance diameter of the source and ρ_0 is the phase coherence length that would be observed for a point source propagating from the receiver to the transmitter. The first term in this equation is the diffraction beam spread, the second is the geometrical optics projection of the transmitter aperture, and the final term is the total turbulence-induced spread.

The phase coherence length is defined as the transverse separation at which the coherence of the field is reduced to exp (-1) . If the coherence length ρ_0 is much larger than the inner scale, the structure function is given by $D_\Psi(r) = 2(r/\rho_0)^{5/3}$ and

$$\rho_0 = \left[1.46 k^2 \int_0^L dz \left(\frac{z}{L} \right)^{5/3} C_n^2(z) \right]^{-3/5} \quad (29)$$

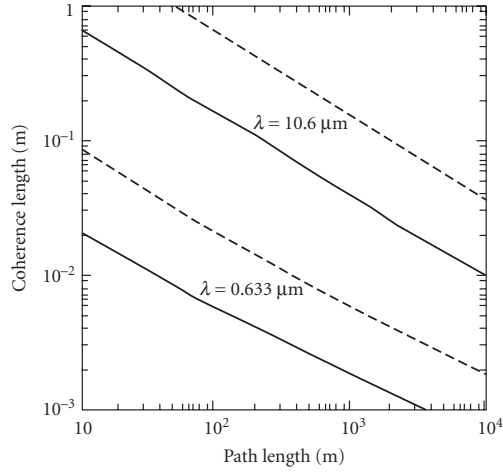


FIGURE 31 Coherence length ρ_0 as a function of path length for two common laser wavelengths. An inner scale of 1 cm, and values of 10^{-12} (solid lines) and 10^{-13} (dashed lines) were used in the calculations.

If the coherence length is much smaller than the inner scale, then $D_\psi(r) = 2(r/\rho_0)^2$ and

$$\rho_0 = \left[1.86k^2 \int_0^L dz \left(\frac{z}{L} \right)^2 C_n^2(z) l_0^{-1/3}(z) \right]^{-1/2} \quad (30)$$

Typical values of ρ_0 for spherical-wave propagation through homogeneous turbulence are presented in Fig. 31. In these plots there is a slope change from $L^{-3/5}$ to $L^{-1/2}$, where ρ_0 is about equal to the inner scale of 1 cm. This small change is nearly imperceptible in the figure.

The extended Huygens-Fresnel principle has been used to calculate the short-term beam spread by explicitly subtracting the beam wander.^{99–102} If ρ_0 and l_0 are much smaller than D , then the short-term beam spread is approximately given by

$$p_s = \left\{ \frac{4}{k^2 D^2} + \frac{D^2}{4L^2} \left(1 - \frac{L}{F} \right)^2 + \frac{4}{k^2 \rho_0^2} \left[1 - 0.62 \left(\frac{\rho_0}{D} \right)^{1/3} \right]^{6/5} \right\}^{1/2} \quad (31)$$

If ρ_0 is much greater than D , the turbulence-induced component of beam spreading can be neglected.

If inner scale and outer scale effects are included, more complicated integral expressions result.¹⁰³ Numerical calculations were performed for truncated Gaussian beams with central obscurations.¹⁰⁴ The following approximation was obtained by a curve fit to the results:

$$p_s = \left(1 + 0.182 \frac{D_{\text{eff}}^2}{r_0^2} \right)^{1/2} p_d \quad \text{for} \quad \frac{D_{\text{eff}}}{r_0} < 3$$

$$p_s = \left[1 + \left(\frac{D_{\text{eff}}}{r_0} \right)^2 - 1.18 \left(\frac{D_{\text{eff}}}{r_0} \right)^{-5/3} \right]^{1/2} p_d \quad \text{for} \quad 3 < \frac{D_{\text{eff}}}{r_0} < \quad (32)$$

where D_{eff} is the effective diameter of the truncated aperture, $r_0 = 2.099 p_0$, and p_d is the diffraction-limited value. These expressions agree fairly well with available data.^{87,105,106}

Imaging and Heterodyne Detection

The problem of imaging through the turbulent atmosphere is similar to the problem of beam propagation through the atmosphere. The dancing of an image in the focal plane of an imaging system is mathematically equivalent to the wander of a beam focused at the object by the same optical system. The resolution of the short-exposure image is equivalent to the short-term beam spread of the same focused beam. The resolution for a long exposure is related to the long-term beam spread.

Thus, the position of the image of a point object will drift in each axis in the focal plane. The variance of that drift is given by¹⁰⁷

$$\sigma_i^2 = 1.10 C_n^2 D^{-1/3} L F^2 \quad (33)$$

where D is the aperture diameter of the imaging system, F is its focal length, and L is the distance to the object.

Fried¹⁰⁸ used the idea of tilt correction to calculate the average image resolution for a short-exposure image in the turbulent atmosphere. This problem is mathematically equivalent to the propagation of a beam in the opposite direction if the imaging aperture replaces the beam width. These results were refined by Lutomirski et al.¹⁰⁹ and applied to a space-to-ground path by Valley.¹⁰³

Image resolution is also related to the signal-to-noise ratio of an optical heterodyne receiver. The long-exposure resolution is equivalent to a staring receiver. The short-exposure resolution is equivalent to a receiver that employs tilt-correction of the signal or of the local oscillator.^{108,110–112}

Scintillation

The refractive index inhomogeneities that distort the optical phase front also produce amplitude scintillation at some distance. The first cases to be considered were plane-wave and spherical-wave propagation through weak path-integrated turbulence where the weak-turbulence condition requires that fluctuations of irradiance be much less than the mean value. Tatarskii¹¹³ used a perturbation approach to the wave equation. Lee and Harp¹¹⁴ used a physical approach to arrive at the same results. These results are summarized in a number of good reviews.^{101,102,115,116}

We will first consider the weak-turbulence results. For propagation from space to the ground, the plane wave formula is generally valid. The variance of irradiance fluctuations (normalized by the mean irradiance value) is given by¹¹⁵

$$\sigma_r^2 = k^{7/6} \sec^{11/6} \theta \int_0^\infty dH H^{5/6} C_n^2(H) \quad (34)$$

where k is the optical wave number, θ is the zenith angle, and H is the height of the receiver above the ground. This expression is valid as long as the path-integrated turbulence is weak enough that the variance is much less than unity. This condition is usually met for near-zenith propagation.

For propagation of diverging waves near the ground, the spherical-wave approximation is often valid. Assuming constant turbulence along the path, the weak-turbulence variance in this case is given by¹¹⁵

$$\begin{aligned} \sigma_r^2 &= \exp [0.5 k^{7/6} L^{11/6} C_n^2] - 1 & \text{for } l_0 < \sqrt{L} \\ \sigma_r^2 &= \exp [1.28 L^3 l_0^{-7/3} C_n^2] - 1 & \text{for } l_0 > \sqrt{L} \end{aligned} \quad (35)$$

where L is the path length and l_0 is the inner scale of turbulence.

For narrow-beam propagation, the effects of the finite beam must be considered. Kon and Tatarskii¹¹⁷ calculated the amplitude fluctuations of a collimated beam using the perturbation technique. Schmeltzer¹¹⁸ extended these results to include focused beams. These results were used to obtain numerical values for a variety of propagation conditions.^{119–121} Ishimaru^{122–124} used a spectral representation to obtain similar results. Under certain conditions, one sees a reduction in the variance on the optical axis^{119,120,125,126} and an increase off of the optical axis.¹²⁷

The spatial scale of weak scintillations is about equal to the larger of either the Fresnel zone size $(L/k)^{1/2}$ or the inner scale.¹¹⁵ Scintillation will be reduced if an aperture larger than the scale size is used to collect the light. If the aperture diameter D is much larger than the scale size, the reduction factor can be expressed as^{128,129}

$$A = C \left(\frac{D}{D_0} \right)^{-7/3} \quad (36)$$

where D_0 is $(\lambda L)^{1/2}$ and C is 4.71 for a plane wave and 23.5 for a spherical wave when the inner scale is much smaller than the Fresnel zone. If the inner scale is much larger than the Fresnel zone, D_0 is the inner scale and C is 0.45 for a plane wave and 9.17 for a spherical wave.

It is generally accepted that the probability density function for weak scintillation is log normal.^{101,113,115,130} This density function has the form

$$p(I) = \frac{1}{\sqrt{2\pi}\sigma_I I} \exp \left[-\frac{1}{2\sigma_r^2} (\ln I + 0.5\sigma_r^2)^2 \right] \quad (37)$$

The weak-turbulence theory is, in essence, a single-scattering theory. As the path-integrated turbulence becomes larger, multiple-scattering effects become important, and this theory breaks down. Actual observed values of irradiance variance are smaller than predicted in this region as shown in Fig. 32. The circles in this figure are 1-minute averages of irradiance variance for 488-nm laser light propagated across 1200 m of flat grassland. The solid line is the weak-turbulence approximation.

In these and other experiments, variance values reach a peak value of between 3 and 5 for a spherical wave and they begin to decrease with increasing turbulence strength.^{131–134} In the limit of infinite path-integrated turbulence, the normalized variance of irradiance is predicted to approach unity.^{135,136} In the intermediate region near the peak irradiance, numerical evaluation of currently available theories is impractical, and variance values are most easily obtained through numerical simulation.^{137–140}

In very strong turbulence two distinct spatial scales are evident in the scintillation pattern.^{136,141–144} The smaller scale is about the size of the coherence length ρ_0 discussed under “Beam Spreading.” The larger scale is the size of the scattering disk which is the ratio of the square of the Fresnel zone to the coherence length $(L/k\rho_0)$. As turbulence increases, the small scale becomes smaller and the large scale becomes larger. The strength of the small-scale fluctuations is constant in this regime and contributes a value of unity to the variance. The large-scale fluctuations contribute the rest of the variance and become weaker with increasing turbulence strength.

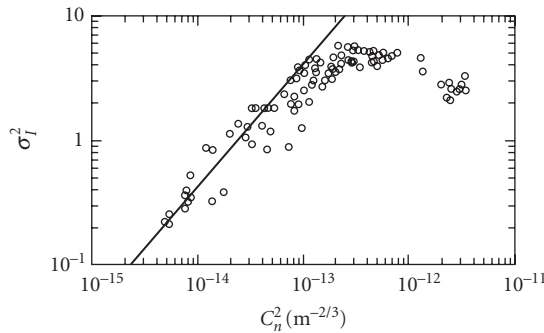


FIGURE 32 Plot of normalized variance of irradiance σ_I^2 as a function of turbulence strength. C_n^2 circles are data taken with a 488-nm wavelength laser over a 1200-m path, and the line is the corresponding weak-turbulence theory.

The two-scale nature of the irradiance fluctuations suggests a two-scale probability density function. The log-normally modulated Rician density function has been shown to agree with experimental data¹³⁰ and simulations.¹⁴⁵ It has the form

$$p(I) = \int_0^\infty dz \frac{(1+r)}{z} \exp\left(-r - \frac{1+r}{z} I\right) I_0 \left\{ 2 \left[\frac{(1+r)r}{z} \right]^{1/2} I \right\} \frac{1}{(2\pi)^{1/2} \sigma_z z} \exp\left[-\frac{(\ln z + \frac{1}{2}\sigma_z^2)^2}{2\sigma_z^2}\right] \quad (38)$$

where r is a coherence parameter that decreases with increasing turbulence, I_0 is the 0-order modified Bessel function of the first kind, and σ_z^2 is the variance of the logarithm of the modulation factor z . For $r \gg 1$ this function reduces to the lognormal; for $r \ll 1$, it reduces to a log-normally modulated exponential.^{134,146}

$$p(I) = \frac{1}{\sqrt{2\pi}\sigma_z} \int_0^\infty \frac{dz}{z^2} \exp\left[-\frac{I}{z} - \frac{(\ln z + \frac{1}{2}\sigma_z^2)^2}{2\sigma_z^2}\right] \quad (39)$$

where the parameter σ_z^2 is related to the irradiance variance by the relationship

$$\sigma_I^2 = 2 \exp(\sigma_z^2) - 1 \quad (40)$$

As this model suggests, the density function of the fluctuations approaches a lognormal, even in strong turbulence, if an aperture $\gg \rho_0$ is used.¹⁴⁷

If the fluctuations at both large and small scales are approximated by gamma distributions, the resulting integral can be evaluated analytically to get the gamma-gamma density function:¹⁴⁸

$$p(I) = \frac{2(\alpha\beta)^{0.5(\alpha+\beta)}}{\Gamma(\alpha)\Gamma(\beta)} I^{0.5(\alpha+\beta)-1} K_{\alpha-\beta}(2\sqrt{\alpha\beta}I) \quad (41)$$

where α and β are related to the variances at the two scales, Γ is the gamma function, and K is the modified Bessel function of the second kind.

3.7 EXAMPLES OF ATMOSPHERIC OPTICAL REMOTE SENSING

One of the more important applications of atmospheric optics is optical remote sensing. Atmospheric optical remote sensing concerns the use of an optical or laser beam to remotely sense information about the atmosphere or a distant target. Optical remote sensing measurements are diverse in nature and include the use of a spectral radiometer aboard a satellite for the detection of trace species in the upper atmosphere, the use of spectral emission and absorption from the earth for the detection of the concentration of water vapor in the atmosphere, the use of lasers to measure the range-resolved distribution of several molecules including ozone in the atmosphere, and Doppler wind measurements. In this section, some typical optical remote sensing experiments will be presented in order to give a flavor of the wide variety of atmospheric optical measurements that are currently being conducted. More in-depth references can be found in several current journal papers, books, and conference proceedings.^{149–156}

The Upper Atmospheric Research Satellite (UARS) was placed into orbit in September 1991 as part of the Earth Observing System. One of the optical remote sensing instruments aboard UARS is the High Resolution Doppler Imager (HRDI) developed by P. Hays' and V. Abreu's group while at the University of Michigan.^{157,158} The HRDI is a triple etalon Fabry-Perot Interferometer designed to measure Doppler shifts of molecular absorption and emission lines in the earth's atmosphere in order to determine the wind velocity of the atmosphere. A wind velocity of 10 m/s causes a Doppler shift of 2×10^{-5} nm for the oxygen lines detected near a wavelength of 600 to 800 nm. A schematic of the instrument is given in Fig. 33a which shows the telescope, triple Fabry-Perots, and unique imaging Photo-Multiplier tubes to detect the Fabry-Perot patterns of the spectral absorption lines. The HRDI instrument is a passive remote sensing system and uses the reflected or scattered sunlight as its illumination source. Figure 33b shows the wind field measured by UARS (HRDI) for an altitude of 90 km.

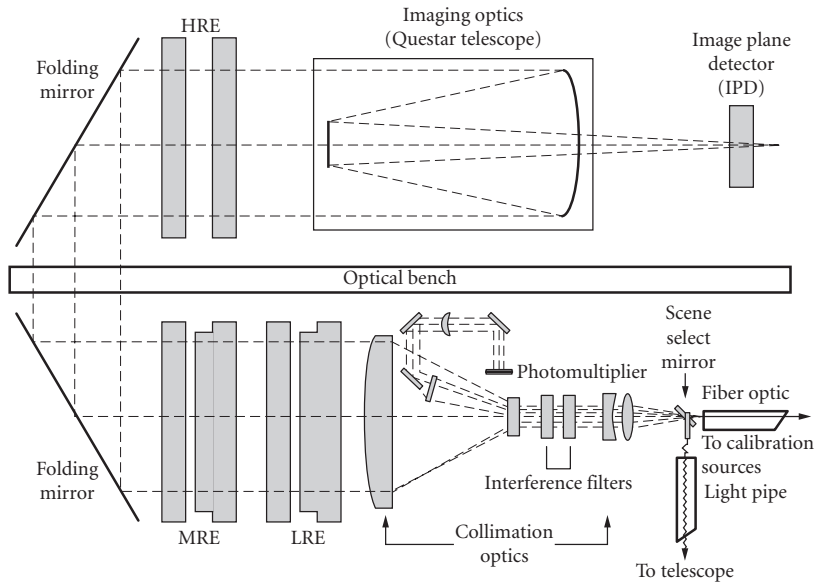


FIGURE 33a Optical layout of the Upper Atmospheric Resolution Satellite (UARS) High Resolution Doppler Imager (HRDI) instrument. FO = fiber optic, LRE = low-resolution etalon, MRE = medium-resolution etalon, HRE = high-resolution etalon. (From Hays, Ref. 157.)

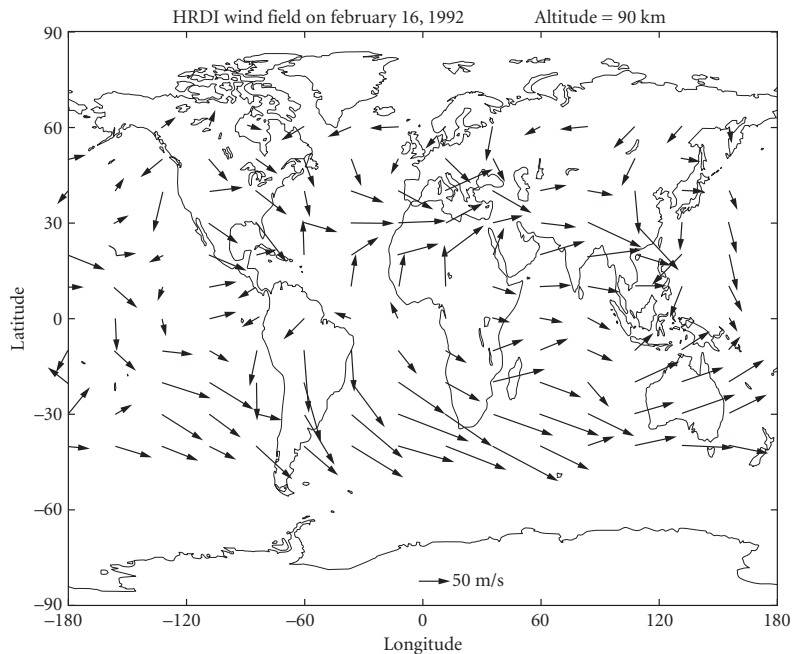


FIGURE 33b Upper atmospheric wind field measured by UARS/HRDI satellite instrument. (From Hays, Ref. 157.)

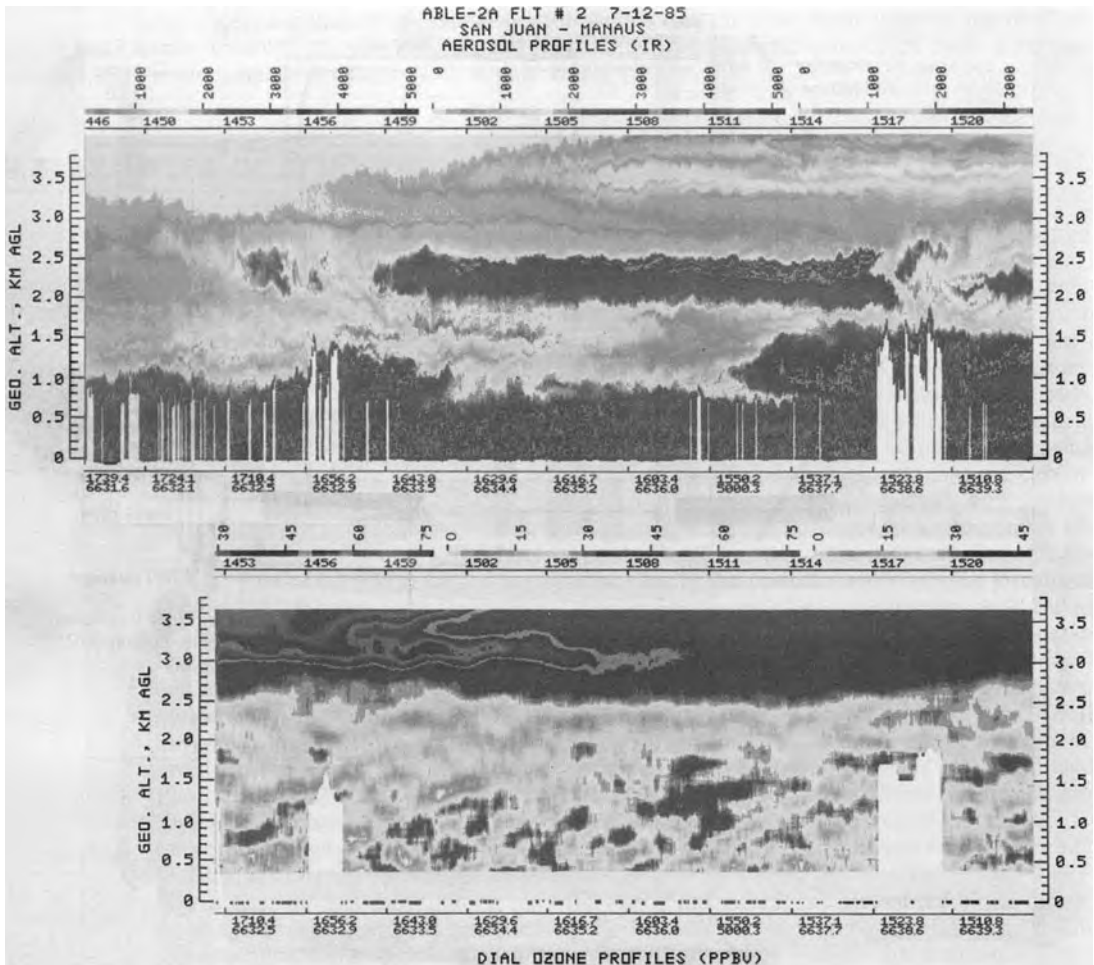


FIGURE 34 Range-resolved lidar measurements of atmospheric aerosols and ozone density. (From Browell, Ref. 159.)

Another kind of atmospheric remote sensing instrument is represented by an airborne laser radar (lidar) system operated by E. Browell's group at NASA/Langley.¹⁵⁹ Their system consists of two pulsed, visible-wavelength dye laser systems that emit short (10 ns) pulses of tunable optical radiation that can be directed toward aerosol clouds in the atmosphere. By the proper tuning of the wavelength of these lasers, the difference in the absorption due to ozone, water vapor, or oxygen in the atmosphere can be measured. Because the laser pulse is short, the timing out to the aerosol scatterers can be determined and range-resolved lidar measurements can be made. Figure 34 shows range-resolved lidar backscatter profiles obtained as a function of the lidar aircraft ground position. The variation in the atmospheric density and ozone distribution as a function of altitude and distance is readily observed.

A Coherent Doppler lidar is one which is able to measure the Doppler shift of the backscattered lidar returns from the atmosphere. Several Doppler lidar systems have been developed which can determine wind speed with an accuracy of 0.1 m/s at ranges of up to 15 km. One such system is operated by M. Hardesty's group at NOAA/WPL for the mapping of winds near airports and for meteorological studies.^{160,161} Figure 35 shows a two-dimensional plot of the measured wind velocity obtained during the approach of a wind gust front associated with colliding thunderstorms; the upper

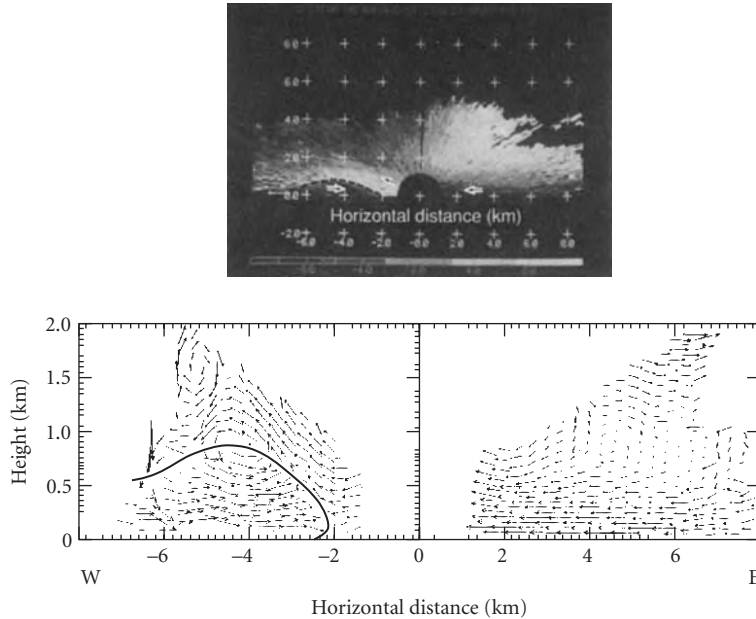


FIGURE 35 Coherent Doppler lidar measurements of atmospheric winds showing velocity profile of gust front. Upper plot is real-time display of Doppler signal and lower plot is range-resolved wind field. (From Hardesty, Ref. 160.)

figure shows the real-time Doppler lidar display of the measured radial wind velocity, and the lower plot shows the computed wind velocity. As seen, a Doppler lidar system is able to remotely measure the wind speed with spatial resolution on the order of 100 m. A similar Doppler lidar system is being considered for the early detection of windshear in front of commercial aircraft.

A further example of atmospheric optical remote sensing is that of the remote measurement of the global concentration and distribution of atmospheric aerosols. P. McCormick's group at NASA/Langley and Hampton University has developed the SAGE II satellite system which is part of a package of instruments to detect global aerosol and selected species concentrations in the atmosphere.¹⁶² This system measures the difference in the optical radiation emitted from the earth's surface and the differential absorption due to known absorption lines or spectral bands of several species in the atmosphere, including ozone. The instrument also provides for the spatial mapping of the concentration of aerosols in the atmosphere, and an example of such a measurement is shown in Fig. 36. This figure shows the measured concentration of aerosols after the eruption of Mt. Pinatubo and demonstrates the global circulation and transport of the injected material into the earth's atmosphere.

More recently, this capability has been refined by David Winker's group at NASA that developed the laser based CALIPSO lidar satellite which has produced continuous high-resolution 3D maps of global cloud and aerosol distributions since its launch in 2006.

There are several ongoing optical remote sensing programs to map and measure the global concentration of CO_2 and other green house gases in the atmosphere. For example, the spaceborne Atmospheric Infrared Sounder (AIRS) from JPL has measured the CO_2 concentration at the mid-troposphere (8 km altitude) beginning in 2003, and the NASA Orbiting Carbon Observatory (OCO) to be launched in 2008 will be the first dedicated spaceborne instrument to measure the sources and sinks of CO_2 globally. Both of these instruments use optical spectroscopy of atmospheric CO_2 lines to measure the concentration of CO_2 in the atmosphere.

Finally, there are related nonlinear optical processes that can also be used for remote sensing. For example, laser-induced-breakdown spectroscopy (LIBS) has been used recently in a lidar system for the remote detection of chemical species by focusing a pulsed laser beam at a remote target, producing a plasma spark at the

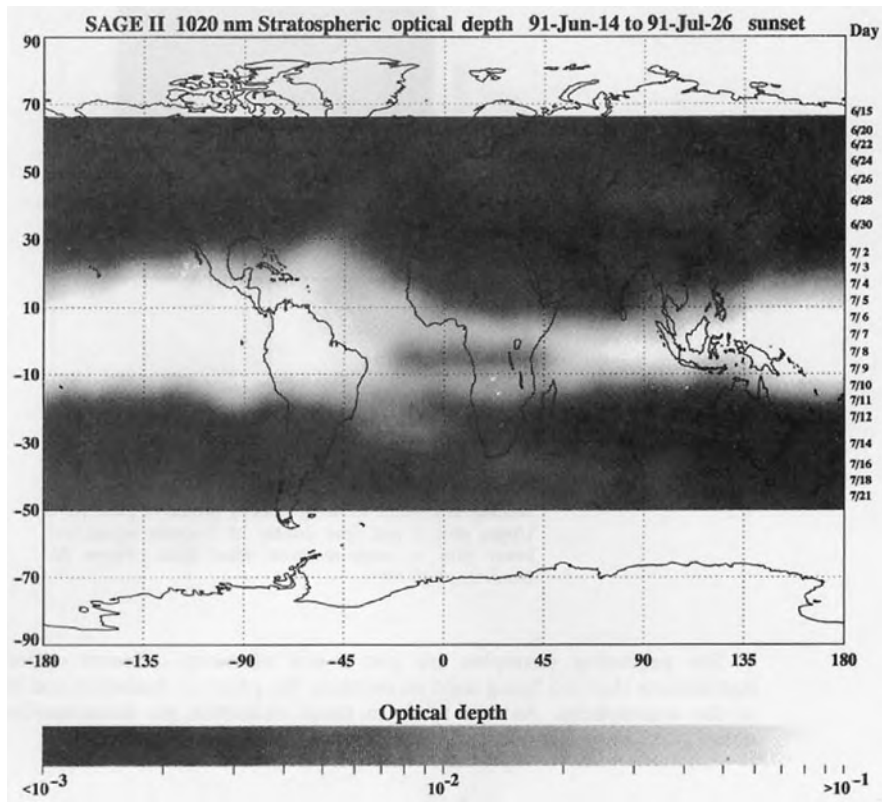


FIGURE 36 Measurement of global aerosol concentration using SAGE II satellite following eruption of Mt. Pinatubo. (From McCormick, Ref. 162.)

target, and analyzing the emitted spectral light after being transmitted back through the atmosphere.^{163,164} Another technique is the use of a high-power femtosecond pulse-length laser to produce a dielectric breakdown (spark) in air that self focuses into a long filament of several 100s of meters in length. The channeling of the laser filament has been used to remotely detect distant targets and atmospheric gases.¹⁶⁵

The preceding examples are just a few of many different optical remote sensing instruments that are being used to measure the physical dynamics and chemical properties of the atmosphere. As is evident in these examples, an understanding of atmospheric optics plays an important and integral part in these measurements.

3.8 METEOROLOGICAL OPTICS

One of the most colorful aspects of atmospheric optics is that associated with meteorological optics. Meteorological optics involves the interplay of light with the atmosphere and the physical origin of the observed optical phenomena. Several excellent books have been written about this subject, and the reader should consult these and the contained references.^{166,167} While it is beyond the scope of this chapter to present an overview of meteorological optics, some specific optical phenomena will be described to give the reader a sampling of some of the interesting effects involved in naturally occurring atmospheric and meteorological optics.

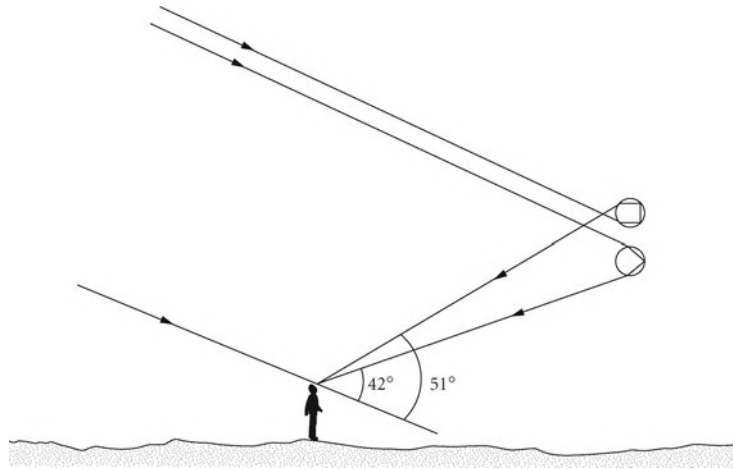


FIGURE 37 Different raindrops contribute to the primary and to the larger, secondary rainbow. (From Greenler, Ref. 167.)

Some of the more common and interesting meteorological optical phenomena involve rainbows, ice-crystal halos, and mirages. The rainbow in the atmosphere is caused by internal reflection and refraction of sunlight by water droplets in the atmosphere. Figure 37 shows the geometry involved in the formation of a rainbow, including both the primary and larger secondary rainbow. Because of the dispersion of light within the water droplet, the colors or wavelengths are separated in the backscattered image. Although rainbows are commonly observed in the visible spectrum, such refraction also occurs in the infrared spectrum. As an example, Fig. 38 shows a natural rainbow in the atmosphere photographed with IR-sensitive film by R. Greenler.¹⁶⁷



FIGURE 38 A natural infrared rainbow. (From Greenler, Ref. 167.)

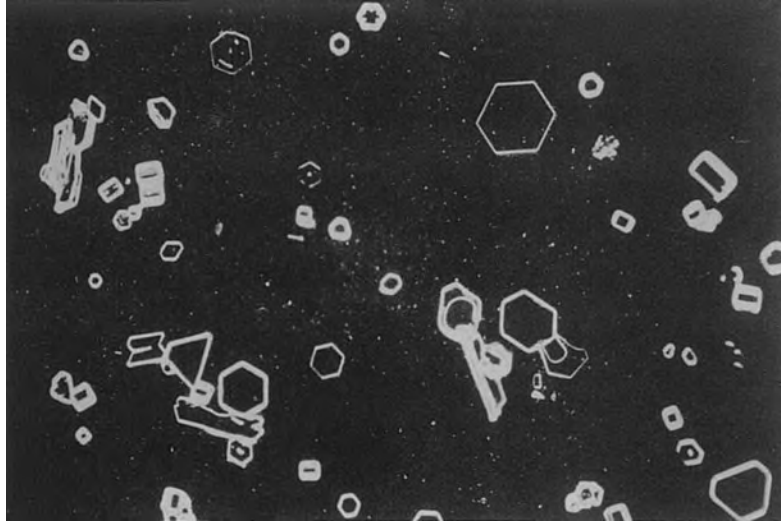


FIGURE 39 Photograph of magnified small ice crystals collected as they fell from the sky. (From Greenler, Ref. 167.)

The phenomena of halos, arcs, and spots are due to the refraction of light by ice crystals suspended in the atmosphere. Figure 39 shows a photograph of collected ice crystals as they fell from the sky. The geometrical shapes, especially the hexagonal (six-sided) crystals, play an important role in the formation of halos and arcs in the atmosphere.

The common optical phenomenon of the mirage is caused by variation in the temperature and thus, the density of the air as a function of altitude or transverse geometrical distance. As an example, Fig. 40 shows the geometry of light-ray paths for a case where the air temperature decreases with height

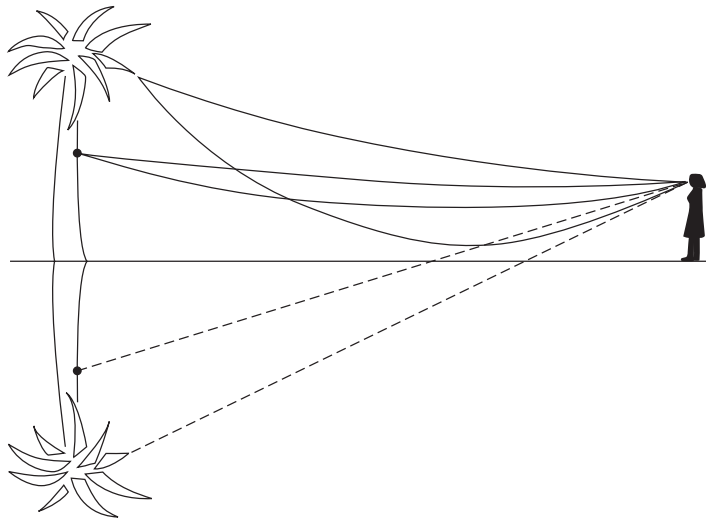


FIGURE 40 The origin of the inverted image in the desert mirage. (From Greenler, Ref. 167.)



FIGURE 41 The desert (or hot-road) mirage. In the inverted part of the image you can see the apparent reflection of motorcycles, cars, painted stripes on the road, and the grassy road edge. (From Greenler, Ref. 167.)

to a sufficient extent over the viewing angle that the difference in the index of refraction can cause a refraction of the image similar to total internal reflection. The heated air (less dense) near the ground can thus act like a mirror, and reflect the light upward toward the viewer. As an example, Fig. 41 shows a photograph taken by Greenler of motorcycles on a hot road surface. The reflected image of the motorcycles “within” the road surface is evident. There are many manifestations of mirages dependent upon the local temperature gradient and geometry of the situation. In many cases, partial and distorted images are observed leading to the almost surreal connotation often associated with mirages.

Finally, another atmospheric meteorological phenomenon is that of the green flash. A green flash is observed under certain conditions just as the sun is setting below the horizon. This phenomenon is easily understood as being due to the different relative displacement of each different wavelength or color in the sun’s image due to spatially distributed refraction of the atmosphere.¹⁶⁷ As the sun sets, the last image to be observed is the shortest wavelength color, blue. However, most of the blue light has been Rayleigh scattered from the image seen by the observer so that the last image observed is closer to a green color. Under extremely clear atmospheric conditions when the Rayleigh scattering is not as preferential in scattering the blue light, the flash has been reported as blue in color. Lastly, one of the authors (DK) has observed several occurrences of the green flash and noticed that the green flash seems to be seen more often from sunsets over water than over land, suggesting that a form of water vapor layer induced ducting of the optical beam along the water’s surface may be involved in enhancing the absorption and scattering process.

3.9 ATMOSPHERIC OPTICS AND GLOBAL CLIMATE CHANGE

Of importance, atmospheric optics largely determines the earth’s climate because incoming sunlight (optical energy) is scattered and absorbed and outgoing thermal radiation is absorbed and reemitted by the atmosphere. This net energy flux, either incoming or outgoing, determines if the earth’s

climate will warm or cool, and is directly related to the radiative transfer calculations mentioned in Sec. 3.4. A convenient way to express changes in this energy balance is in terms of the radiative forcing, which is defined as the change in the incoming and outgoing radiative flux at the top of the troposphere. To a good approximation, individual forcing terms add linearly to produce a linear surface-temperature response at regional to global scales. Reference 168 by the Intergovernmental Panel on Climate Change (2007) provides a detailed description of radiative forcing and the scientific basis for these findings.

As an example, Fig. 42 shows that the major radiative forcing mechanisms for changes between the years of 1750 and 2005 are the result of anthropogenic changes to the atmosphere.¹⁶⁸ As can be seen, the largest effect is that of infrared absorption by the greenhouse gases, principally CO_2 . This is the reason that the measurement of the global concentration of CO_2 is important for accurate predictions for future warming or cooling trends of the earth. Other forcing effects are also shown in Fig. 42. Water vapor remains the largest absorber of infrared radiation, but changes in water vapor caused

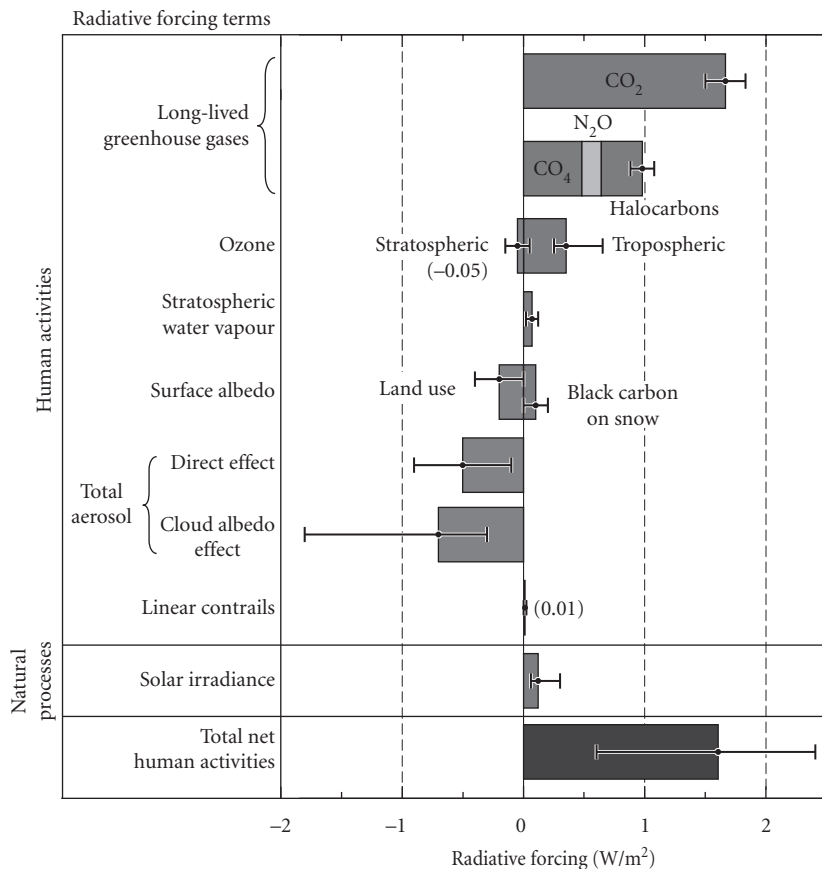


FIGURE 42 Effect of atmospheric optics on global climate change represented by the radiative forcing terms between the years of 1750 and 2005. The change in the energy (W/m^2) balance for the earth over this time period is shown resulting in a net positive energy flow onto the earth, with a potential warming effect. Contributions due to individual terms, such as changes in the CO_2 gas concentration or changes in land use, are shown. (Reprinted with permission from Ref. 168.)

by climate changes are considered a response rather than a forcing term. The exception is a small increase in stratospheric water vapor produced by methane emissions. Similarly, the effects of clouds are part of a climate response, except for the increase in cloudiness that is a direct result of increases in atmospheric aerosols. The linearity of climate response to radiative forcing implies that the most efficient radiative transfer calculations for each term can be used. For example, a high resolution spectral absorption calculation is not needed to calculate the radiative transfer through clouds, and a multiple-scattering calculation is not needed to calculate the effects of absorption by gases.

In summary, Fig. 42 shows the important contributions of radiative forcing effects that contribute to changes in the heat balance of the earth's atmosphere. All of the noted terms are influenced by the optical properties of the atmosphere whether due to absorption of sunlight by the line or band spectrum of molecules in the air, the reflection of light by the earth's surface or oceans, or reabsorption of thermal radiation by greenhouse gases. The interested reader is encouraged to study Ref. 168 for more information, and references therein.

3.10 ACKNOWLEDGMENTS

We would like to acknowledge the contributions and help received in the preparation of this chapter and in the delineation of the authors' work. The authors divided the writing of the chapter sections as follows: D. K. Killinger served as lead author and wrote Secs. 3.2 through 3.4 and Secs. 3.7 and 3.8 on atmospheric interactions with light, remote sensing, and meteorological optics. L. S. Rothman wrote and provided the extensive background information on HITRAN, FASCODE, and LOWTRAN in Sec. 3.5. The comprehensive Secs. 3.6 and 3.9 were written by J. H. Churnside. The data of Fig. 29 were provided by G. R. Ochs of NOAA/WPL and the data in Fig. 30 were provided by R. R. Beland of the Geophysics Directorate, Phillips Laboratory. We wish to thank Prof. Robert Greenler for providing original photographs of the meteorological optics phenomena; Paul Hays, Vincent Abreu, and Wilbert Skinner for information on the HRDI instrument; P. McCormick and D. Winker for SAGE II data; Mike Hardesty for Doppler lidar wind profiles; and Ed Browell for lidar ozone mapping data. We want to thank A. Jursa for providing a copy of the *Handbook of Geophysics*. R. Measures for permission to use diagrams from his book *Laser Remote Sensing*, and M. Thomas and D. Duncan for providing a copy of their chapter on atmospheric optics from *The Infrared Handbook*.

Finally, we wish to thank many of our colleagues who have suggested topics and technical items added to this work. We hope that the reader will gain an overall feeling of atmospheric optics from reading this chapter, and we encourage the reader to use the references cited for further in-depth study.

3.11 REFERENCES

1. R. M. Goody and Y. L. Young, *Atmospheric Radiation*, Oxford University Press, London, 1989.
2. W. G. Driscoll (ed.), Optical Society of America, *Handbook of Optics*, McGraw-Hill, New York, 1978.
3. A. S. Jursa (ed.), *Handbook of Geophysics and the Space Environment*, Air Force Geophysics Lab., NTIS Doc#ADA16700, 1985.
4. W. Wolfe and G. Zissis, *The Infrared Handbook*, Office of Naval Research, Washington D.C., 1978.
5. R. Measures, *Laser Remote Sensing*, Wiley-Interscience, John Wiley & Sons, New York, 1984.
6. "Major Concentration of Gases in the Atmosphere," NOAA S/T 76-1562, 1976; "AFGL Atmospheric Constituent Profiles (0-120 km)," AFGL-TR-86-0110, 1986; U.S. Standard Atmosphere, 1962 and 1976; Supplement 1966, U.S. Printing Office, Washington D.C., 1976.
7. E. P. Shettle and R. W. Fenn, "Models of the Aerosols of the Lower Atmosphere and the Effects of Humidity Variations on Their Optical Properties," AFGL TR-79-0214; ADA 085951, 1979.
8. A. Force, D. K. Killinger, W. DeFeo, and N. Menyuk, "Laser Remote Sensing of Atmospheric Ammonia Using a CO₂ Lidar System," *Appl. Opt.* **24**:2837 (1985).

9. B. Nilsson, "Meteorological Influence on Aerosol Extinction in the 0.2–40 μm Range," *Appl. Opt.* **18**:3457 (1979).
10. J. F. Luhr, "Volcanic Shade Causes Cooling," *Nature* **354**:104 (1991).
11. L. S. Rothman, R. R. Gamache, A. Goldman, L. R. Brown, R. A. Toth, H. Pickett, R. Poynter, et al., "The HITRAN Database: 1986 Edition," *Appl. Optics* **26**:4058 (1986).
12. L. S. Rothman, R. R. Gamache, R. H. Tipping, C. P. Rinsland, M. A. H. Smith, D. C. Benner, V. Malathy Devi, et al., "The HITRAN Molecular Database: Editions of 1991 and 1992," *J. Quant. Spectrosc. Radiat. Transfer* **48**:469 (1992).
13. L. S. Rothman, I. E. Gordan, A. Barbe, D. Chris Benner, P. F. Bernath, M. Birk, L. R. Brown, et al., "The HITRAN 2008 Molecular Spectroscopic Database," *J. Quant. Spectrosc. Radiat. Transfer* **110**:533 (2009).
14. E. E. Whiting, "An Empirical Approximation to the Voigt Profile," *J. Quant. Spectrosc. Radiat. Transfer* **8**:1379 (1968).
15. J. Olivero and R. Longbothum, "Empirical Fits to the Voigt Linewidth: A Brief Review," *J. Quant. Spectrosc. Radiat. Transfer* **17**:233 (1977).
16. F. Schreir, "The Voigt and Complex Error Function—A Comparison of Computational Methods," *J. Quant. Spectrosc. Radiat. Transfer* **48**:734 (1992).
17. J. -M. Hartmann, C. Boulet, and D. Robert, *Collision Effects on Molecular Spectra. Laboratory Experiments and Models, Consequences for Applications*, Elsevier, Paris, 2008.
18. D. E. Burch, "Continuum Absorption by H_2O ," AFGL-TR-81-0300; ADA 112264, 1981; S. A. Clough, F. X. Kneizys, and R. W. Davies, "Lineshape and the Water Vapor Continuum," *Atmospheric Research* **23**:229 (1989).
19. S. Shardanand and A. D. Prasad Rao, "Absolute Rayleigh Scattering Cross Section of Gases and Freons of Stratospheric Interest in the Visible and Ultraviolet Region," NASA TN 0-8442 (1977); R. T. H. Collins and P. B. Russell, "Lidar Measurement of Particles and Gases by Elastic and Differential Absorption," in D. Hinkley (ed.), *Laser Monitoring of the Atmosphere*, Springer-Verlag, New York, 1976.
20. E. U. Condon and H. Odishaw (eds.), *Handbook of Physics*, McGraw-Hill, New York, 1967.
21. S. R. Pal and A. I. Carswell, "Polarization Properties of Lidar Backscattering from Clouds," *Appl. Opt.* **12**:1530 (1973).
22. G. Mie, "Bertrage Z. Phys. TruberMedien, Spezeziell kolloidaler Metallosungen," *Ann. Physik* **25**:377 (1908).
23. D. Deirrendjian, "Scattering and Polarization Properties of Water Clouds and Hazes in the Visible and Infrared," *Appl. Opt.* **2**:187 (1964).
24. E. J. McCartney, *Optics of the Atmosphere*, Wiley, New York, 1976.
25. M. Wright, E. Proctor, L. Gasiorsek, and E. Liston, "A Preliminary Study of Air Pollution Measurement by Active Remote Sensing Techniques," NASA CR-132724, 1975.
26. P. McCormick and D. Winker, "NASA/LaRC: 1 μm Lidar Measurement of Aerosol Distribution," Private communication, 1991.
27. D. K. Killinger and N. Menyuk, "Laser Remote Sensing of the Atmosphere," *Science* **235**:37 (1987).
28. R. W. Boyd, *Nonlinear Optics*, Academic Press, Orlando, Fla., 1992.
29. M. D. Levenson and S. Kano, *Introduction to Nonlinear Laser Spectroscopy*, Academic Press, Boston, 1988.
30. S. A. Clough, F. X. Kneizys, E. P. Shettle, and G. P. Anderson, "Atmospheric Radiance and Transmittance: FASCOD2," *Proc. of Sixth Conf. on Atmospheric Radiation*, Williamsburg, Va., published by Am. Meteorol. Soc., Boston, 1986.
31. J. A. Dowling, W. O. Gallery, and S. G. O'Brian, "Analysis of Atmospheric Interferometer Data," AFGL-TR-84-0177, 1984.
32. R. Isaacs, S. Clough, R. Worsham, J. Moncet, B. Lindner, and L. Kaplan, "Path Characterization Algorithms for FASCOD," Tech. Report GL-TR-90-0080, AFGL, 1990; ADA#231914.
33. F. X. Kneizys, E. Shettle, W. O. Gallery, J. Chetwynd, L. Abreu, J. Selby, S. Clough, and R. Fenn, "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN6," AFGL TR-83-0187, 1983; ADA#137786.
34. F. X. Kneizys, E. Shettle, L. Abreu, J. Chetwynd, G. Anderson, W. O. Gallery, J. E. A. Selby, and S. Clough, "Users Guide to LOWTRAN7," AFGL TR-88-0177, 1988; ADA#206773.
35. The HITRAN database compilation is available on the internet and can be accessed by filling out a request form at the HITRAN web site <http://cfa.harvard.edu/hitrان>. The MODTRAN code (with LOWTRAN7 embedded within), and a PC version of FASCOD, can be obtained from the ONTAR Corp, 9 Village Way, North Andover, MA 01845-2000.

36. D. K. Killinger and W. Wilcox, Jr., HITRAN-PC Program; can be obtained from ONTAR Corp. at www.ontar.com.
37. NIST/EPA Gas Phase Infrared Database, U.S. Dept. of Commerce, NIST, Standard Ref. Data, Gaithersburg, MD 20899.
38. Vapor phase infrared spectral library, Pacific Northwest National Laboratory, Richland, WA 99352.
39. LAB_CALC, Galactic Industries, 395 Main St., Salem, NH, 03079 USA; Infrared Analytics, 1424 N. Central Park Ave, Anaheim, CA 92802; Aldrich Library of Spectra, Aldrich Co., Milwaukee, WI 53201; Sadtler Spectra Data, Philadelphia, PA 19104-2596; Coblenz Society, P.O. Box 9952, Kirkwood, MO 63122.
40. L. C. Andrews and R. L. Phillips, *Laser Beam Propagation through Random Media*, 2nd ed., SPIE Press, Washington, 2005.
41. R. J. Hill, "Models of the Scalar Spectrum for Turbulent Advection," *J. Fluid Mech.* **88**:541–662 (1978).
42. J. H. Churnside, "A Spectrum of Refractive-Index Turbulence in the Turbulent Atmosphere," *J. Mod. Opt.* **37**:13–16 (1990).
43. L. C. Andrews, "An Analytical Model for the Refractive Index Power Spectrum and Its Application to Optical Scintillations in the Atmosphere," *J. Mod. Opt.* **39**:1849–1853 (1992).
44. R. S. Lawrence, G. R. Ochs, and S. F. Clifford, "Measurements of Atmospheric Turbulence Relevant to Optical Propagation," *J. Opt. Soc. Am.* **60**:826–830 (1970).
45. M. A. Kallistratova and D. F. Timanovskiy, "The Distribution of the Structure Constant of Refractive Index Fluctuations in the Atmospheric Surface Layer," *Iz. Atmos. Ocean. Phys.* **7**:46–48 (1971).
46. A. S. Monin and A. M. Obukhov, "Basic Laws of Turbulent Mixing in the Ground Layer of the Atmosphere," *Trans. Geophys. Inst. Akad. Nauk. USSR* **151**:163–187 (1954).
47. A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics: Mechanics of Turbulence*, MIT Press, Cambridge, 1971.
48. J. C. Wyngaard, Y. Izumi, and S. A. Collins, Jr., "Behavior of the Refractive-Index-Structure Parameter Near the Ground," *J. Opt. Soc. Am.* **61**:1646–1650 (1971).
49. L. R. Tsang, "Microstructure of Temperature Fields in the Free Atmosphere," *Radio Sci.* **4**:1175–1177 (1969).
50. A. S. Frisch and G. R. Ochs, "A Note on the Behavior of the Temperature Structure Parameter in a Convective Layer Capped by a Marine Inversion," *J. Appl. Meteorol.* **14**:415–419 (1975).
51. K. L. Davidson, T. M. Houlihan, C. W. Fairall, and G. E. Schader, "Observation of the Temperature Structure Function Parameter, C_n^2 , over the Ocean," *Boundary-Layer Meteorol.* **15**:507–523 (1978).
52. K. E. Kunkel, D. L. Walters, and G. A. Ely, "Behavior of the Temperature Structure Parameter in a Desert Basin," *J. Appl. Meteorol.* **15**:130–136 (1981).
53. W. Kohsiek, "Measuring C_T^2 , C_Q^2 , and C_{TQ} in the Unstable Surface Layer, and Relations to the Vertical Fluxes of Heat and Moisture," *Boundary-Layer Meteorol.* **24**:89–107 (1982).
54. M. S. Belen'kiy, V. V. Boronyev, N. Ts. Gomboyev, and V. L. Mironov, *Sounding of Atmospheric Turbulence*, Nauka, Novosibirsk, p. 114, 1986.
55. A. A. M. Holtzlag and A. P. Van Ulden, "A Simple Scheme for Daytime Estimates of the Surface Fluxes from Routine Weather Data," *J. Clim. Appl. Meteorol.* **22**:517–529 (1983).
56. T. Thiermann and A. Kohnle, "A Simple Model for the Structure Constant of Temperature Fluctuations in the Lower Atmosphere," *J. Phys. D: Appl. Phys.* **21**:S37–S40 (1988).
57. E. A. Andreas, "Estimating C_n^2 over Snow and Sea Ice from Meteorological Data," *J. Opt. Soc. Am. A* **5**:481–494 (1988).
58. J. L. Bufton, P. O. Minott, and M. W. Fitzmaurice, "Measurements of Turbulence Profiles in the Troposphere," *J. Opt. Soc. Am.* **62**:1068–1070 (1972).
59. F. W. Eaton, W. A. Peterson, J. R. Hines, K. R. Peterman, R. E. Good, R. R. Beland, and J. W. Brown, "Comparisons of VHF Radar, Optical, and Temperature Fluctuation Measurements of C_n^2 , r_0 , and Θ_0 ," *Theor. Appl. Climatol.* **39**:17–29 (1988).
60. F. Dalaudier, M. Crochet, and C. Sidi, "Direct Comparison between in situ and Radar Measurements of Temperature Fluctuation Spectra: A Puzzling Result," *Radio Sci.* **24**:311–324 (1989).
61. D. W. Beran, W. H. Hooke, and S. F. Clifford, "Acoustic Echo-Sounding Techniques and Their Application to Gravity-Wave, Turbulence, and Stability Studies," *Boundary-Layer Meteorol.* **4**:133–153 (1973).
62. M. Fukushima, K. Akita, and H. Tanaka, "Night-Time Profiles of Temperature Fluctuations Deduced from Two-Year Solar Observation," *J. Meteorol. Soc. Jpn.* **53**:487–491 (1975).

63. D. N. Asimakopoulis, R. S. Cole, S. J. Caughey, and B. A. Crease, "A Quantitative Comparison between Acoustic Sounder Returns and the Direct Measurement of Atmospheric Temperature Fluctuations," *Boundary-Layer Meteorol.* **10**:137–147 (1976).
64. T. E. VanZandt, J. L. Green, K. S. Gage, and W. L. Clark, "Vertical Profiles of Refractivity Turbulence Structure Constant: Comparison of Observations by the Sunset Radar with a New Theoretical Model," *Radio Sci.* **13**:819–829 (1978).
65. K. S. Gage and B. B. Balsley, "Doppler Radar Probing of the Clear Atmosphere," *Bull. Am. Meteorol. Soc.* **59**:1074–1093 (1978).
66. R. B. Chadwick and K. P. Moran, "Long-Term measurements of C_n^2 in the Boundary Layer," *Radio Sci.* **15**:355–361 (1980).
67. B. B. Balsley and V. L. Peterson, "Doppler-Radar Measurements of Clear Air Turbulence at 1290 MHz," *J. Appl. Meteorol.* **20**:266–274 (1981).
68. E. E. Gossard, R. B. Chadwick, T. R. Detman, and J. Gaynor, "Capability of Surface-Based Clear-Air Doppler Radar for Monitoring Meteorological Structure of Elevated Layers," *J. Clim. Appl. Meteorol.* **23**:474 (1984).
69. G. D. Nastrom, W. L. Ecklund, K. S. Gage, and R. G. Strauch, "The Diurnal Variation of Backscattered Power from VHF Doppler Radar Measurements in Colorado and Alaska," *Radio Sci.* **20**:1509–1517 (1985).
70. D. L. Fried, "Remote Probing of the Optical Strength of Atmospheric Turbulence and of Wind Velocity," *Proc. IEEE* **57**:415–420 (1969).
71. J. W. Strohbehn, "Remote Sensing of Clear-air Turbulence," *J. Opt. Soc. Am.* **60**:948 (1970).
72. J. Vernin and F. Roddier, "Experimental Determination of Two-Dimensional Power Spectra of Stellar Light Scintillation. Evidence for a Multilayer Structure of the Air Turbulence in the Upper Troposphere," *J. Opt. Soc. Am.* **63**:270–273 (1973).
73. G. R. Ochs, T. Wang, R. S. Lawrence, and S. F. Clifford, "Refractive Turbulence Profiles Measured by One-Dimensional Spatial Filtering of Scintillations," *Appl. Opt.* **15**:2504–2510 (1976).
74. R. E. Hufnagel and N. R. Stanley, "Modulation Transfer Function Associated with Image Transmission through Turbulent Media," *J. Opt. Soc. Am.* **54**:52–61 (1964).
75. R. E. Hufnagel, "Variations of Atmospheric Turbulence," in *Technical Digest of Topical Meeting on Optical Propagation through Turbulence*, Optical Society of America, Washington, D.C., 1974.
76. R. J. Sasiela, *A Unified Approach to Electromagnetic Wave Propagation in Turbulence and the Evaluation of Multiparameter Integrals*, Technical Report 807, MIT Lincoln Laboratory, Lexington, 1988.
77. V. A. Banakh and V. L. Mironov, *Lidar in a Turbulent Atmosphere*, Artech House, Boston, 1987.
78. C. W. Fairall and R. Markson, "Aircraft Measurements of Temperature and Velocity Microturbulence in the Stably Stratified Free Troposphere," *Proceedings of the Seventh Symposium on Turbulence and Diffusion*, November 12–15, Boulder, Co. 1985.
79. J. C. Kaimal, *The Atmospheric Boundary Layer—Its Structure and Measurement*, Indian Institute of Tropical Meteorology, Pune, 1988.
80. J. Barat and F. Bertin, "On the Contamination of Stratospheric Turbulence Measurements by Wind Shear," *J. Atmos. Sci.* **41**:819–827 (1984).
81. A. Ziad, R. Conan, A. Tokovinin, F. Martin, and J. Borgnino, "From the Grating Scale Monitor to the Generalized Seeing Monitor," *Appl. Opt.* **39**:5415–5425 (2000).
82. A. Ziad, M. Schöck, G. A. Chanan, M. Troy, R. Dekany, B. F. Lane, J. Borgnino, and F. Martin, "Comparison of Measurements of the Outer Scale of Turbulence by Three Different Techniques," *Appl. Opt.* **43**:2316–2324 (2004).
83. L. A. Chernov, *Wave Propagation in a Random Medium*, Dover, New York, p. 26, 1967.
84. P. Beckmann, "Signal Degeneration in Laser Beams Propagated through a Turbulent Atmosphere," *Radio Sci.* **69D**:629–640 (1965).
85. T. Chiba, "Spot Dancing of the Laser Beam Propagated through the Atmosphere," *Appl. Opt.* **10**:2456–2461 (1971).
86. J. H. Churnside and R. J. Lataitis, "Wander of an Optical Beam in the Turbulent Atmosphere," *Appl. Opt.* **29**:926–930 (1990).
87. G. A. Andreev and E. I. Gelfer, "Angular Random Walks of the Center of Gravity of the Cross Section of a Diverging Light Beam," *Radiophys. Quantum Electron.* **14**:1145–1147 (1971).

88. M. A. Kallistratova and V. V. Pokasov, "Defocusing and Fluctuations of the Displacement of a Focused Laser Beam in the Atmosphere," *Radiophys. Quantum Electron.* **14**:940–945 (1971).
89. J. A. Dowling and P. M. Livingston, "Behavior of Focused Beams in Atmospheric Turbulence: Measurements and Comments on the Theory," *J. Opt. Soc. Am.* **63**:846–858 (1973).
90. J. R. Dunphy and J. R. Kerr, "Turbulence Effects on Target Illumination by Laser Sources: Phenomenological Analysis and Experimental Results," *Appl. Opt.* **16**:1345–1358 (1977).
91. V. I. Klyatskin and A. I. Kon, "On the Displacement of Spatially Bounded Light Beams in a Turbulent Medium in the Markovian-Random-Process Approximation," *Radiophys. Quantum Electron.* **15**:1056–1061 (1972).
92. A. I. Kon, V. L. Mironov, and V. V. Nosov, "Dispersion of Light Beam Displacements in the Atmosphere with Strong Intensity Fluctuations," *Radiophys. Quantum Electron.* **19**:722–725 (1976).
93. V. L. Mironov and V. V. Nosov, "On the Theory of Spatially Limited Light Beam Displacements in a Randomly Inhomogeneous Medium," *J. Opt. Soc. Am.* **67**:1073–1080 (1977).
94. R. F. Lutmirski and H. T. Yura, "Propagation of a Finite Optical Beam in an Inhomogeneous Medium," *Appl. Opt.* **10**:1652–1658 (1971).
95. R. F. Lutmirski and H. T. Yura, "Wave Structure Function and Mutual Coherence Function of an Optical Wave in a Turbulent Atmosphere," *J. Opt. Soc. Am.* **61**:482–487 (1971).
96. H. T. Yura, "Atmospheric Turbulence Induced Laser Beam Spread," *Appl. Opt.* **10**:2771–2773 (1971).
97. H. T. Yura, "Mutual Coherence Function of a Finite Cross Section Optical Beam Propagating in a Turbulent Medium," *Appl. Opt.* **11**:1399–1406 (1972).
98. H. T. Yura, "Optical Beam Spread in a Turbulent Medium: Effect of the Outer Scale of Turbulence," *J. Opt. Soc. Am.* **63**:107–109 (1973).
99. H. T. Yura, "Short-Term Average Optical-Beam Spread in a Turbulent Medium," *J. Opt. Soc. Am.* **63**:567–572 (1973).
100. M. T. Tavis and H. T. Yura, "Short-Term Average Irradiance Profile of an Optical Beam in a Turbulent Medium," *Appl. Opt.* **15**:2922–2931 (1976).
101. R. L. Fante, "Electromagnetic Beam Propagation in Turbulent Media," *Proc. IEEE* **63**:1669–1692 (1975).
102. R. L. Fante, "Electromagnetic Beam Propagation in Turbulent Media: An Update," *Proc. IEEE* **68**:1424–1443 (1980).
103. G. C. Valley, "Isoplanatic Degradation of Tilt Correction and Short-Term Imaging Systems," *Appl. Opt.* **19**:574–577 (1980).
104. H. J. Breaux, *Correlation of Extended Huygens-Fresnel Turbulence Calculations for a General Class of Tilt Corrected and Uncorrected Laser Apertures*, Interim Memorandum Report No. 600, U.S. Army Ballistic Research Laboratory, 1978.
105. D. M. Cordray, S. K. Searles, S. T. Hanley, J. A. Dowling, and C. O. Gott, "Experimental Measurements of Turbulence Induced Beam Spread and Wander at 1.06, 3.8, and 10.6 μm ," *Proc. SPIE* **305**:273–280 (1981).
106. S. K. Searles, G. A. Hart, J. A. Dowling, and S. T. Hanley, "Laser Beam Propagation in Turbulent Conditions," *Appl. Opt.* **30**:401–406 (1991).
107. J. H. Churnside and R. J. Lataitis, "Angle-of-Arrival Fluctuations of a Reflected Beam in Atmospheric Turbulence," *J. Opt. Soc. Am. A* **4**:1264–1272 (1987).
108. D. L. Fried, "Optical Resolution through a Randomly Inhomogeneous Medium for Very Long and Very Short Exposures," *J. Opt. Soc. Am.* **56**:1372–1379 (1966).
109. R. F. Lutmirski, W. L. Woodie, and R. G. Buser, "Turbulence-Degraded Beam Quality: Improvement Obtained with a Tilt-Correcting Aperture," *Appl. Opt.* **16**:665–673 (1977).
110. D. L. Fried, "Statistics of a Geometrical Representation of Wavefront Distortion," *J. Opt. Soc. Am.* **55**:1427–1435 (1965); **56**:410 (1966).
111. D. M. Chase, "Power Loss in Propagation through a Turbulent Medium for an Optical-Heterodyne System with Angle Tracking," *J. Opt. Soc. Am.* **56**:33–44 (1966).
112. J. H. Churnside and C. M. McIntyre, "Partial Tracking Optical Heterodyne Receiver Arrays," *J. Opt. Soc. Am.* **68**:1672–1675 (1978).
113. V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation*, Israel Program for Scientific Translations, Jerusalem, 1971.
114. R. W. Lee and J. C. Harp, "Weak Scattering in Random Media, with Applications to Remote Probing," *Proc. IEEE* **57**:375–406 (1969).

115. R. S. Lawrence and J. W. Strohbehn, "A Survey of Clear-Air Propagation Effects Relevant to Optical Communications," *Proc. IEEE* **58**:1523–1545 (1970).
116. S. F. Clifford, "The Classical Theory of Wave Propagation in a Turbulent Medium," in J. W. Strohbehn (ed.), *Laser Beam Propagation in the Atmosphere*, Springer-Verlag, New York, pp. 9–43, 1978.
117. A. I. Kon and V. I. Tatarskii, "Parameter Fluctuations of a Space-Limited Light Beam in a Turbulent Atmosphere," *Izv. VUZ Radiofiz.* **8**:870–875 (1965).
118. R. A. Schmeltzer, "Means, Variances, and Covariances for Laser Beam Propagation through a Random Medium," *Quart. J. Appl. Math.* **24**:339–354 (1967).
119. D. L. Fried and J. B. Seidman, "Laser-Beam Scintillation in the Atmosphere," *J. Opt. Soc. Am.* **57**:181–185 (1967).
120. D. L. Fried, "Scintillation of a Ground-to-Space Laser Illuminator," *J. Opt. Soc. Am.* **57**:980–983 (1967).
121. Y. Kinoshita, T. Asakura, and M. Suzuki, "Fluctuation Distribution of a Gaussian Beam Propagating through a Random Medium," *J. Opt. Soc. Am.* **58**:798–807 (1968).
122. A. Ishimaru, "Fluctuations of a Beam Wave Propagating through a Locally Homogeneous Medium," *Radio Sci.* **4**:295–305 (1969).
123. A. Ishimaru, "Fluctuations of a Focused Beam Wave for Atmospheric Turbulence Probing," *Proc. IEEE* **57**:407–414 (1969).
124. A. Ishimaru, "The Beam Wave Case and Remote Sensing," in J. W. Strohbehn (ed.), *Laser Beam Propagation in the Atmosphere*, Springer-Verlag, New York, pp. 129–170, 1978.
125. P. J. Titterton, "Scintillation and Transmitter-Aperture Averaging over Vertical Paths," *J. Opt. Soc. Am.* **63**:439–444 (1973).
126. H. T. Yura and W. G. McKinley, "Optical Scintillation Statistics for IR Ground-to-Space Laser Communication Systems," *Appl. Opt.* **22**:3353–3358 (1983).
127. P. A. Lightsey, J. Anspach, and P. Sydney, "Observations of Uplink and Retroreflected Scintillation in the Relay Mirror Experiment," *Proc. SPIE* **1482**:209–222 (1991).
128. D. L. Fried, "Aperture Averaging of Scintillation," *J. Opt. Soc. Am.* **57**:169–175 (1967).
129. J. H. Churnside, "Aperture Averaging of Optical Scintillations in the Turbulent Atmosphere," *Appl. Opt.* **30**:1982–1994 (1991).
130. J. H. Churnside and R. G. Frehlich, "Experimental Evaluation of Log-Normally Modulated Rician and IK Models of Optical Scintillation in the Atmosphere," *J. Opt. Soc. Am. A* **6**:1760–1766 (1989).
131. G. Parry and P. N. Pusey, "K Distributions in Atmospheric Propagation of Laser Light," *J. Opt. Soc. Am.* **69**:796–798 (1979).
132. G. Parry, "Measurement of Atmospheric Turbulence Induced Intensity Fluctuations in a Laser Beam," *Opt. Acta* **28**:715–728 (1981).
133. W. A. Coles and R. G. Frehlich, "Simultaneous Measurements of Angular Scattering and Intensity Scintillation in the Atmosphere," *J. Opt. Soc. Am.* **72**:1042–1048 (1982).
134. J. H. Churnside and S. F. Clifford, "Lognormal Rician Probability-Density Function of Optical Scintillations in the Turbulent Atmosphere," *J. Opt. Soc. Am. A* **4**:1923–1930 (1987).
135. R. Dashen, "Path Integrals for Waves in Random Media," *J. Math. Phys.* **20**:894–920 (1979).
136. K. S. Gochelashvily and V. I. Shishov, "Multiple Scattering of Light in a Turbulent Medium," *Opt. Acta* **18**:767–777 (1971).
137. S. M. Flatté and G. Y. Wang, "Irradiance Variance of Optical Waves Through Atmospheric Turbulence by Numerical Simulation and Comparison with Experiment," *J. Opt. Soc. Am. A* **10**:2363–2370 (1993).
138. R. Frehlich, "Simulation of Laser Propagation in a Turbulent Atmosphere," *Appl. Opt.* **39**:393–397 (2000).
139. S. M. Flatté and J. S. Gerber, "Irradiance-Variance Behavior by Numerical Simulation for Plane-Wave and Spherical-Wave Optical Propagation through Strong Turbulence," *J. Opt. Soc. Am. A* **17**:1092–1097 (2000).
140. R. Rao, "Statistics of the Fractal Structure and Phase Singularity of a Plane Light Wave Propagation in Atmospheric Turbulence," *Appl. Opt.* **47**:269–276 (2008).
141. K. S. Gochelashvily, V. G. Pevgov, and V. I. Shishov, "Saturation of Fluctuations of the Intensity of Laser Radiation at Large Distances in a Turbulent Atmosphere (Fraunhofer Zone of Transmitter)," *Sov. J. Quantum Electron.* **4**:632–637 (1974).
142. A. M. Prokhorov, F. V. Bunkin, K. S. Gochelashvily, and V. I. Shishov, "Laser Irradiance Propagation in Turbulent Media," *Proc. IEEE* **63**:790–810 (1975).

143. R. L. Fante, "Inner-Scale Size Effect on the Scintillations of Light in the Turbulent Atmosphere," *J. Opt. Soc. Am.* **73**:277–281 (1983).
144. R. G. Frehlich, "Intensity Covariance of a Point Source in a Random Medium with a Kolmogorov Spectrum and an Inner Scale of Turbulence," *J. Opt. Soc. Am. A.* **4**:360–366 (1987).
145. R. J. Hill and R. G. Frehlich, "Probability Distribution of Irradiance for the Onset of Strong Scintillation," *J. Opt. Soc. Am. A.* **14**:1530–1540 (1997).
146. J. H. Churnside and R. J. Hill, "Probability Density of Irradiance Scintillations for Strong Path-Integrated Refractive Turbulence," *J. Opt. Soc. Am. A.* **4**:727–733 (1987).
147. F. S. Vetelino, C. Young, L. Andrews, and J. Reclons, "Aperture Averaging Effects on the Probability Density of Irradiance Fluctuations in Moderate-to-Strong Turbulence," *Appl. Opt.* **46**: 2099–2108 (2007).
148. M. A. Al-Habash, L. C. Andrews, and R. L. Phillips, "Mathematical Model for the Irradiance PDF of a Laser Beam Propagating through Turbulent Media," *Opt. Eng.* **40**:1554–1562 (2001).
149. D. K. Killinger and A. Mooradian (eds.), *Optical and Laser Remote Sensing*, Springer-Verlag, New York, Optical Sciences, vol. 39, 1983.
150. L. J. Radziemski, R. W. Solarz, and J. A. Paisner (eds.), *Laser Spectroscopy and Its Applications*, Marcel Dekker, New York, Optical Eng., vol. 11, 1987.
151. T. Kobayashi, "Techniques for Laser Remote Sensing of the Environment," *Remote Sensing Reviews*, **3**:1–56 (1987).
152. E. D. Hinkley (ed.), *Laser Monitoring of the Atmosphere*, Springer-Verlag, Berlin, 1976.
153. W. B. Grant and R. T. Menzies, "A Survey of Laser and Selected Optical Systems for Remote Measurement of Pollutant Gas Concentrations," *APCA Journal* **33**:187 (1983).
154. "Optical Remote Sensing of the Atmosphere," *Conf. Proceedings, OSA Topical Meeting*, Williamsburg, 1991.
155. Dennis K. Killinger, "Lidar and Laser Remote Sensing," *Handbook of Vibrational Spectroscopy*, John Wiley & Sons, Chichester, 2002.
156. Claus Weitkamp, (ed.), *Lidar: Range Resolved Optical Remote Sensing of the Atmosphere*, Springer-Verlag, New York, 2005.
157. P. B. Hays, V. J. Abreu, D. A. Gell, H. J. Grassl, W. R. Skinner, and M. E. Dobbs, "The High Resolution Doppler Imager on the Upper Atmospheric Research Satellite," *J. Geophys. Res. (Atmosphere)* **98**:10713 (1993).
158. P. B. Hays, V. J. Abreu, M. D. Burrage, D. A. Gell, A. R. Marshall, Y. T. Morton, D. A. Ortland, W. R. Skinner, D. L. Wu, and J. H. Yee, "Remote Sensing of Mesospheric Winds with the High Resolution Imager," *Planet. Space Sci.* **40**:1599 (1992).
159. E. Browell, "Differential Absorption Lidar Sensing of Ozone," *Proc. IEEE* **77**:419 (1989).
160. J. M. Intrieri, A. J. Dedard, and R. M. Hardesty, "Details of Colliding Thunderstorm Outflow as Observed by Doppler Lidar," *J. Atmospheric Sciences* **47**:1081 (1990).
161. R. M. Hardesty, K. Elmore, M. E. Jackson, in *21st Conf. on Radar Meteorology*, American Meteorology Society, Boston, 1983.
162. M. P. McCormick and R. E. Veiga, "Initial Assessment of the Stratospheric and Climatic Impact of the 1991 Mount Pinatubo Eruption—Prolog," *Geophysical Research Lett.* **19**:155 (1992).
163. D. K. Killinger, S.D. Allen, R.D. Waterbury, C. Stefano, and E. L. Dottery, "Enhancement of Nd:YAG LIBS Emission of a Remote Target Using a Simultaneous CO₂ Laser Pulse," *Optics Express* **15**:12905 (2007).
164. A. Miziolek, V. Palleschi, and I. Schechter, (eds.), *Laser Induced Spectroscopy*, Cambridge University Press, Cambridge, 2006.
165. J. Kasparian, R. Sauerbrey, and S.L. Chin, "The Critical Laser Intensity of Self-Guided Light Filaments in Air," *Appl. Phys. B* **71**:877 (2000).
166. R. A. R. Tricker, *Introduction to Meteorological Optics*, American Elsevier, New York, 1970.
167. R. Greenler, *Rainbows, Halos, and Glories*, Cambridge University Press, Cambridge, 1980.
168. P. Forster, V. Ramaswamy, P. Artaxo, T. Bernsten, R. Betts, D.W. Fahey, J. Haywood, et al. "Changes in Atmospheric Constituents and in Radiative Forcing," In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the IV Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (eds.), Cambridge University Press, 2007. Online at <http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-chapter2.pdf>.

This page intentionally left blank

IMAGING THROUGH ATMOSPHERIC TURBULENCE

Virendra N. Mahajan*

*The Aerospace Corporation
El Segundo, California*

Guang-ming Dai

*Laser Vision Correction Group
Advanced Medical Optics
Milpitas, California*

ABSTRACT

In this chapter, how the random aberrations introduced by atmospheric turbulence degrade the image formed by a ground-based telescope with an annular pupil is considered. The results for imaging with a circular pupil are obtained as a special case of the annular pupil. Both the long- and short-exposure images are discussed in terms of their Strehl ratio, point-spread function (PSF), and transfer function. The discussion given is equally applicable to laser beams propagating through turbulence. An atmospheric coherence length is defined and it is shown that, for fixed power of a beam and regardless of the size of its diameter, the central irradiance in the focal plane is smaller than the corresponding aberration-free value for a beam of diameter equal to that of the coherence length. The aberration function is decomposed into Zernike annular polynomials and the autocorrelation and crosscorrelations of the expansion coefficients are given for Kolmogorov turbulence. It is shown that the aberration variance increases with the obscuration ratio of the annular pupil. The angle of arrival is also discussed, both in terms of the wavefront tilt as well as the centroid of the aberrated PSF. It is shown that the difference between the two is small, and the obscuration has only a second-order effect.

4.1 GLOSSARY

a	outer radius of the pupil
a_j	expansion coefficients
$A_L(\vec{r}_p)$	amplitude function of the lens (or imaging system) at a pupil point with position vector $\vec{r}_p$
C_n^2	refractive index structure parameter
$\mathcal{D}$	structure function ($\mathcal{D}_w$ —wave, $\mathcal{D}_\phi$ —phase, $\mathcal{D}_I$ —log amplitude, $\mathcal{D}_n$ —refractive index)
D	diameter of exit pupil or aperture
F	focal ratio of the image-forming light cone ($F = R/D$)

*The author is also an adjunct professor at the College of Optical Sciences at the University of Arizona, Tucson and the Department of Optics and Photonics, National Central University, Chung Li, Taiwan.

$I_i(\vec{r}_i)$	irradiance at a point $\vec{r}_i$ in the image plane
$I(r)$	normalized irradiance in the image plane such that its aberration-free central value is $I(0) = 1$
$\langle I_0 \rangle$	time-averaged irradiance at the exit pupil
j	Zernike aberration mode number
J	number of Zernike aberration modes
$l(\vec{r}_p)$	log-amplitude function introduced by atmospheric turbulence
L	path length through atmosphere or from source to receiver
$n(\vec{r})$	fluctuating part of refractive index $N(\vec{r})$
$P(r_c)$	encircled power in a circle of radius r_c in the image plane
P_{ex}	power in the exit pupil
$P_L(\vec{r}_p)$	lens pupil function
$P_R(\vec{r}_p)$	complex amplitude variation introduced by atmospheric turbulence
r_0	Fried's atmospheric coherence length
R	radius of curvature of the reference sphere with respect to which the aberration is defined
$R_n^m(\rho)$	Zernike circle radial polynomial of degree n and azimuthal frequency m
$\langle S \rangle$	time-averaged Strehl ratio
S_a	coherent area $\pi r_0^2/4$ of atmospheric turbulence
$\langle S_t \rangle$	tilt-corrected time-averaged Strehl ratio
S_{ex}	area of exit pupil
$Z_n^m(\rho, \theta)$	Zernike circle polynomial of degree n and azimuthal frequency m
ϵ	obscuration ratio of an annular pupil
Δ_j	phase aberration variance after correcting $J = j$ aberration modes
η	central irradiance in the image plane normalized by the aberration-free value for a pupil with area S_a but containing the same total power
λ	wavelength of object radiation
$\vec{v}_i$	spatial frequency vector in the image plane
$\vec{v}$	normalized spatial frequency vector
ϑ_0	isoplanatic angle of turbulence
$\sigma_\alpha, \sigma_\beta$	tip and tilt angle standard deviations
σ_Φ^2	phase aberration variance
$\tau(\vec{v})$	optical transfer function
ρ	radial variable normalized by the pupil radius a
$\tau_a(\nu)$	long-exposure (LE) atmospheric MTF reduction factor
$\Phi(\vec{r}_p)$	phase aberration
$\Phi_R(\vec{r}_p)$	random phase aberration introduced by atmospheric turbulence

4.2 INTRODUCTION

The resolution of a telescope forming an aberration-free image is determined by its diameter D ; larger the diameter, better the resolution. However, in ground-based astronomy, the resolution is degraded considerably because of the aberrations introduced by atmospheric turbulence. A plane wave of uniform amplitude and phase representing the light from a star propagating through the atmosphere undergoes both amplitude and phase variations due to the random inhomogeneities

in its refractive index. The amplitude variations, called scintillations, result in the twinkling of stars. The purpose of a large ground-based telescope has therefore generally not been better resolution but to collect more light so that dim objects may be observed. Of course, with the advent of adaptive optics,¹⁻³ the resolution can be improved by correcting the phase aberrations with a deformable mirror. The amplitude variations are negligible in near-field imaging, that is, when the far-field distance $D^2/\lambda \gg L$ or $D \gg \sqrt{\lambda L}$, where λ is the wavelength of the starlight and L is the propagation path length through the turbulence.⁴ In principle, a diffraction-limited image can be obtained if the aberrations are corrected completely in real time by the deformable mirror. However, in far-field imaging, that is, when $D^2/\lambda \ll L$ or $D \ll \sqrt{\lambda L}$, significant amplitude variations are introduced in addition to the phase variations. Since the amplitude variations cannot be corrected by adaptive optics, even a complete correction of the phase errors does not yield a diffraction-limited image.

In this chapter, we consider the effect of random aberrations introduced by atmospheric turbulence on the image quality. The discussion given is applicable equally to laser beams propagating through turbulence. First, we derive expressions for the degraded time-averaged point-spread function (PSF) and optical transfer function (OTF). The pupil function of the overall imaging system is written as the product of the pupil function of the optical system and a complex amplitude factor introduced by turbulence. As a result, the time-averaged OTF of the overall system is also equal to the product of the OTF of the optical system and an OTF reduction factor representing the mutual coherence function of the wave propagating through the turbulent atmosphere. The time-averaged images thus obtained are referred to as the *long-exposure images*.⁴ The exposure time may be 1 to 10 s.

Next, the structure functions for refractive index and phase fluctuations are given for the *Kolmogorov turbulence*. We introduce the notion of *atmospheric coherence length* r_0 , and show that it limits the resolution of the image regardless of how large the telescope diameter is.⁴ Since large telescopes have annular pupils, we consider systems with such pupils and describe the Strehl ratio, PSF, and encircled power as a function of the ratio of the pupil diameter and the coherence length. The phase aberration introduced by Kolmogorov turbulence is expanded in terms of the Zernike annular polynomials,⁵⁻⁸ and autocorrelation and crosscorrelation of the expansion coefficients are given.⁹ It is shown that atmospheric turbulence dominates the degradation of the system modulation transfer function (MTF), Strehl ratio, or the angle of arrival; the effect of the pupil obscuration is small even for weak turbulence. The piston-removed aberration variance increases monotonically as the obscuration of the pupil increases. A large portion (87 percent for a circular pupil and somewhat larger for an annular pupil) of the aberration variance is a random wavefront tilt, resulting in a random displacement of the image over a long exposure. Hence, a better quality image is obtained if the exposure time is short enough so that the image does not wander. Such an image is referred to as a *short-exposure image*.⁴ The exposure time may be 0.1 s or less. The location of such an image is random which is not relevant in astronomy. The image wander, sometimes referred to as the *angle of arrival*, is considered in terms of the wavefront tilt (called the Z-tilt) or the *centroid* of the image (called the G-tilt), and expressions are given for their standard deviations.

The fluctuating image can be stabilized if the tilt is corrected in real time with a steering mirror. A tilt-corrected image is equivalent to a time-averaged short-exposure image. The characteristics of such an image are considered and compared with those of a long-exposure image. The angular resolution of the system for a long-exposure image is about λ/r_0 compared to the diffraction-limited resolution λ/D . For a short-exposure image the resolution is substantially better than that for a long-exposure image. When the coherence length is much smaller than the pupil diameter, the image breaks up into small spots, called *speckles*, whose size depends on the latter, while the overall size of the image is determined by the former. An example of a speckled short-exposure image of a point object is illustrated and compared with the corresponding aberration-free image.

4.3 LONG-EXPOSURE IMAGE

When the medium between an object and the optical system imaging it is homogeneous, a spherical wavefront of uniform amplitude centered on an object point is incident on the entrance pupil of the system. If the system is aberration free and has uniform transmittance, a spherical wavefront of

uniform amplitude centered at the Gaussian image point emerges from its exit pupil. However, if the system is aberrated and has nonuniform transmittance, then the emerging wavefront is distorted with a nonuniform amplitude. Let $\Phi_L(\vec{r}_p)$ and $A_L(\vec{r}_p)$ be the phase and amplitude transmittance corresponding to a point $\vec{r}_p$ on the exit pupil. A dimensionless pupil function of the system may be written

$$P_L(\vec{r}_p) = A_L(\vec{r}_p) \exp[i\Phi_L(\vec{r}_p)] \quad (1)$$

When the medium between the object and the imaging system is inhomogeneous as in ground-based astronomy, it introduces random phase and amplitude variations across the wavefront propagating through it. Let $\Phi_R(\vec{r}_{en})$ and $\ell(\vec{r}_{en})$ be the random phase and (dimensionless) log amplitude introduced by atmospheric turbulence at a point $\vec{r}_{en}$ on the entrance pupil. As the wave propagates through the imaging system, it undergoes additional phase and amplitude variations. The total phase aberration at a point $\vec{r}_p$ on the exit pupil may be written

$$\Phi(\vec{r}_p) = \Phi_L(\vec{r}_p) + \Phi_R(\vec{r}_p) \quad (2)$$

where the position vector $\vec{r}_p$ is related to the position vector $\vec{r}_{en}$ by the pupil magnification m_p (i.e., $\vec{r}_p = m_p \vec{r}_{en}$). The pupil function of the overall system (i.e., including the effects of atmospheric turbulence) representing the wavefront at the exit pupil may be written

$$P(\vec{r}_p) = P_L(\vec{r}_p) P_R(\vec{r}_p) \quad (3)$$

where

$$P_R(\vec{r}_p) = \exp[\ell(\vec{r}_p) + i\Phi_R(\vec{r}_p)] \quad (4)$$

is truncated by the pupil function $P_L(\vec{r}_p)$ and represents the complex amplitude variation introduced by turbulence.

The instantaneous irradiance distribution of the star image formed by the overall system is given by^{10,11}

$$I_i(\vec{r}_i) = \frac{P_{ex}}{S_{ex} \lambda^2 R^2} \left| \int P(\vec{r}_p) \exp\left(-\frac{2\pi i}{\lambda R} \vec{r}_p \cdot \vec{r}_i\right) d\vec{r}_p \right|^2 \quad (5)$$

where $\vec{r}_i$ is the position vector of a point in the image plane, P_{ex} is the total power in the image, S_{ex} is the area of the exit pupil, λ is the wavelength of object radiation, and R is the radius of curvature of the reference sphere with respect to which the aberration is defined.

Substituting Eq. (3) into Eq. (5), the time-averaged distribution, representing a long-exposure image, may be written

$$\begin{aligned} \langle I_i(\vec{r}_i) \rangle &= \frac{\langle P_{ex} \rangle}{S_{ex} \lambda^2 R^2} \iint P_L(\vec{r}_p) P_L^*(\vec{r}'_p) \langle P_R(\vec{r}_p) P_R^*(\vec{r}'_p) \rangle \\ &\quad \times \exp\left[-\frac{2\pi i}{\lambda R} (\vec{r}_p - \vec{r}'_p) \cdot \vec{r}_i\right] d\vec{r}_p d\vec{r}'_p \end{aligned} \quad (6)$$

where * denotes a complex conjugate. Thus, considering Eq. (4), we may write the *mutual coherence function* (MCF) of the wave incident on the system, using abbreviated notation

$$\begin{aligned} \langle P_R(\vec{r}_p) P_R^*(\vec{r}'_p) \rangle &\equiv \langle P_1 P_2^* \rangle \\ &= \langle \exp[(\ell_1 + \ell_2) + i(\Phi_1 - \Phi_2)] \rangle \end{aligned} \quad (7)$$

If the refractive index fluctuations are statistically stationary, that is, they are statistically homogeneous and isotropic, then so are the fluctuations in ℓ and Φ which they generate. Therefore,

$$\langle (\ell_1 + \ell_2) (\Phi_1 - \Phi_2) \rangle = \langle \ell_1 \Phi_1 \rangle - \langle \ell_2 \Phi_2 \rangle + \langle \ell_2 \Phi_1 \rangle - \langle \ell_1 \Phi_2 \rangle \quad (8a)$$

$$= 0 \quad (8b)$$

where the two averaged quantities in the first term on the right-hand side of Eq. (8a) cancel each other due to homogeneity of turbulence, and those in the second term cancel due to its isotropy. Thus, $\ell_1 + \ell_2$ and $\Phi_1 - \Phi_2$ are uncorrelated random variables. Hence, Eq. (7) may be written

$$\langle P_R(\vec{r}_p) P_R^*(\vec{r}_p') \rangle = \langle \exp(\ell_1 + \ell_2) \rangle \langle \exp[i(\Phi_1 - \Phi_2)] \rangle \quad (9)$$

For a Gaussian random variable x with a mean value of $\langle x \rangle$ and a standard deviation of σ , it is easy to show that

$$\begin{aligned} \langle \exp(bx) \rangle &= \frac{1}{2\pi\sigma} \int_{-\infty}^{\infty} \exp(bx) \exp[-(x - \langle x \rangle)^2 / 2\sigma^2] dx \\ &= \exp \left[\frac{1}{2} b^2 \langle (x - \langle x \rangle)^2 \rangle + b \langle x \rangle \right] \end{aligned} \quad (10)$$

where b is an arbitrary constant. As the wave propagates through the random atmosphere, it accumulates randomly both phase and amplitude variations. Each of these two random variables can be expressed as an integral along the propagation path. Since the atmospheric path length is much longer than the correlation length of these variable, the accumulation consists of a sum of a large number of terms that are statistically independent. By the central limit theorem, we can infer that they obey Gaussian statistics. Letting both ℓ and Φ_R be Gaussian random variables where Φ_R has a mean value of zero, it can be shown from conservation of power, that^{4,11}

$$\langle \exp(\ell_1 + \ell_2) \rangle = \exp \left[-\frac{1}{2} \mathcal{D}_\ell(|\vec{r}_p - \vec{r}_p'|) \right] \quad (11)$$

where

$$\mathcal{D}_\ell(|\vec{r}_p - \vec{r}_p'|) = \langle [\ell(\vec{r}_p) - \ell(\vec{r}_p')]^2 \rangle \quad (12)$$

is the *log-amplitude structure function*.

Since Φ_1 and Φ_2 are Gaussian random variables with zero mean values, $\Phi_1 - \Phi_2$ is also a Gaussian random variable. Hence, following Eq. (10), we obtain

$$\begin{aligned} \langle \exp[i(\Phi_1 - \Phi_2)] \rangle &= \exp \left[-\frac{1}{2} \langle (\Phi_1 - \Phi_2)^2 \rangle \right] \\ &= \exp \left[-\frac{1}{2} \mathcal{D}_\Phi(|\vec{r}_p - \vec{r}_p'|) \right] \end{aligned} \quad (13)$$

where

$$\mathcal{D}_\Phi(|\vec{r}_p - \vec{r}_p'|) = \langle [\Phi(\vec{r}_p) - \Phi(\vec{r}_p')]^2 \rangle \quad (14)$$

is the *phase structure function* of turbulence. Substituting Eqs. (11) and (13) into Eq. (9), we may write the MCF

$$\langle P_R(\vec{r}_p) P_R^*(\vec{r}_p') \rangle = \exp \left[-\frac{1}{2} \mathcal{D}_w(|\vec{r}_p - \vec{r}_p'|) \right] \quad (15)$$

where

$$\mathcal{D}_w(|\vec{r}_p - \vec{r}'_p|) = \mathcal{D}_\ell(|\vec{r}_p - \vec{r}'_p|) + \mathcal{D}_\Phi(|\vec{r}_p - \vec{r}'_p|) \quad (16)$$

is called the *wave structure function* of turbulence.

Substituting Eq. (15) into Eq. (6), we obtain

$$\begin{aligned} \langle I_i(\vec{r}_i) \rangle &= \frac{\langle P_{\text{ex}} \rangle}{S_{\text{ex}} \lambda^2 R^2} \iint P_L(\vec{r}_p) P_L^*(\vec{r}'_p) \exp \left[-\frac{1}{2} \mathcal{D}_w(|\vec{r}_p - \vec{r}'_p|) \right] \\ &\quad \times \exp \left[-\frac{2\pi i}{\lambda R} (\vec{r}_p - \vec{r}'_p) \cdot \vec{r}_i \right] d\vec{r}_p d\vec{r}'_p \end{aligned} \quad (17)$$

Because of a Fourier transform relationship between the PSF and the OTF, we identify $(\vec{r}_p - \vec{r}'_p)/\lambda R$ with a spatial frequency $\vec{v}_i$. Thus, we let

$$\vec{r}_p - \vec{r}'_p = \lambda R \vec{v}_i \quad (18)$$

Substituting Eq. (18) into Eq. (17) and carrying out the integration over $\vec{r}_p$, we may write

$$\langle I_i(\vec{r}_i) \rangle = \langle P_{\text{ex}} \rangle \int \tau_L(\vec{v}_i) \exp \left[-\frac{1}{2} \mathcal{D}_w(\lambda R \vec{v}_i) \right] \exp(-2\pi i \vec{v}_i \cdot \vec{r}_i) d\vec{v}_i \quad (19)$$

where

$$\tau_L(\vec{v}_i) = S_{\text{ex}}^{-1} \int P_L(\vec{r}_p) P_L^*(\vec{r}_p - \lambda R \vec{v}_i) d\vec{r}_p \quad (20)$$

is the OTF of the (turbulence-free) optical system and $v_i = |\vec{v}_i|$. Now, we introduce normalized quantities

$$I(r) = I_i(\vec{r}_i)/I(0) \quad (21a)$$

$$\vec{r} = \vec{r}_i/\lambda F \quad (21b)$$

and

$$\vec{v} = \vec{v}_i/(1/\lambda F) \quad (21c)$$

where

$$I(0) = \frac{\langle P_{\text{ex}} \rangle S_{\text{ex}}}{\lambda^2 R^2} \quad (22)$$

is the aberration-free central irradiance for a uniform-amplitude wavefront with a total power $\langle P_{\text{ex}} \rangle$ and $1/\lambda F$ is the cutoff spatial frequency of the optical system. Here, $F = R/D$ is the focal ratio of the image-forming light cone. In terms of the normalized quantities, Eq. (19) for the time-averaged PSF may be written

$$\langle I(\vec{r}) \rangle = (4/\pi) \int \langle \tau(\vec{v}) \rangle \exp(-2\pi i \vec{v} \cdot \vec{r}) d\vec{v} \quad (23)$$

where

$$\langle \tau(\vec{v}) \rangle = \tau_L(\vec{v}) \exp \left[-\frac{1}{2} \mathcal{D}_w(\nu D) \right] \quad (24)$$

is the time-averaged OTF. This OTF being equal to the product of the OTF of the optical system and the MCF associated with atmospheric turbulence is a consequence of the fact that the pupil function of the overall system is equal to the product of the pupil function of the optical system and the complex amplitude variation introduced by turbulence. It should be evident that, since $\nu D = \nu_i \lambda R$, the modification of the system OTF due to turbulence is independent of the diameter D of the exit pupil of the system. The MCF acts as a *reduction factor* for the OTF of the overall system.¹¹ It may also be referred to as the Hopkins ratio of the image if $\tau_L(\vec{v})$ is the aberration-free OTF.¹¹

4.4 KOLMOGOROV TURBULENCE AND ATMOSPHERIC COHERENCE LENGTH

The refractive index $N(\vec{r})$ of the turbulent atmosphere at a point $\vec{r}$ in space fluctuates due to fluctuations of its temperature. It can be written as the sum of its mean value $\langle N(\vec{r}) \rangle$ at a point $\vec{r}$ and a fluctuating part $n(\vec{r})$ at that point, in the form

$$N(\vec{r}) = \langle N(\vec{r}) \rangle + n(\vec{r}) \quad (25)$$

Whereas $\langle N(\vec{r}) \rangle \approx 1$, $n(\vec{r})$ is only on the order of 10^{-6} . It should be evident that $n(\vec{r})$ has a mean value $\langle n(\vec{r}) \rangle = 0$. The structure function $\mathcal{D}_n(\vec{r}_1, \vec{r}_2)$ of the refractive index fluctuations represents the mean square value of the difference of refractive index at two points $\vec{r}_1$ and $\vec{r}_2$, that is,

$$\mathcal{D}_n(\vec{r}_1, \vec{r}_2) = \langle [n(\vec{r}_1) - n(\vec{r}_2)]^2 \rangle \quad (26)$$

The turbulent atmosphere consists of packets of air called *eddies* each with a characteristic value of its refractive index. For Kolmogorov turbulence, the structural function is given by

$$\mathcal{D}_n(r) = C_n^2 r^{2/3}, \quad \ell_0 \ll r \ll L_0 \quad (27)$$

where $r = |\vec{r}_2 - \vec{r}_1|$ and C_n^2 (in units of $\text{m}^{-2/3}$) is called the *refractive index structure parameter*. The quantities ℓ_0 and L_0 are called the *inner* and *outer scales of turbulence* representing the smallest and the largest eddies, respectively. Typical values of C_n^2 vary from 10^{-13} $\text{m}^{-2/3}$ for strong turbulence to 10^{-17} $\text{m}^{-2/3}$ for weak turbulence. Values of ℓ_0 are on the order of a few millimeters, and those of L_0 vary from 1 to 100 m.

The wave structure function for a spherical wave propagating through Kolmogorov turbulence is given by¹²

$$\mathcal{D}_w(r) = 2.914 k^2 r^{5/3} \int_0^L C_n^2(z) (z/L)^{5/3} dz \quad (28)$$

where $k = 2\pi/\lambda$ and z varies along the atmospheric path of total length L from a value of zero at the source (or the object plane) to a value of L at the receiver (or the image plane). The wave structure function can also be written in the form^{4,12}

$$\mathcal{D}_w(r) = 6.88 (r/r_0)^{5/3} \quad (29)$$

where, for a spherical wave,

$$r_0 = 0.185 \lambda^{6/5} \left\{ \int_0^L C_n^2(z) (z/L)^{5/3} dz \right\}^{-3/5} \quad (30)$$

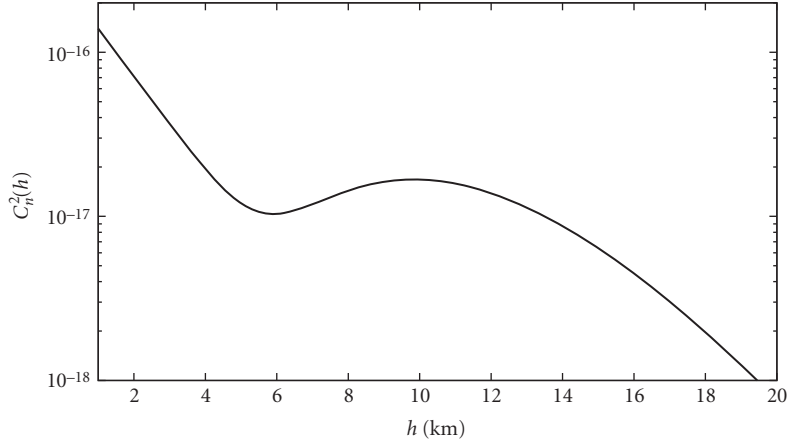


FIGURE 1 Variation of the refractive index structure parameter $C_n^2(h)$ with the altitude h for the Hufnagle-Valley model $H - V_{5/7}$ of Eq. (31).

is a characteristic length of Kolmogorov turbulence called its *coherence length* or *diameter*, often referred to as *Fried's coherence length*. We note that r_0 varies with the optical wavelength as $\lambda^{6/5}$. If the line of sight makes an angle θ with the zenith, then the path length L and the integral in Eq. (30) are increased by $\sec \theta$, or r_0 decreases by a factor of $(\sec \theta)^{3/5}$.

A commonly used model for the $C_n^2(z)$ variation is the Hufnagle-Valley model, often referred to as the $H - V_{5/7}$ model, given by¹³

$$C_n^2(h) = 8.2 \times 10^{-26} W^2 h^{10} e^{-h} + 2.7 \times 10^{-16} e^{-h/1.5} + A e^{-10h} \quad (31)$$

where h is the height from ground in kilometers, $A = 1.7 \times 10^{-14}$, and $W = 21$ m/s is the wind speed. Its variation with h is shown in Fig. 1. It decreases rapidly for the first few kilometers, dips at about 5.9 km, rises slightly due to turbulence in the jet stream peaking at about 9.8 km, and decreases monotonically to negligible values beyond about 20 km.

For a point source on the ground observed by an observer on an aircraft, space shuttle, or a satellite, light propagates upward in a manner similar to the $C_n^2(h)$ profile of increasing height h . The observer looks at the point source down through the atmosphere.¹⁴ Substituting Eq. (31) into Eq. (30) with h replaced by z (keeping in mind that C_n^2 is in units of $\text{m}^{-2/3}$), we obtain the value of r_0 in the plane of the observer as a function of his/her height L , as shown in Fig. 2 by the solid curve. The observation is assumed to be along the zenith at a wavelength of $\lambda = 0.5 \mu\text{m}$. We note that r_0 increases slowly, goes through a peak at about 7.2 km and a valley at about 13.7 km, and then increases linearly with L beyond a height of about 20 km above which there is very little atmosphere.

For a ground observer looking up at a point source on an aircraft, space shuttle, or a satellite, z increases downward with a value of zero at the source and L on the ground. To determine r_0 from Eq. (30) using the $C_n^2(h)$ profile such as given by Eq. (31), we must either replace h by $L - z$, or replace z/L by $1 - z/L$ as may be seen by a change of variable. Thus, Eq. (30) in this case is replaced by

$$\begin{aligned} r_0 &= 0.185 \lambda^{6/5} \left[\int_0^L C_n^2(L - z) (z/L)^{5/3} dz \right]^{-3/5} \\ &= 0.185 \lambda^{6/5} \left[\int_0^L C_n^2(z) \left(1 - \frac{z}{L} \right)^{5/3} dz \right]^{-3/5} \end{aligned} \quad (32)$$

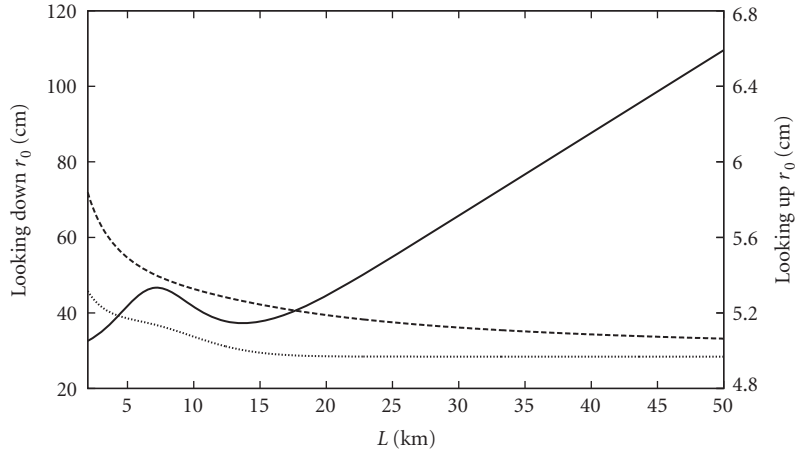


FIGURE 2 Variation of r_0 in the plane of the observer with the path length L from the source to the observer. The propagation of light is along the zenith at a wavelength $\lambda = 0.5 \mu\text{m}$. The solid curve is for a point source on ground such that the spherical wave propagates upward to a space observer looking down directly at the source. The dashed curve is for a point source in space so that a spherical wave propagates down and an observer on ground looks at it directly above. The dotted curve is for plane wave propagation with an observer on ground (or in space) looking up (or down) at a space (or ground) object.

As shown in Fig. 2 by the dashed curve, the value of r_0 in the plane of the observer on ground does not change very much with the height L of the point source.

For a plane wave propagating up or down through the atmosphere, the factor z/L in Eq. (28) and in turn, in Eq. (30), is replaced by unity. It is easy to see this for starlight observed from ground. A plane wave can be thought of as a spherical wave originating at an infinite distance and traveling through a uniform medium for which $C_n^2 = 0$, except for the propagation path through the atmosphere. The value of z/L in the region for which $C_n^2 \neq 0$ is infinitesimally different from unity. Hence, the starlight propagation in ground-based astronomy can be considered as a plane wave propagation with $(z/L)^{5/3}$ in Eq. (30) replaced by unity, or a spherical wave propagating an infinite distance to reach the earth's atmosphere with nonzero C_n^2 value only near the end of its path for which z/L is negligibly different from unity. The variation of r_0 with the separation L of the plane-wave source and the observer is also shown in Fig. 2 by the dotted curve. We note that the variation is small and occurs for small values of L . For large values of L , approaching infinity as in ground-based astronomy, the value of r_0 is close to 5 cm. Moreover, the value of r_0 for a ground observer looking up through the atmosphere at an incoming plane wave is smaller than its value for a corresponding spherical wave, as expected from the smaller factor of $1 - z/L$ compared to z/L in Eq. (32).

We also note that r_0 is smaller when a satellite is observed from ground compared to when a ground object is observed from a satellite. As an observer moves to a higher altitude above the atmosphere, the value of r_0 increases linearly with the altitude L . Consequently, the image degradation is much smaller when a ground object is observed from space than when a space object is observed from ground. In the first case, the object is near the region of turbulence and it is observed from far away. In the second case, the object is away from the region of turbulence, but it is observed from nearby. The fact that the image quality is superior in the first case is similar to when an object behind a diffuse shower glass is observed. One can see some detail in the object when it is in contact with the shower glass. However, as soon as the object is moved slightly away from the shower glass, it appears only as a halo, illustrating complete loss of image resolution. This does not, however, mean that *reciprocity* of wave propagation does not hold. For example, if the wavefront errors of a wave from a point source

in space propagating downward are measured on ground and a conjugate correction is introduced in a beam transmitted upward with a deformable mirror, the beam focus in space will be diffraction limited (neglecting any measurement or correction error), illustrating that the atmosphere introduces the same wavefront errors whether a beam is propagating up or down through it.

Neglecting the variation of C_n^2 for horizontal propagation, we obtain

$$\mathcal{D}_w(r) = \begin{cases} 2.91C_n^2 L k^2 r^{5/3} & \text{Plane wave} \\ (3/8)2.91C_n^2 L k^2 r^{5/3} & \text{Spherical wave} \end{cases} \quad (33a)$$

and

$$r_0 = \begin{cases} 1.68(C_n^2 L k^2)^{-3/5} & \text{Plane wave} \\ 3.02(C_n^2 L k^2)^{-3/5} & \text{Spherical wave} \end{cases} \quad (34a)$$

In the *near field* ($L \ll D^2/\lambda$), the amplitude variations are negligible and, therefore, $\mathcal{D}_\phi(r) = \mathcal{D}_w(r)$. In the *far field* ($L \gg D^2/\lambda$), $\mathcal{D}_\phi(r) = (1/2)\mathcal{D}_w(r)$. For in-between ranges, the multiplying factor between the two structure functions varies smoothly between 1 and 1/2.

Substituting Eq. (29) into Eq. (24), we may write the time-averaged OTF

$$\langle \tau(\vec{v}; D/r_0) \rangle = \tau_L(\vec{v}) M(v; D/r_0) \quad (35)$$

where

$$M(v; D/r_0) = \exp[-3.44(vD/r_0)^{5/3}] \quad (36)$$

is the *long-exposure MCF*. Since $vD = v_i \lambda R$ and r_0 varies as $\lambda^{6/5}$, the factor in the exponent in Eq. (36) varies with wavelength as $\lambda^{-1/3}$.

Since $\exp(-3.44) \approx 0.03$, atmospheric turbulence reduces the overall system MTF corresponding to a spatial frequency $v = r_0/D$ by a factor of 0.03. From Eqs. (15) and (29), we find that r_0 represents a correlation length such that the correlation of complex amplitudes at two points on a wave separated by a distance r_0 is 0.03. Moreover, by definition, the MCF represents the *degree of spatial coherence* of the wave at the receiver, and thus the visibility of the fringes formed in a two-pinhole experiment and observed in the vicinity of a point that is equidistant from the two pinholes. Note that because of the random nature of atmospheric turbulence, the time-averaged irradiances at the two pinholes are equal to each other. Hence, r_0 represents a *partial coherence length* of the wave so that its degree of coherence corresponding to two points on it separated by r_0 is only 0.03, or that the visibility of the fringes formed by the secondary waves from these points is 0.03. The value of r_0 on a mountain site may vary from 5 to 10 cm in the visible region of the spectrum, and increases with wavelength as $\lambda^{6/5}$.

4.5 APPLICATION TO SYSTEMS WITH ANNULAR PUPILS

Now we apply Eq. (23) to a system with an annular pupil of outer radius $a = D/2$ and inner radius $a\epsilon$, and thus an obscuration ratio ϵ and an area $S_{\text{ex}}(\epsilon) = \pi(1 - \epsilon^2)a^2$ (see Fig. 3). For simplicity, we assume that the system is aberration free. Accordingly, its OTF is given by^{11,16}

$$\tau(v; \epsilon) = \frac{1}{1 - \epsilon^2} [\tau(v) + \epsilon^2 \tau(v/\epsilon) - \tau_{12}(v; \epsilon)], \quad 0 \leq v \leq 1 \quad (37)$$

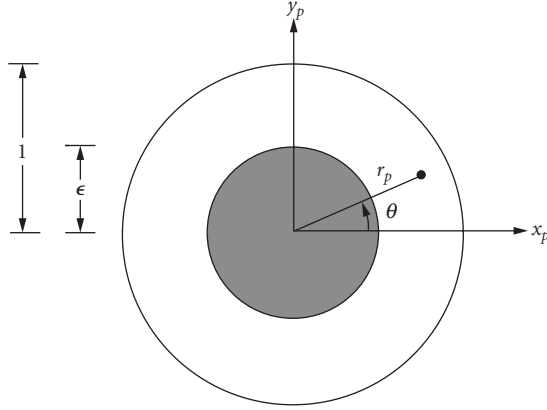


FIGURE 3 Annular pupil of obscuration ratio ϵ , representing the ratio of the inner and outer radii of the pupil.

where $\tau(\nu)$ is given by

$$\begin{aligned} \tau(\nu) &= (2/\pi) [\cos^{-1} \nu - \nu(1 - \nu^2)^{1/2}], \quad 0 \leq \nu \leq 1 \\ &= 0, \text{ otherwise} \end{aligned} \quad (38)$$

and represents the OTF of a corresponding system without any obscuration (i.e., for a circular pupil) and

$$\tau_{12}(\nu; \epsilon) = 2\epsilon^2, \quad 0 \leq \nu \leq (1 - \epsilon)/2 \quad (39a)$$

$$= (2/\pi)(\theta_1 + \epsilon^2\theta_2 - 2\nu\sin\theta_1), \quad (1 - \epsilon)/2 \leq \nu \leq (1 + \epsilon)/2 \quad (39b)$$

$$= 0, \text{ otherwise} \quad (39c)$$

In Eq. (39b), θ_1 and θ_2 are given by

$$\cos\theta_1 = \frac{4\nu^2 + 1 - \epsilon^2}{4\nu} \quad (39d)$$

and

$$\cos\theta_2 = \frac{4\nu^2 - 1 + \epsilon^2}{4\epsilon\nu} \quad (39e)$$

respectively. Since the OTF is equal to the autocorrelation of the pupil function,^{10,11,17} the cutoff frequency, which corresponds to the separation of two pupils by their diameter, is independent of ϵ . From Eq. (37)

$$\frac{\tau(\nu; \epsilon)}{\tau(\nu)} = \frac{1}{1 - \epsilon^2} \quad (40)$$

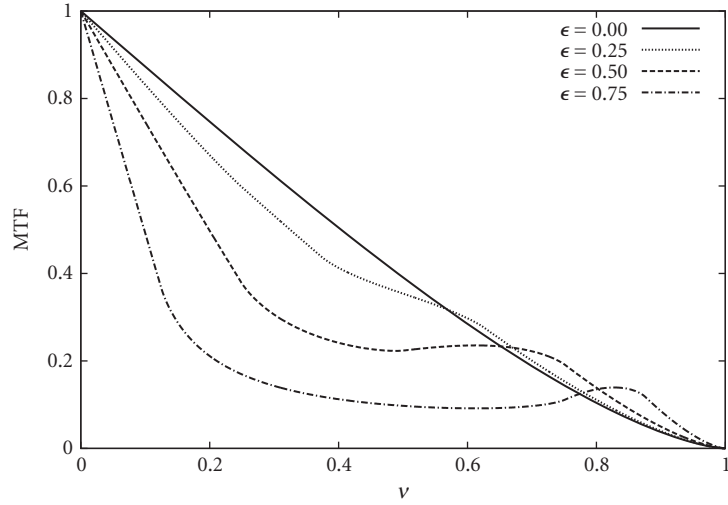


FIGURE 4 Aberration-free OTF $\tau(\nu; \epsilon)$ of an annular pupil illustrating the effect of its obscuration.

for spatial frequencies in the range $(1 + \epsilon)/2 < \nu < 1$. The overlap area of two annular pupils displaced from each other by an amount corresponding to a frequency in this frequency range is independent of ϵ , but the fractional area is larger owing to the smaller area of the obscured exit pupil.¹¹

How $\tau(\nu; \epsilon)$ varies with ν is shown in Fig. 4 for various values of ϵ , including zero. It is evident that the OTF of an annular pupil is significantly lower at low frequencies but somewhat higher at high frequencies, compared to that for a corresponding circular ($\epsilon = 0$) pupil. The slope of the OTF for an annular pupil at the origin is given by

$$\tau'(0; \epsilon) = -4/\pi(1 - \epsilon) \quad (41)$$

This slope does not change when aberrations are introduced.¹¹ Moreover,

$$\int_0^1 \tau(\nu; \epsilon) \nu d\nu = (1 - \epsilon^2)/8 \quad (42)$$

The time-averaged irradiance distribution and encircled power (in a circle of radius r_c in units of λF) of the image of a point object are given by

$$\langle I(r; \epsilon; D/r_0) \rangle = \frac{8}{1 - \epsilon^2} \int_0^1 \langle \tau(\nu; \epsilon; D/r_0) \rangle J_0(2\pi r \nu) \nu d\nu \quad (43)$$

and

$$\langle P(r_c; \epsilon; D/r_0) \rangle = 2\pi r_c \int_0^1 \langle \tau(\nu; \epsilon; D/r_0) \rangle J_1(2\pi r_c \nu) d\nu \quad (44)$$

where

$$\langle \tau(\nu; \epsilon; D/r_0) \rangle = \tau(\nu; \epsilon) \exp \left[-3.44(\nu D/r_0)^{5/3} \right] \quad (45)$$

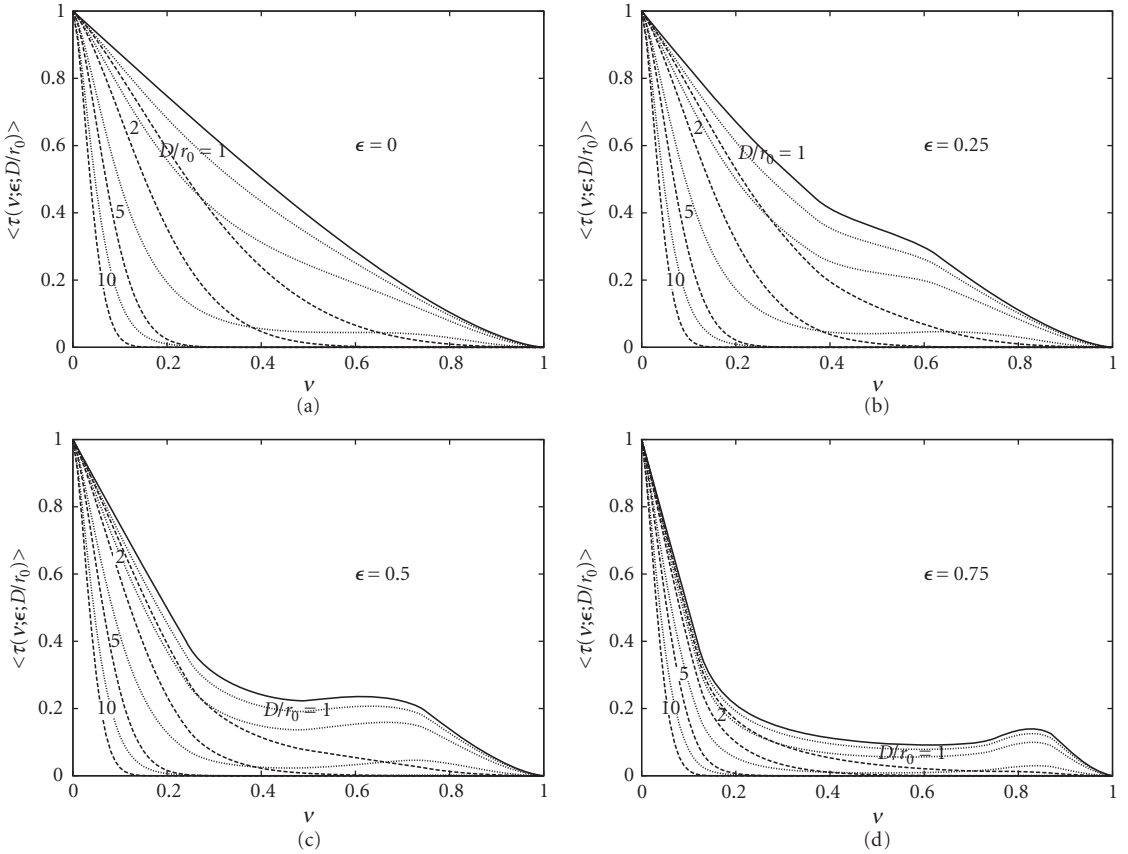


FIGURE 5 Time-averaged OTF for various values of ϵ and D/r_0 . The solid curves represent the aberration-free OTF, and the dashed and dotted curves represent the corresponding long- and short-exposure OTFs.

is the OTF of the system degraded by atmospheric turbulence. The irradiance is normalized by the aberration-free central irradiance $\langle P_{\text{ex}} \rangle S_{\text{ex}}(\epsilon)/\lambda^2 R^2$ and the encircled power is normalized by the total power $\langle P_{\text{ex}} \rangle$. Figure 5 shows the time-averaged OTF for several values of D/r_0 . The OTF gain at high frequencies disappears even for weak turbulence, as is evident from Fig. 5b for $D/r_0 = 1$. The turbulence dominates the OTF for large values of D/r_0 , and the effect of obscuration becomes small, as illustrated in Fig. 5c. Not only is the MTF at any frequency reduced, but the effective cutoff frequency is also reduced, for example, from a value of 1 to about 0.1 when $D/r_0 = 10$.

Letting $r = 0$ in Eq. (43) yields the Strehl ratio of the image

$$\begin{aligned} \langle S(\epsilon; D/r_0) \rangle &\equiv \langle I(0; \epsilon; D/r_0) \rangle \\ &= \frac{8}{1 - \epsilon^2} \int_0^1 \langle \tau(v; \epsilon; D/r_0) \rangle v dv \end{aligned} \quad (46)$$

Figure 6 shows how the time-averaged Strehl ratio varies with D/r_0 for several values of ϵ . The Strehl ratio decreases monotonically as D/r_0 increases. It is slightly lower for an annular pupil for values of $D/r_0 \lesssim 3$ and somewhat higher for $D/r_0 \gtrsim 3$. As shown in Fig. 6b, it decreases monotonically with increasing

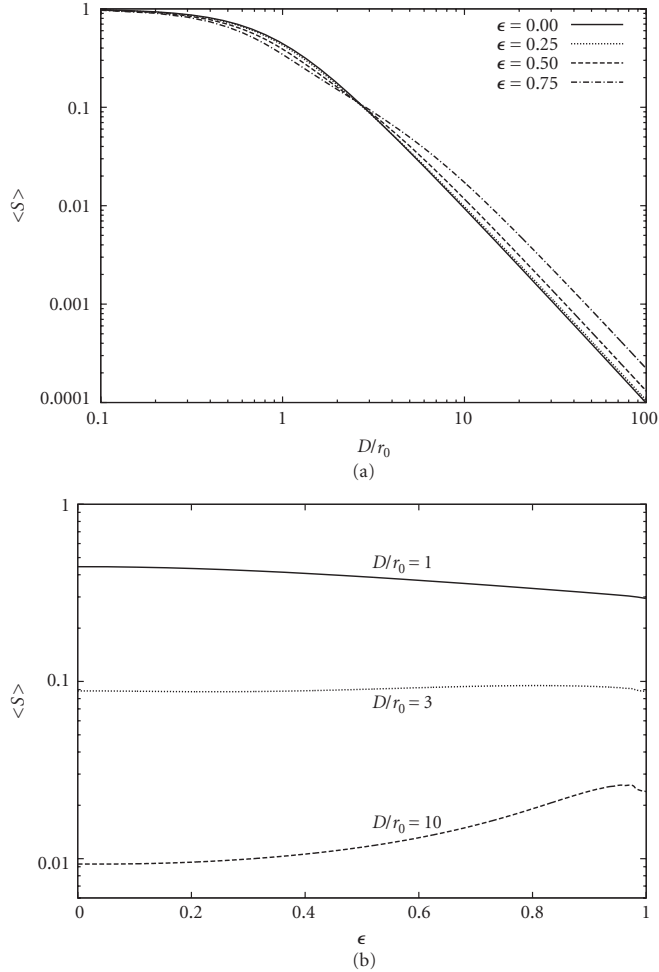


FIGURE 6 Variation of time-averaged long-exposure Strehl ratio $\langle S(\epsilon; D/r_0) \rangle$, representing the central irradiance normalized by its corresponding aberration-free value. (a) As a function of D/r_0 and (b) as a function of ϵ .

value of ϵ for small values of D/r_0 , remains nearly constant for $D/r_0 = 3$, and increases monotonically for $D/r_0 = 10$. Some typical values of the Strehl ratio are listed in Table 1 for several values of ϵ and D/r_0 .

Since $S_{\text{ex}}(\epsilon) = \pi(1 - \epsilon^2)D^2/4$, the aberration-free central irradiance $\langle P_{\text{ex}} \rangle S_{\text{ex}}(\epsilon)/\lambda^2 R^2$ for a fixed total power $\langle P_{\text{ex}} \rangle$ increases with D as D^2 . To see how the aberrated central irradiance varies with D , we consider a quantity¹⁸

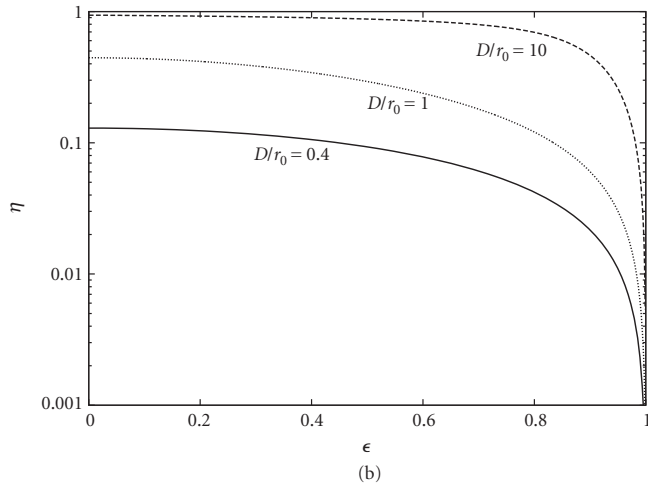
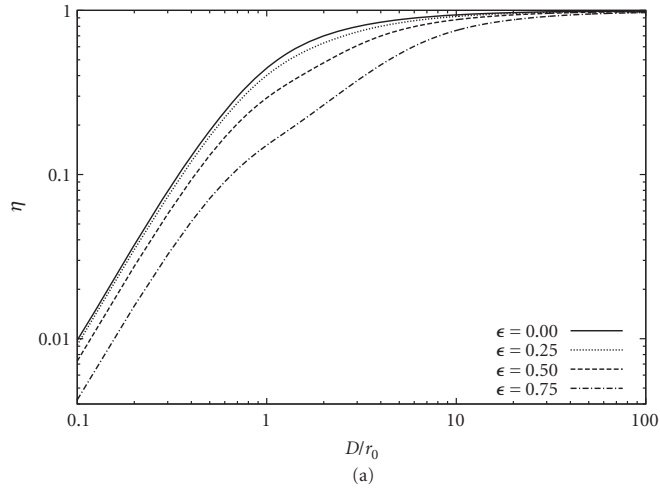
$$\eta(\epsilon; D/r_0) = (1 - \epsilon^2)(D/r_0)^2 \langle S(\epsilon; D/r_0) \rangle \quad (47a)$$

$$= 8(D/r_0)^2 \int_0^1 \langle \tau(v; \epsilon; D/r_0) \rangle \exp[-3.44(vD/r_0)^{5/3}] v dv \quad (47b)$$

Figure 7 shows how η varies with D/r_0 . Its aberration-free (or diffraction-limited) value varies as $(1 - \epsilon^2)(D/r_0)^2$. For very small values of D/r_0 , the atmospheric MTF reduction factor is approximately

TABLE 1 Long (LE) and Short-Exposure (SE) Time-Averaged Strehl Ratio for Various Values of ϵ and D/r_0

ϵ	$D/r_0 = 1$		$D/r_0 = 2$		$D/r_0 = 5$		$D/r_0 = 10$	
	LE	SE	LE	SE	LE	SE	LE	SE
0	0.446	0.889	0.175	0.691	0.035	0.204	0.009	0.023
0.25	0.429	0.890	0.169	0.694	0.036	0.207	0.010	0.024
0.50	0.391	0.892	0.160	0.699	0.040	0.216	0.012	0.027
0.75	0.345	0.889	0.152	0.691	0.050	0.211	0.017	0.033

**FIGURE 7** Variation of long-exposure $\eta(\epsilon; D/r_0)$, representing the time-averaged central irradiance normalized by its aberration-free value for a circular pupil of diameter r_0 but with the same total power as the annular pupil. (a) As a function of D/r_0 and (b) as a function of ϵ .

equal to unity. Accordingly, $\langle S(\epsilon; D/r_0) \rangle$ is also approximately equal to unity, and the aberrated value of η increases with D/r_0 as in the aberration-free case. However, it increases slowly as D/r_0 increases, with a negligible increase beyond a certain value of D/r_0 , depending on the value of ϵ . The contribution to the integral in Eq. (47b) comes from values of v small enough that the factor in the exponent is not vanishingly small. Since, $\tau(v; \epsilon) \approx 1$ near the origin irrespective of the value of ϵ , Eq. (47b) for large values of D/r_0 may be written

$$\begin{aligned} \eta(\epsilon; D/r_0) &= 8(D/r_0)^2 \int_0^1 \exp[-3.44(vD/r_0)^{5/3}] v dv \\ &= 8(3.44)^{-6/5} (3/5) \int_0^\infty x^{(6/5)-1} \exp(-x) dx \\ &= 1 \end{aligned} \quad (48)$$

where $x = 3.44(vD/r_0)^{5/3}$ and the integral over x is the gamma function $\Gamma(6/5)$. Thus,

$$\eta(\epsilon; D/r_0) \rightarrow 1 \text{ as } D/r_0 \rightarrow \infty \quad (49)$$

independent of the value of ϵ . As is evident from Fig. 7, the saturation effects of atmospheric turbulence occur at larger and larger values of D/r_0 as ϵ increases. The two asymptotes of $\eta(\epsilon; D/r_0)$ for a given value of ϵ intersect at the point given by $(1 - \epsilon^2)(D/r_0)^2 = 1$ or $D/r_0 = (1 - \epsilon^2)^{-1/2}$.¹⁸

The unnormalized aberrated central irradiance is given by

$$\langle I_i(0; \epsilon; D/r_0) \rangle = \frac{\langle P_{\text{ex}} \rangle S_{\text{ex}}(\epsilon)}{\lambda^2 R^2} \langle S(\epsilon; D/r_0) \rangle \quad (50a)$$

$$= \frac{\langle P_{\text{ex}} \rangle S_a}{\lambda^2 R^2} \eta(\epsilon; D/r_0) \quad (50b)$$

where $S_a = \pi r_0^2/4$ is the *coherent area* of the atmosphere. Hence, regardless of how large D is, the central irradiance is less than or equal to the aberration-free central irradiance for a system with a circular exit pupil of diameter r_0 , equality approaching as $D/r_0 \rightarrow \infty$. The limiting value of the central irradiance is independent of the value of ϵ . Since $S_a \sim r_0^2 \sim \lambda^{12/5}$, the limiting value varies with wavelength as $\lambda^{2/5}$. It is evident from Eq. (50b) that η represents the aberrated central irradiance normalized by the aberration-free value $\langle P_{\text{ex}} \rangle S_a / \lambda^2 R^2$ for a circular pupil of diameter r_0 .

In astronomical observations, the time-averaged image power given by $\langle P_{\text{ex}} \rangle = \pi(1 - \epsilon^2)D^2 \langle I_0 \rangle$, where $\langle I_0 \rangle$ is the time-averaged irradiance across the exit pupil, increases as D increases. However, if the observation is made against a uniform background, then the background irradiance in the image also increases as D^2 . Hence, regardless of the value of ϵ , the detectability of a point object is limited by turbulence to a value corresponding to an exit pupil of diameter r_0 , no matter how large the diameter D of the exit pupil is. In the case of a laser transmitter with a fixed value of laser power P_{ex} , the central irradiance on a target will again be limited to its aberration-free value for an exit pupil of diameter r_0 , regardless of how large its beam diameter is. Similarly, the ratio of the signal and noise powers in an optical heterodyne detection of a turbulence-degraded signal is limited to the aberration-free value corresponding to an exit pupil of diameter r_0 .

Figure 8 shows how the time-averaged irradiance distribution or the PSF and encircled power change as D/r_0 increases for several values of ϵ . The PSFs are normalized to unity at the center, the actual central value being the LE Strehl ratio given in Table 1. As ϵ increases, power flows from the central bright spot into the diffraction rings. As D/r_0 increases, the diffraction rings disappear and the PSFs become smooth, and a given fraction of total power is contained in a circle of larger and larger radius.

4.6 MODAL EXPANSION OF ABERRATION FUNCTION

So far we have considered the degradation of system performance by propagation through atmospheric turbulence without any knowledge of the magnitude or the type of the phase aberrations introduced by it. In this section, we expand the Kolmogorov turbulence-degraded aberration function into a complete set of orthonormal polynomials and determine its time-averaged variance as well as the variance and covariance of the expansion coefficients.

We expand the aberration function $\Phi(\rho, \theta; \epsilon)$ in terms of a complete set of Zernike annular polynomials $Z_j(\rho, \theta; \epsilon)$ that are orthonormal over the unit annulus in the form⁵⁻⁸

$$\Phi(\rho, \theta; \epsilon) = \sum_j a_j(\epsilon) Z_j(\rho, \theta; \epsilon) \quad (51)$$

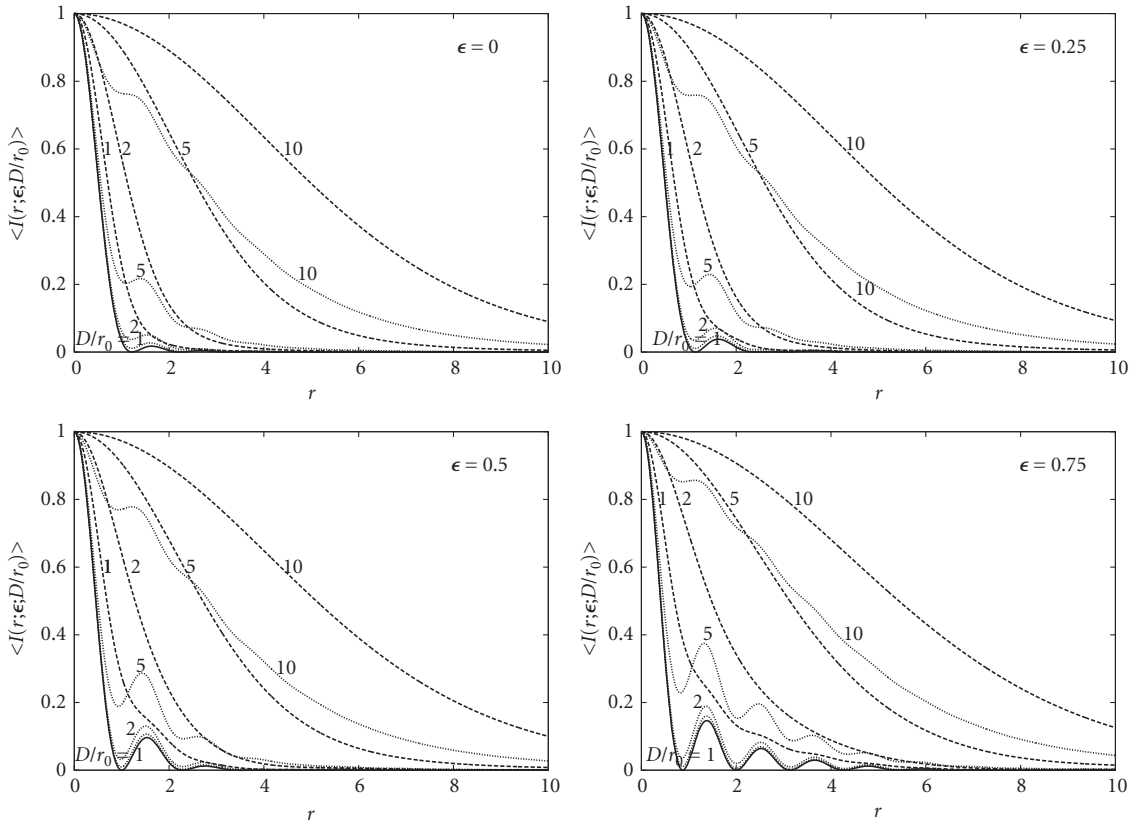


FIGURE 8 Time-averaged irradiance $\langle I(r; \epsilon; D/r_0) \rangle$ and encircled power $\langle P(r; \epsilon; D/r_0) \rangle$ distributions for circular and annular pupils. The solid curves represent the aberration-free distributions, and the dashed and dotted curves represent the corresponding long- and short-exposure distributions.

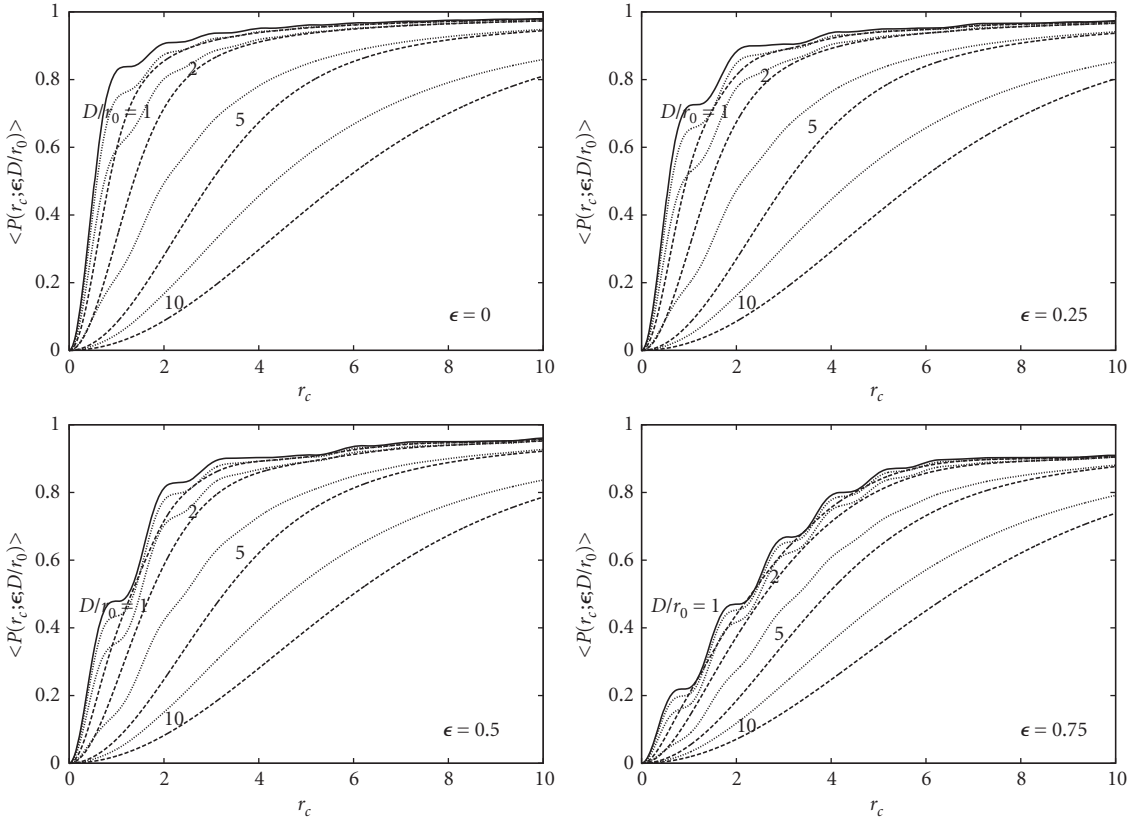


FIGURE 8 (Continued)

where $a_j(\epsilon)$ are the expansion coefficients of the polynomials, $\epsilon \leq \rho \leq 1$ and $0 \leq \theta < 2\pi$. Since the phase aberration is a Gaussian random variable with a zero time-averaged value, so are the expansion coefficients. The annular polynomials are given by

$$Z_{\text{even } j}(\rho, \theta; \epsilon) = \sqrt{2(n+1)} R_n^m(\rho; \epsilon) \cos m\theta, \quad m \neq 0 \quad (52a)$$

$$Z_{\text{odd } j}(\rho, \theta; \epsilon) = \sqrt{2(n+1)} R_n^m(\rho; \epsilon) \sin m\theta, \quad m \neq 0 \quad (52b)$$

$$Z_j(\rho, \theta; \epsilon) = \sqrt{n+1} R_n^0(\rho; \epsilon), \quad m = 0 \quad (52c)$$

where n and m are positive integers (including zero) and $n - m \geq 0$ and even.

The index n represents the radial degree or the order of a polynomial, since it represents the highest power of ρ in the polynomial, and m is called the azimuthal frequency. The index j is a polynomial-ordering number and is a function of both n and m . The polynomials are ordered such that an even j corresponds to a symmetric polynomial varying as $\cos m\theta$, while an odd j corresponds to an antisymmetric polynomial varying as $\sin m\theta$. A polynomial with a lower value of n is ordered first, while for a given value of n , a polynomial with a lower value of m is ordered first.

TABLE 2 Relationship Among the Polynomial Indices n , m , and j

n	m	j	n	m	j
0	0	1	6	0	22
1	1	2, 3	6	2	23, 24
2	0	4	6	4	25, 26
2	2	5, 6	6	6	27, 28
3	1	7, 8	7	1	29, 30
3	3	9, 10	7	3	31, 32
4	0	11	7	5	33, 34
4	2	12, 13	7	7	35, 36
4	4	14, 15	8	0	37
5	1	16, 17	8	2	38, 39
5	3	18, 19	8	4	40, 41
5	5	20, 21	8	6	42, 43
			8	8	44, 45

The relationships among the indices j , n , and m are given in Table 2. For example, when $n = 4$ and $m = 2$, then $j = 12$ for the $\cos 2\theta$ polynomial and $j = 13$ for the $\sin 2\theta$ polynomial. The number of polynomials for a given value of n is $n + 1$, and the number of polynomials up to and including a certain order n is given by

$$N_n = (n + 1)(n + 2)/2 \quad (53)$$

The radial annular polynomials $R_n^m(\rho; \epsilon)$ obey the orthogonality relation

$$\int_{\epsilon}^1 R_n^m(\rho; \epsilon) R_{n'}^m(\rho; \epsilon) \rho \, d\rho = \frac{1 - \epsilon^2}{2(n + 1)} \delta_{nn'} \quad (54)$$

where $\delta_{nn'}$ is a Kronecker delta. Accordingly, the annular polynomials obey the orthonormality condition

$$\int_{\epsilon=0}^1 \int_{\theta=0}^{2\pi} Z_j(\rho, \theta; \epsilon) Z_{j'}(\rho, \theta; \epsilon) \rho \, d\rho \, d\theta \bigg/ \int_{\epsilon=0}^1 \int_{\theta=0}^{2\pi} \rho \, d\rho \, d\theta = \delta_{jj'} \quad (55)$$

The Zernike expansion coefficients are given by

$$a_j(\epsilon) = \frac{1}{\pi(1 - \epsilon^2)} \int_{\epsilon=0}^1 \int_{\theta=0}^{2\pi} \Phi(\rho, \theta; \epsilon) Z_j(\rho, \theta; \epsilon) \rho \, d\rho \, d\theta \quad (56)$$

as may be seen by substituting Eq. (51) into Eq. (56) and using the orthonormality Eq. (55).

The first eleven annular polynomials are listed in Table 3, where

$$R_1^1(\rho; \epsilon) = \frac{\rho}{(1 + \epsilon^2)^{1/2}} \quad (57a)$$

$$R_2^0(\rho; \epsilon) = \frac{(2\rho^2 - 1 - \epsilon^2)}{(1 - \epsilon^2)} \quad (57b)$$

$$R_2^2(\rho; \epsilon) = \frac{\rho^2}{(1 + \epsilon^2 + \epsilon^4)^{1/2}} \quad (57c)$$

TABLE 3 Orthonormal Zernike Annular Polynomials $Z_j(\rho, \theta; \epsilon)$

j	n	m	$Z_j(\rho, \theta; \epsilon)$	Aberration Name*
1	0	0	$R_0^0(\rho; \epsilon) = 1$	Piston
2	1	1	$2R_1^1(\rho; \epsilon) \cos \theta$	x tilt
3	1	1	$2R_1^1(\rho; \epsilon) \sin \theta$	y tilt
4	2	0	$\sqrt{3}R_2^0(\rho; \epsilon)$	Defocus
5	2	2	$\sqrt{6}R_2^2(\rho; \epsilon) \sin 2\theta$	Primary astigmatism at 45°
6	2	2	$\sqrt{6}R_2^2(\rho; \epsilon) \cos 2\theta$	Primary astigmatism at 0°
7	3	1	$\sqrt{8}R_3^1(\rho; \epsilon) \sin \theta$	Primary y coma
8	3	1	$\sqrt{8}R_3^1(\rho; \epsilon) \cos \theta$	Primary x coma
9	3	3	$\sqrt{8}R_3^3(\rho; \epsilon) \sin 3\theta$	
10	3	3	$\sqrt{8}R_3^3(\rho; \epsilon) \cos 3\theta$	
11	4	0	$\sqrt{5}R_4^0(\rho; \epsilon)$	Primary spherical

*The name must precede with "orthonormal annular," e.g., orthonormal annular primary x coma.

$$R_3^1(\rho; \epsilon) = \frac{3(1 + \epsilon^2)\rho^3 - 2(1 + \epsilon^2 + \epsilon^4)\rho}{(1 - \epsilon^2)[(1 + \epsilon^2)(1 + 4\epsilon^2 + \epsilon^4)]^{1/2}} \quad (57d)$$

$$R_3^3(\rho; \epsilon) = \frac{\rho^3}{(1 + \epsilon^2 + \epsilon^4 + \epsilon^6)^{1/2}} \quad (57e)$$

and

$$R_4^0(\rho; \epsilon) = \frac{6\rho^4 - 6(1 + \epsilon^2)\rho^2 + 1 + 4\epsilon^2 + \epsilon^4}{(1 - \epsilon^2)^2} \quad (57f)$$

Of course, they reduce to the corresponding circle polynomials as $\epsilon \rightarrow 0$.

4.7 COVARIANCE AND VARIANCE OF EXPANSION COEFFICIENTS

From Eq. (56), the covariance or cross-correlation of the expansion coefficients may be written

$$\langle a_j(\epsilon) a_{j'}(\epsilon) \rangle = \frac{1}{\pi^2(1 - \epsilon^2)^2} \iint Z_j(\rho, \epsilon) Z_{j'}(\rho'; \epsilon) \langle \Phi(\rho; \epsilon) \Phi(\rho'; \epsilon) \rangle d\rho d\rho' \quad (58)$$

where ρ is the normalized position vector of a pupil point. It can be obtained by using the Fourier transforms of the annular polynomials.⁹ The autocorrelation, that is, the mean square value or the variance of an expansion coefficient, may be obtained from Eq. (58) by letting $j = j'$, that is, by letting $n = n'$ and $m = m'$.

A closed-form analytical solution is obtained when $m = n = n'$

$$\begin{aligned} \left\langle \left[a_n^n(\epsilon) \right]^2 \right\rangle &= \frac{0.0229(n+1)\Gamma(n-5/6)\pi^{8/3}}{2^{5/3}\Gamma(17/6)(1-\epsilon^2)[1-\epsilon^{2(n+1)}]} \left(\frac{D}{r_0} \right)^{5/3} \\ &\times \left[\frac{(1 + \epsilon^{2n+17/3})\Gamma(14/3)}{\Gamma(17/6)\Gamma(n+23/6)} - \frac{2\epsilon^{2(n+1)}}{(n+1)!} F_1 \left(n - \frac{5}{6}, -\frac{11}{6}; n+2; \epsilon^2 \right) \right] \end{aligned} \quad (59)$$

where $F_1(a, b; c; z)$ is a hypergeometric function. Since $r_0 \propto \lambda^{6/5}$, the variance in terms of the optical path-length errors [obtained by multiplying $\langle [a_n''(\epsilon)]^2 \rangle$ by $(\lambda/2\pi)^2$] is independent of λ , as expected in the absence of atmospheric dispersion.

For a circular pupil,¹⁹ the following results can be obtained from the corresponding results for an annular pupil by letting $\epsilon \rightarrow 0$:

$$\begin{aligned} \langle a_j a_{j'} \rangle &= 0.1534(-1)^{(n+n'-2m)/2} [(n+1)(n'+1)]^{1/2} (D/r_0)^{5/3} \delta_{mm'} \\ &\times \frac{\Gamma(14/3)\Gamma[(n+n'-5/3)/2]}{\Gamma[(n-n'+17/3)/2]\Gamma[(n'-n+17/3)/2]\Gamma[(n+n'+23/3)/2]}, \quad n, n' \neq 0 \end{aligned} \quad (60)$$

and

$$\langle a_j^2 \rangle = 0.7587 \frac{\Gamma(n-5/6)}{\Gamma(n+23/6)} (n+1) \left(\frac{D}{r_0} \right)^{5/3}, \quad n \neq 0 \quad (61)$$

Numerical values of variance and covariance are given in Tables 4 and 5 for various values of ϵ including zero. We note that when $\epsilon = 0$, the variance of an aberration coefficient depends on

TABLE 4 Variance $\langle a_j^2(\epsilon) \rangle$ of Expansion Coefficients in Units of $(D/r_0)^{5/3}$

$\langle a_j^2 \rangle$	$\epsilon=0$	0.05	0.1	0.15	0.2	0.25	0.35	0.5	0.75
$\langle a_2^2 \rangle, \langle a_5^2 \rangle$	0.44888	0.45000	0.45336	0.45893	0.46670	0.47660	0.50260	0.55610	0.67917
$\langle a_4^2 \rangle, \langle a_6^2 \rangle$	0.02322	0.02302	0.02245	0.02153	0.02033	0.01891	0.01560	0.01028	0.00292
$\langle a_5^2 \rangle, \langle a_6^2 \rangle$	0.02322	0.02328	0.02345	0.02375	0.02418	0.02475	0.02635	0.02997	0.03888
$\langle a_7^2 \rangle, \langle a_8^2 \rangle$	0.00619	0.00621	0.00624	0.00628	0.00630	0.00626	0.00593	0.00468	0.00162
$\langle a_9^2 \rangle, \langle a_{10}^2 \rangle$	0.00619	0.00621	0.00625	0.00633	0.00645	0.00660	0.00705	0.00815	0.01132
$\langle a_{11}^2 \rangle, \langle a_{12}^2 \rangle$	0.00245	0.00241	0.00227	0.00207	0.00183	0.00158	0.00109	0.00052	0.00007
$\langle a_{12}^2 \rangle, \langle a_{13}^2 \rangle$	0.00245	0.00246	0.00248	0.00251	0.00255	0.00260	0.00266	0.00247	0.00108
$\langle a_{14}^2 \rangle, \langle a_{15}^2 \rangle$	0.00245	0.00246	0.00248	0.00251	0.00256	0.00262	0.00280	0.00326	0.00480

TABLE 5 Covariance $\langle a_j(\epsilon) a_{j'}(\epsilon) \rangle$ of Expansion Coefficients in Units of $(D/r_0)^{5/3}$

$\langle a_j a_{j'} \rangle$	$\epsilon = 0$	0.05	0.1	0.15	0.2	0.25	0.35	0.5	0.75
$\langle a_2 a_8 \rangle, \langle a_3 a_7 \rangle$	-0.01416	-0.01420	-0.01430	-0.01444	-0.01460	-0.01474	-0.01481	-0.01392	-0.00885
$\langle a_4 a_{11} \rangle$	-0.00388	-0.00381	-0.00363	-0.00334	-0.00300	-0.00262	-0.00186	-0.00091	-0.00012
$\langle a_4 a_{22} \rangle$	0.00032	0.00030	0.00024	0.00017	0.00009	0.00002	-0.00008	-0.00011	-0.00003
$\langle a_5 a_{13} \rangle, \langle a_6 a_{12} \rangle$	-0.00388	-0.00389	-0.00392	-0.00397	-0.00404	-0.00412	-0.00433	-0.00451	-0.00341
$\langle a_7 a_{17} \rangle, \langle a_8 a_{16} \rangle$	-0.00156	-0.00158	-0.00164	-0.00172	-0.00178	-0.00173	-0.00095	0.00365	0.01180
$\langle a_{11} a_{22} \rangle$	-0.00076	-0.00073	-0.00067	-0.00058	-0.00048	-0.00038	-0.00022	-0.00007	0.00000
$\langle a_9 a_{19} \rangle, \langle a_{10} a_{18} \rangle$	-0.00156	-0.00156	-0.00157	-0.00159	-0.00162	-0.00166	-0.00176	-0.00196	-0.00177
$\langle a_{14} a_{26} \rangle, \langle a_{15} a_{25} \rangle$	-0.00076	-0.00076	-0.00077	-0.00078	-0.00079	-0.00081	-0.00086	-0.00099	-0.00104

the value of n , but is independent of the value of m . Thus, for example, $\langle a_4^2 \rangle = \langle a_5^2 \rangle = \langle a_6^2 \rangle$, or $\langle a_7^2 \rangle = \langle a_8^2 \rangle = \langle a_9^2 \rangle = \langle a_{10}^2 \rangle$. However, when $\epsilon \neq 0$, the variance depends on the values of both n and m , but it is independent of whether it is a cosine or a sine mode. Thus, for example, $\langle a_4^2(\epsilon) \rangle \neq \langle a_5^2(\epsilon) \rangle = \langle a_6^2(\epsilon) \rangle$, or $\langle a_7^2(\epsilon) \rangle = \langle a_8^2(\epsilon) \rangle \neq \langle a_9^2(\epsilon) \rangle = \langle a_{10}^2(\epsilon) \rangle$. Similarly, the covariance for given values of n and n' are independent of their m values when $\epsilon = 0$. For example, $\langle a_4 a_{11} \rangle = \langle a_5 a_{13} \rangle = \langle a_6 a_{12} \rangle$, or $\langle a_7 a_{17} \rangle = \langle a_8 a_{16} \rangle = \langle a_9 a_{19} \rangle = \langle a_{10} a_{18} \rangle$. However, when $\epsilon \neq 0$, equality is obtained only when the m values are also equal. For example, $\langle a_4(\epsilon) a_{11}(\epsilon) \rangle \neq \langle a_5(\epsilon) a_{13}(\epsilon) \rangle = \langle a_6(\epsilon) a_{12}(\epsilon) \rangle$, or $\langle a_7(\epsilon) a_{17}(\epsilon) \rangle = \langle a_8(\epsilon) a_{16}(\epsilon) \rangle \neq \langle a_9(\epsilon) a_{19}(\epsilon) \rangle = \langle a_{10}(\epsilon) a_{18}(\epsilon) \rangle$. The first nonzero cross-correlations are the tilt-coma cross-correlations $\langle a_2(\epsilon) a_8(\epsilon) \rangle$ and $\langle a_3(\epsilon) a_7(\epsilon) \rangle$. Moreover, for a given value of n , the correlation values decrease rapidly as the order difference $n' - n$ increases.

Figures 9 and 10 show the variance and covariance of Zernike coefficients as a function of ϵ for some low-order terms. The variance of terms with $m = n$ increase monotonically with an increasing value of ϵ , e.g., tip and tilt variances $\langle a_2^2 \rangle$ and $\langle a_3^2 \rangle$, or astigmatism variances $\langle a_5^2 \rangle$ and $\langle a_6^2 \rangle$. For other terms, e.g., coma $\langle a_7^2 \rangle$ and $\langle a_8^2 \rangle$, or defocus $\langle a_4^2 \rangle$ and spherical aberration $\langle a_{11}^2 \rangle$, they generally decrease monotonically and approach zero as $\epsilon \rightarrow 1$. Similarly, the covariances of tilt with coma, for example, $\langle a_2(\epsilon) a_8(\epsilon) \rangle$, and defocus with primary or secondary spherical aberration, that is, $\langle a_4(\epsilon) a_{11}(\epsilon) \rangle$ or $\langle a_4(\epsilon) a_{22}(\epsilon) \rangle$, approach zero as $\epsilon \rightarrow 1$.

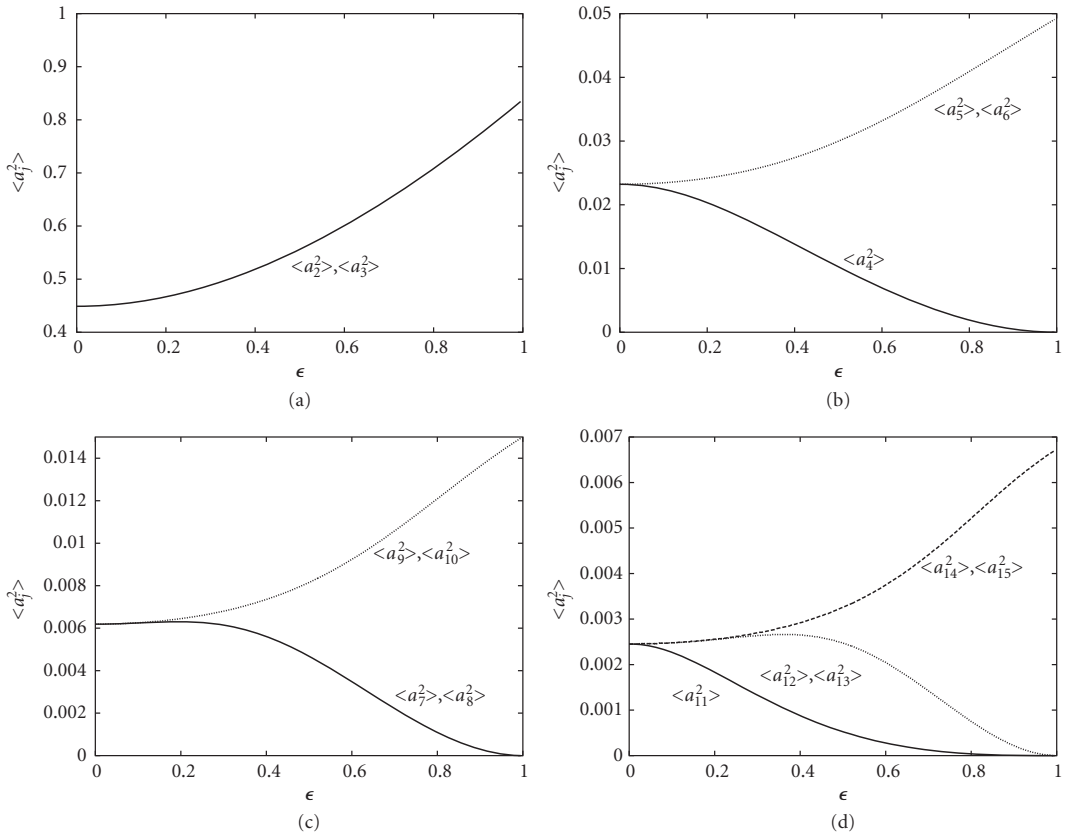


FIGURE 9 Variance $\langle a_j^2(\epsilon) \rangle$ of Zernike coefficients. (a) Tip and tilt; (b) defocus and astigmatism; (c) coma and tilt, etc.; and (d) spherical and higher orders.

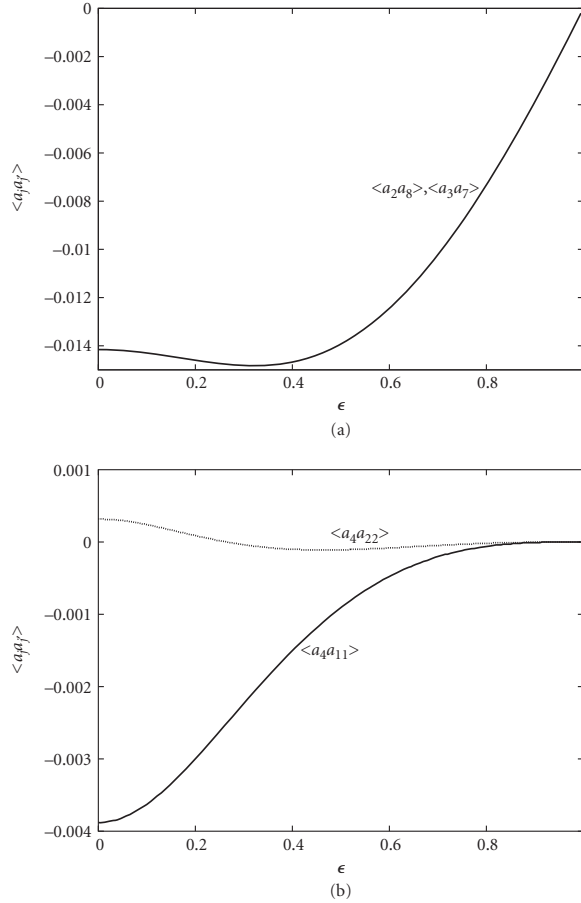


FIGURE 10 Covariance $\langle a_i(\epsilon)a_j(\epsilon) \rangle$ of Zernike coefficients. (a) Tilt and coma and (b) defocus and spherical.

4.8 ANGLE OF ARRIVAL FLUCTUATIONS

Depending on the measurement process, the angle of arrival can be defined by the Zernike tilts of the wavefront or the centroid of the aberrated PSF. The first is referred to as the Z-tilt and represents the least square fit of the linear terms to the aberration function or equivalently the gradient of the wavefront at the center of the pupil. The PSF centroid, however, represents the mean value of the wavefront gradient across the pupil. The tilt associated with the centroid is referred to as the G- or the gradient-tilt.^{1,20}

From Eq. (51), the wavefront tilt aberration introduced by turbulence is given by

$$\begin{aligned}
 \Phi_t(\rho, \theta; \epsilon) &= a_2(\epsilon)Z_2(\rho, \theta; \epsilon) + a_3(\epsilon)Z_3(\rho, \theta; \epsilon) \\
 &= 2[a_2(\epsilon)R_1^1(\rho; \epsilon)\cos\theta + a_3(\epsilon)R_1^1(\rho; \epsilon)\sin\theta] \\
 &= \frac{2}{(1 + \epsilon^2)^{1/2}}[a_2(\epsilon)\rho\cos\theta + a_3(\epsilon)\rho\sin\theta]
 \end{aligned} \tag{62}$$

Or, in Cartesian coordinates,

$$\Phi_t(x, y; \epsilon) = \frac{2}{(1 + \epsilon^2)^{1/2}} [a_2(\epsilon)x + a_3(\epsilon)y] \quad (63)$$

The wavefront tilt displaces the image to a point $[x_t(\epsilon), y_t(\epsilon)]$ given by²¹

$$\begin{aligned} [x_t(\epsilon), y_t(\epsilon)] &= 2F \frac{\lambda}{2\pi} \left(\frac{\partial \Phi_t}{\partial x}, \frac{\partial \Phi_t}{\partial y} \right) \\ &= \frac{2\lambda F}{\pi(1 + \epsilon^2)^{1/2}} [a_2(\epsilon), a_3(\epsilon)] \end{aligned} \quad (64)$$

Accordingly, the angle of arrival of the wave is given by

$$\begin{aligned} [\alpha(\epsilon), \beta(\epsilon)] &= \frac{1}{R} [x_t(\epsilon), y_t(\epsilon)] \\ &= \frac{2\lambda}{\pi D(1 + \epsilon^2)^{1/2}} [a_2(\epsilon), a_3(\epsilon)] \end{aligned} \quad (65)$$

Or, the mean square fluctuation of the angle of arrival is given by

$$[\sigma_\alpha^2(\epsilon), \sigma_\beta^2(\epsilon)] = \left(\frac{2\lambda}{\pi D} \right)^2 \frac{1}{(1 + \epsilon^2)} \langle a_2^2(\epsilon), a_3^2(\epsilon) \rangle \quad (66)$$

where from Eq. (59)

$$\begin{aligned} \langle a_2^2(\epsilon) \rangle &= \langle a_3^2(\epsilon) \rangle \\ &= \frac{0.9858}{(1 + \epsilon^2)(1 - \epsilon^2)^2} \left[0.4554(1 + \epsilon^{23/3}) - \epsilon^4 F_1 \left(\frac{1}{6}, -\frac{11}{6}; 3; \epsilon^2 \right) \right] \left(\frac{D}{r_0} \right)^{5/3} \end{aligned} \quad (67)$$

Accordingly, the standard deviation of the angular fluctuations based on the Zernike tilts is given by

$$[\sigma_\alpha(\epsilon)]_Z = [\sigma_\beta(\epsilon)]_Z = \frac{2}{\pi(1 + \epsilon^2)^{1/2}} \frac{\lambda}{D} \left[\langle a_2^2(\epsilon) \rangle \right]^{1/2} \quad (68)$$

Since $\langle a_2^2(\epsilon) \rangle \sim (D/r_0)^{5/3}$, Eq. (68) shows that the mean square fluctuation of the image displacement decreases with the pupil diameter D as $D^{-1/3}$. Moreover, since $r_0 \sim \lambda^{6/5}$, we find that the fluctuation is independent of the wavelength, as expected in the absence of any atmospheric dispersion. For a circular pupil, substituting $\langle a_2^2 \rangle = \langle a_3^2 \rangle = 0.449(D/r_0)^{5/3}$, we obtain

$$\begin{aligned} (\sigma_\alpha)_Z &= (\sigma_\beta)_Z = \frac{2}{\pi} \frac{\lambda}{D} \left[\langle a_2^2 \rangle \right]^{1/2} \\ &= 0.4265 \frac{\lambda}{D} \left(\frac{D}{r_0} \right)^{5/6} \end{aligned} \quad (69)$$

Since the wavefront tilt produced by turbulence is independent of the shape of the pupil, except for its correlation with the coma terms, $\langle a_2^2(\epsilon) \rangle$ increases with ϵ approximately as $1 + \epsilon^2$. This is illustrated in Fig. 11, where $\langle a_2^2(\epsilon) \rangle$ and $(1 + \epsilon^2)\langle a_2^2 \rangle$ are compared with each other as a function of ϵ . Accordingly, the angular fluctuation depends very weakly on the obscuration ratio. For example, when $\epsilon = 0.5$, the constant 0.4265 on the right-hand side of Eq. (69) is replaced by 0.4246.

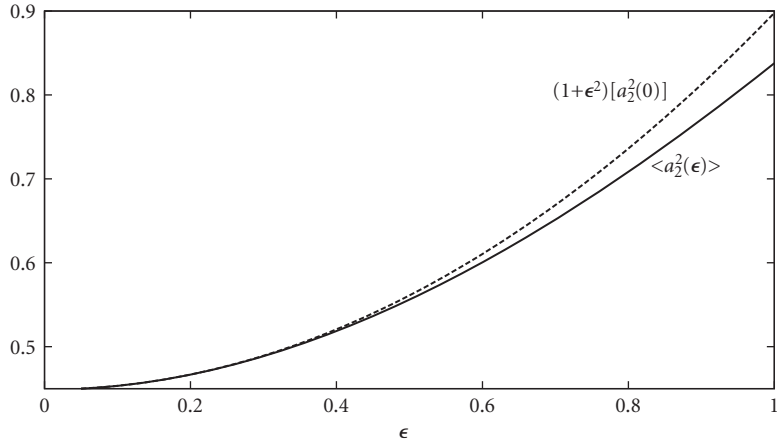


FIGURE 11 Comparison of tilt variance $\langle a_2^2(\epsilon) \rangle$ with $(1 + \epsilon^2)\langle a_2^2 \rangle$.

Tatarski gives an expression for the fluctuations of the angle of arrival based on the centroid of the aberrated PSF for a circular pupil.²² His analysis can be extended to systems with annular pupils. However, we consider the centroid in terms of the Zernike annular-polynomial expansion coefficients of the aberration function. This approach gives insight into why the G-tilt is different from the Z-tilt, why the two tilts are approximately equal to each other, and why the G-tilt is smaller.

Neglecting the amplitude variations introduced by turbulence, the centroid (x_c, y_c) of an aberrated PSF is given by the mean value of the gradient of the aberration function^{11,23}

$$(x_c, y_c) = \frac{R}{S_p} \iint \left[\frac{\partial W(x, y)}{\partial x}, \frac{\partial W(x, y)}{\partial y} \right] dx dy \quad (70)$$

where $W(x, y)$ is the wave aberration at a point (x, y) on a pupil of area S_p . If we write the aberration function in terms of the Zernike annular polynomials in the form

$$W(\rho, \theta; \epsilon) = \sum_{n=0}^{\infty} \sum_{m=0}^n [2(n+1)/(1 + \delta_{m0})]^{1/2} R_n^m(\rho; \epsilon) [c_{nm}(\epsilon) \cos m\theta + s_{nm}(\epsilon) \sin m\theta] \quad (71)$$

where $c_{nm}(\epsilon)$ and $s_{nm}(\epsilon)$ are the Zernike annular expansion coefficients, it can be shown that the centroid of the corresponding PSF is given by^{11,23}

$$[x_c(\epsilon), y_c(\epsilon)] = \frac{2F}{1 - \epsilon^2} \sum_{n=1}^{\infty} \sqrt{2(n+1)} [R_n^1(1; \epsilon) - \epsilon R_n^1(\epsilon; \epsilon)] [c_{n1}(\epsilon), s_{n1}(\epsilon)] \quad (72)$$

where a prime on the summation sign indicates a summation over only the odd integral values of n . Thus, the only aberrations that contribute to the centroid are those with $m = 1$. Aberrations of the type $R_n^1(\rho; \epsilon) \cos \theta$ contribute to x_c and those of the type $R_n^1(\rho; \epsilon) \sin \theta$ contribute to y_c , which also follows from the symmetry considerations of the aberrations. Keeping the tilt and only the primary coma terms (and thus neglecting higher orders of coma owing to their small magnitudes) in Eq. (72), we obtain

$$\begin{aligned} [\sigma_\alpha(\epsilon)]_G &= [\sigma_\beta(\epsilon)]_G \\ &= \frac{1}{\pi(1 - \epsilon^2)} \frac{\lambda}{D} \left[4A^2(\epsilon) \langle a_2^2(\epsilon) \rangle + 8B^2(\epsilon) \langle a_8^2(\epsilon) \rangle + 8\sqrt{2} \frac{1 - \epsilon^2}{(1 + \epsilon^2)^{1/2}} B(\epsilon) \langle a_2(\epsilon) a_8(\epsilon) \rangle \right]^{1/2} \end{aligned} \quad (73)$$

where, from Eqs. (57a) and (57d),

$$A(\epsilon) = R_1^1(1; \epsilon) - \epsilon R_1^1(\epsilon; \epsilon) = \frac{1 - \epsilon^2}{(1 + \epsilon^2)^{1/2}} \quad (74a)$$

and

$$B(\epsilon) = R_3^1(1; \epsilon) - \epsilon R_3^1(\epsilon; \epsilon) = \frac{1 + 3\epsilon^2 - 3\epsilon^4 - \epsilon^6}{(1 - \epsilon^2)[(1 + \epsilon^2)(1 + 4\epsilon^2 + \epsilon^4)]^{1/2}} \quad (74b)$$

If we retain only the Zernike tilt terms, we obtain the Z-tilt given by Eq. (68).

For a circular pupil, $A(\epsilon)$ and $B(\epsilon)$ both reduce to unity, and Eq. (73) reduces to

$$(\sigma_\alpha)_G = (\sigma_\beta)_G = \frac{\lambda}{\pi D} \left[4\langle a_2^2 \rangle + 8\langle a_8^2 \rangle + 8\sqrt{2}\langle a_2 a_8 \rangle \right]^{1/2} \quad (75)$$

Substituting for $\langle a_2^2 \rangle$, $\langle a_8^2 \rangle$, and $\langle a_2 a_8 \rangle$ from Tables 4 and 5, we obtain

$$(\sigma_\alpha)_G = 0.4132 \frac{\lambda}{D} \left(\frac{D}{r_0} \right)^{5/6} \quad (76)$$

The dependence of G- and Z-tilts on ϵ is illustrated in Fig. 12. The G-tilt is smaller than the Z-tilt by approximately 3 percent. It is not surprising that the Z- and G-tilts are approximately equal to each other. Since the Zernike tilt aberrations account for 87 percent or more of the aberration variance (see Sec. 4.11 and Fig. 16), the variance of the coma aberration(s) is relatively small. The G-tilt is smaller than the Z-tilt because of the negative correlations of the tilt coefficients with the coma coefficients.

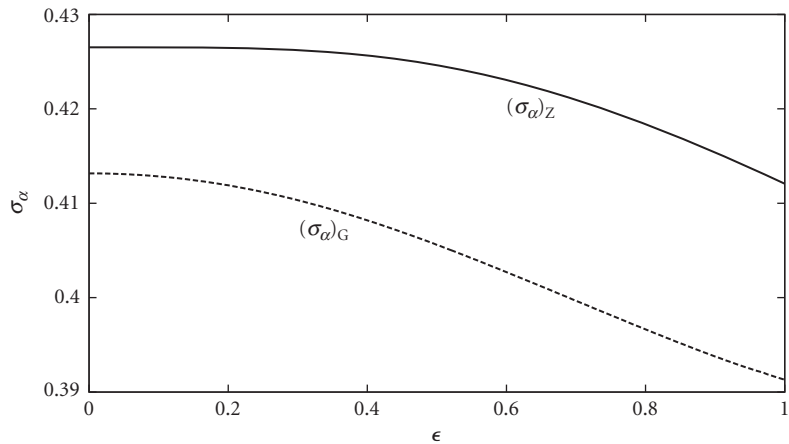


FIGURE 12 Dependence of the angle of arrival fluctuations on the obscuration ratio ϵ , illustrating the small difference between the Zernike wavefront tilt $[\sigma_\alpha(\epsilon)]_Z$ and the centroid-based G-tilt $[\sigma_\alpha(\epsilon)]_G$.

4.9 ABERRATION VARIANCE AND APPROXIMATE STREHL RATIO

The piston autocorrelation $\langle a_1^2(\epsilon) \rangle$ for Kolmogorov turbulence, obtained by letting $j = j' = 1$ in Eq. (58), so that both n and n' are equal to zero, is infinity.⁹ The mean square value $\langle \Phi^2(\epsilon) \rangle$ of the aberration is also infinite, but the difference of the two is finite. Since the mean value $\langle \Phi(\epsilon) \rangle$ of the aberration is zero, $\langle \Phi^2(\epsilon) \rangle$ is also the variance of the aberration. However, the piston aberration $Z_1(\rho, \theta; \epsilon)$, being a constant, does not affect the aberration variance or the image quality. Accordingly, we refer to the piston-removed variance as simply the aberration variance. It is given by

$$\begin{aligned} \sigma_\Phi^2(\epsilon) &\equiv \langle \Phi^2(\epsilon) \rangle - \langle a_1^2(\epsilon) \rangle \\ &= \frac{0.046\pi^{11/3}}{2^{5/3}\Gamma(17/6)\Gamma(11/6)(1-\epsilon^2)^2} \left[\frac{\Gamma(14/3)(1+\epsilon)^{17/3}}{\Gamma(17/6)\Gamma(23/6)} - 2\epsilon^2 F_1\left(-\frac{5}{6}, -\frac{11}{6}, 2; \epsilon^2\right) \right] \left(\frac{D}{r_0}\right)^{5/3} \end{aligned} \quad (77)$$

Letting $\epsilon = 0$ for a circular pupil, Eq. (77) reduces to

$$\begin{aligned} \sigma_\Phi^2 &\equiv \langle \Phi^2 \rangle - \langle a_1^2 \rangle \\ &= 1.0324(D/r_0)^{5/3} \end{aligned} \quad (78)$$

Figure 13 shows how $\Delta_1(\epsilon) \equiv \sigma_\Phi^2(\epsilon)$ in units of $(D/r_0)^{5/3}$ varies with ϵ . Its value increases monotonically from 1.0324 for a circular pupil ($\epsilon = 0$) to 1.843 for an infinitesimally thin ring pupil ($\epsilon \rightarrow 1$).

The approximate expression $S \approx \exp(-\sigma_\Phi^2)$ for the Strehl ratio^{11,24,25} is not suitable for calculating the time-averaged Strehl ratio for a random aberration, especially when its value is small. For example, the Strehl ratio given by

$$\langle S_1(D/r_0) \rangle \approx \exp[-1.03(D/r_0)^{5/3}] \quad (79)$$

is illustrated in Fig. 14 by the dashed curve, which is quite steep owing to the $(D/r_0)^{5/3}$ dependence of σ_Φ^2 . Even for a small value of $D/r_0 = 1$, it gives a Strehl ratio of 0.357, compared to a true value of

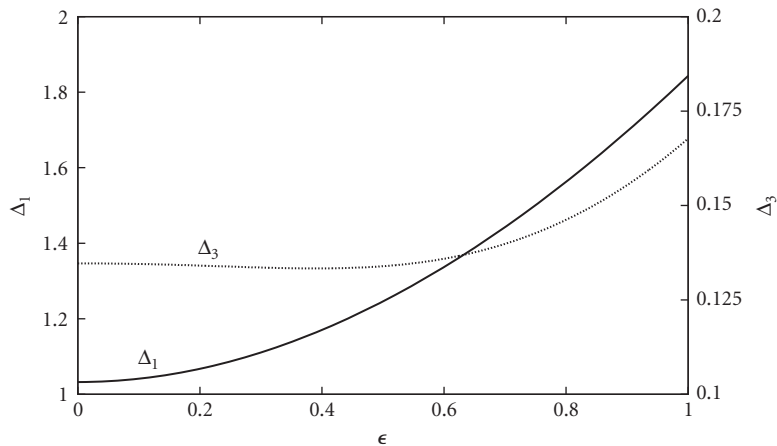


FIGURE 13 Piston-removed aberration variance $\Delta_1(\epsilon) \equiv \sigma_\Phi^2(\epsilon)$ and tilt-corrected variance $\Delta_3(\epsilon)$ in units of $(D/r_0)^{5/3}$ a function of ϵ .

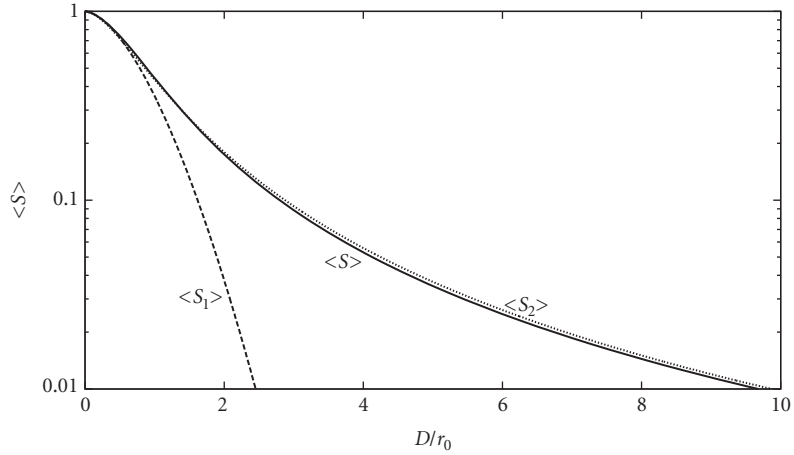


FIGURE 14 Comparison of exact and approximate long-exposure Strehl ratios for a circular pupil.

0.446. For larger values of D/r_0 , it underestimates the Strehl ratio by larger factors. It is not surprising that Eq. (79) yields an underestimate of the true value of the Strehl ratio. If we write Eq. (14) in the form

$$\mathcal{D}_\Phi(r) = \langle [\Phi(0) - \Phi(r)]^2 \rangle = 2[\sigma_\Phi^2 - R_\Phi(r)] \quad (80)$$

where

$$R_\Phi(r) = \langle \Phi(0)\Phi(r) \rangle \quad (81)$$

is the autocorrelation of the phase aberration and substitute in Eq. (24), we find that the Strehl ratio is given by Eq. (79) only when the phase correlation is zero. Otherwise, it will yield a larger value. Unfortunately, the phase correlation function for Kolmogorov turbulence is not defined; only the phase structure function is. A much better approximation is given by

$$\langle S_2(D/r_0) \rangle = [1 + (D/r_0)^{5/3}]^{-6/5} \quad (82)$$

as may be seen from Fig. 14, where it is illustrated by the dotted curve. It overestimates the true value only slightly.

4.10 MODAL CORRECTION OF ATMOSPHERIC TURBULENCE

If the first J modes of the phase aberration are corrected, the corrected phase error may be written

$$\Phi_c(\rho, \theta; \epsilon) = \sum_{j=1}^J a_j(\epsilon) Z_j(\rho, \theta; \epsilon) \quad (83)$$

Accordingly, the residual phase aberration is given by

$$\Phi_J(\rho, \theta; \epsilon) = \Phi(\rho, \theta; \epsilon) - \Phi_c(\rho, \theta; \epsilon) = \sum_{j=J+1}^{\infty} a_j(\epsilon) Z_j(\rho, \theta; \epsilon) \quad (84)$$

Hence, the mean square value of the residual aberration is given by

$$\begin{aligned} \Delta_J(\epsilon) &= \langle [\Phi_J(\rho, \theta; \epsilon)]^2 \rangle \\ &= \pi^{-1} \int_0^1 \int_0^{2\pi} \left\langle \left[\Phi(\rho, \theta; \epsilon) - \sum_{j=1}^J a_j(\epsilon) Z_j(\rho, \theta; \epsilon) \right]^2 \right\rangle \rho d\rho d\theta \\ &= \Delta_1(\epsilon) - \sum_{j=2}^J \langle a_j^2(\epsilon) \rangle \end{aligned} \quad (85)$$

where $\Delta_1(\epsilon) \equiv \sigma_\Phi^2(\epsilon)$. Numerical values of $\Delta_J(\epsilon)$ are given in Table 6 for several values of ϵ . Figure 13 shows how $\Delta_J(\epsilon)$ (the piston-removed and tilt-corrected variance) varies with ϵ . We note that it is nearly constant until ϵ is larger than approximately 0.5. However, as shown in Fig. 15, $\Delta_6(\epsilon)$ piston removed and aberrations corrected up to and including astigmatism) and $\Delta_{11}(\epsilon)$ (the piston removed and aberrations corrected up to and including primary spherical aberration) first increase slightly with ϵ and then decrease considerably, and finally increase rapidly as $\epsilon \rightarrow 1$.

TABLE 6 Variance $\Delta_J(\epsilon)$ of Residual Phase Aberration in Units of $(D/r_0)^{5/3}$

Δ_J	$\epsilon = 0$	0.25	0.5	0.75
Δ_1	1.0324	1.0870	1.2461	1.5010
Δ_2	0.5835	0.6104	0.6900	0.8218
Δ_3	0.1347	0.1338	0.1339	0.1426
Δ_4	0.1115	0.1149	0.1236	0.1397
Δ_5	0.0882	0.0902	0.0937	0.1008
Δ_6	0.0650	0.0654	0.0637	0.0620
Δ_7	0.0588	0.0591	0.0590	0.0603
Δ_8	0.0526	0.0529	0.0543	0.0587
Δ_9	0.0464	0.0463	0.0462	0.0474
Δ_{10}	0.0402	0.0397	0.0380	0.0361
Δ_{11}	0.0378	0.0381	0.0375	0.0360
Δ_{12}	0.0353	0.0355	0.0350	0.0349
Δ_{13}	0.0329	0.0329	0.0325	0.0338
Δ_{14}	0.0304	0.0303	0.0293	0.0290
Δ_{15}	0.0280	0.0277	0.0260	0.0242
Δ_{16}	0.0268	0.0264	0.0244	0.0218
Δ_{17}	0.0256	0.0251	0.0229	0.0193
Δ_{18}	0.0244	0.0239	0.0215	0.0186
Δ_{19}	0.0232	0.0226	0.0201	0.0178
Δ_{20}	0.0220	0.0213	0.0185	0.0153
Δ_{21}	0.0208	0.0200	0.0169	0.0129
Δ_{22}	0.0202	0.0197	0.0168	0.0129
Δ_{23}	0.0195	0.0190	0.0159	0.0115
Δ_{24}	0.0189	0.0183	0.0151	0.0101
Δ_{25}	0.0182	0.0176	0.0142	0.0095
Δ_{26}	0.0176	0.0169	0.0134	0.0089
Δ_{27}	0.0169	0.0162	0.0125	0.0075
Δ_{28}	0.0162	0.0155	0.0117	0.0061

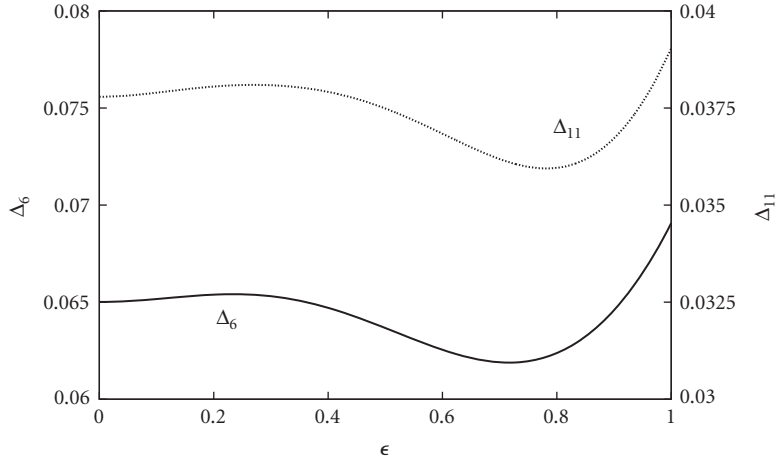


FIGURE 15 $\Delta_J(\epsilon)$ of the residual aberration in units of $(D/r_0)^{5/3}$ after correction of $J = 6$ and $J = 11$ modes.

The residual phase structure function after correction of the first J modes is given by

$$\mathcal{D}_J(\vec{\rho}, \vec{\rho}'; \epsilon) = \langle [\Phi_J(\vec{\rho}; \epsilon) - \Phi_J(\vec{\rho}'; \epsilon)]^2 \rangle \quad (86)$$

$$\begin{aligned} &= \langle \{ [\Phi(\vec{\rho}; \epsilon) - \Phi(\vec{\rho}'; \epsilon)] - [\Phi_c(\vec{\rho}; \epsilon) - \Phi_c(\vec{\rho}'; \epsilon)] \}^2 \rangle \\ &= \mathcal{D}_\Phi(\vec{\rho}, \vec{\rho}'; \epsilon) - \langle [\Phi_c(\vec{\rho}; \epsilon) - \Phi_c(\vec{\rho}'; \epsilon)]^2 \rangle \\ &\quad - 2\langle [\Phi_c(\vec{\rho}; \epsilon) - \Phi_c(\vec{\rho}'; \epsilon)] [\Phi_J(\vec{\rho}; \epsilon) - \Phi_J(\vec{\rho}'; \epsilon)] \rangle \end{aligned} \quad (87)$$

where, as in Eq. (14),

$$\mathcal{D}_\Phi(\vec{\rho}, \vec{\rho}'; \epsilon) = \langle [\Phi(\vec{\rho}; \epsilon) - \Phi(\vec{\rho}'; \epsilon)]^2 \rangle \quad (88)$$

is the long-exposure phase structure function. From Eq. (14), $\mathcal{D}_\Phi(\vec{\rho}, \vec{\rho}'; \epsilon)$ depends on $|\vec{\rho} - \vec{\rho}'|$, or from Eqs. (18) and (21c) on $v = |\vec{\rho} - \vec{\rho}'|/2$. For Kolmogorov turbulence in the near field, since $\mathcal{D}_\Phi = \mathcal{D}_w$, we may write from Eq. (29)

$$\mathcal{D}_\Phi(v; \epsilon) = 6.88(vD/r_0)^{5/3} \quad (89)$$

It is independent of ϵ , as expected, since it is a characteristic of turbulence.

The MCF for the residual phase aberration after correction of the first J modes is given by

$$M_J(v; \epsilon) = \exp\left\{i\langle [\Phi_J(\vec{\rho}; \epsilon) - \Phi_J(\vec{\rho}'; \epsilon)]^2 \rangle\right\} = \exp[-(1/2)\mathcal{D}_J(v; \epsilon)] \quad (90)$$

The corresponding time-averaged MTF is accordingly given by

$$\langle \tau(v; \epsilon; D/r_0) \rangle_J = \tau(v; \epsilon) \exp[-(1/2)\mathcal{D}_J(v; \epsilon)] \quad (91)$$

The MTF at a certain frequency or the Strehl ratio for a given value of D/r_0 improves as more and more modes are corrected. Of course, $J = \infty$ represents complete correction, that is the aberration-free case.

4.11 SHORT-EXPOSURE IMAGE

It is evident from Tables 4 and 5 that a large portion of the phase aberration introduced by turbulence is a random wavefront tilt caused by large eddies and represented by the coefficients $a_2(\epsilon)$ and $a_3(\epsilon)$. Figure 16 shows the relative tilt-corrected variance $[(\Delta_1(\epsilon) - \Delta_3(\epsilon))/\Delta_1(\epsilon)]$ as a function of ϵ . It increases from a value of approximately 87 percent for a circular pupil to nearly 91 percent as $\epsilon \rightarrow 1$. Thus, if the tilt is corrected in (near) real time with a steering mirror, the variance of the phase aberration for a circular pupil is reduced by a factor of $1.03/0.134 \approx 7.7$. Hence, unless D/r_0 is very large, a significant improvement in the quality of an image is obtained by correcting the wavefront tilt. A short-exposure image is equivalent to a tilt-free image, and a tilt-corrected image yields a time-averaged short-exposure image. For a circular pupil, the phase aberration with a standard deviation of 1 radian is obtained without any correction (except for the piston mode) when $D = r_0$. However, if the x and y tilts are corrected, we find from the expression for Δ_3 , D can be as large as $3.34 r_0$ for one radian of phase aberration.

Now the tilt coefficients $[a_2(\epsilon), a_3(\epsilon)]$ correlate with the primary coma coefficients $[a_8(\epsilon), a_7(\epsilon)]$, secondary coma coefficients $[a_{16}(\epsilon), a_{17}(\epsilon)]$, tertiary coma coefficients $[a_{30}(\epsilon), a_{29}(\epsilon)]$, and so on. Neglecting their correlation with coefficients other than the primary coma coefficients, the tilt-corrected phase structure function may be written⁹

$$\mathcal{D}_3(v; \epsilon) = \mathcal{D}_\Phi(v; \epsilon) - \frac{16v^2}{1 + \epsilon^2} \langle a_2^2(\epsilon) \rangle - \frac{32\sqrt{2}v^2[(1 + \epsilon^2)(6v^2 - 6v + 1) - 2\epsilon^4]}{(1 - \epsilon^2)(1 + 6\epsilon^2 + 10\epsilon^4 + 6\epsilon^6 + \epsilon^8)^{1/2}} \langle a_2(\epsilon)a_8(\epsilon) \rangle \quad (92)$$

It is evident that unlike the uncorrected phase structure function, the tilt-corrected structure function does depend on the diameter D . For a circular pupil, Eq. (92) reduces to

$$\begin{aligned} \mathcal{D}_3(v) &= \mathcal{D}_\Phi(v) - 16v^2 \langle a_2^2 \rangle - 32\sqrt{2}v^2(6v^2 - 6v + 1) \langle a_2 a_8 \rangle \\ &= 6.8839v^{5/3}(1 - 0.9503v^{1/3} - 0.5585v^{4/3} + 0.5585v^{7/3})(D/r_0)^{5/3} \end{aligned} \quad (93)$$

$\mathcal{D}_3(v)$ has a maximum value of $0.344(D/r_0)^{5/3}$ at $v = 1$, which is 1/20 of the value 6.88 $(D/r_0)^{5/3}$ without any correction (i.e., for $J = 1$), illustrating a significant benefit of the tilt correction. When $\epsilon \neq 0$, $\mathcal{D}_3(v; \epsilon)$ is larger for larger values of ϵ , indicating more severe fluctuations of the wavefront.

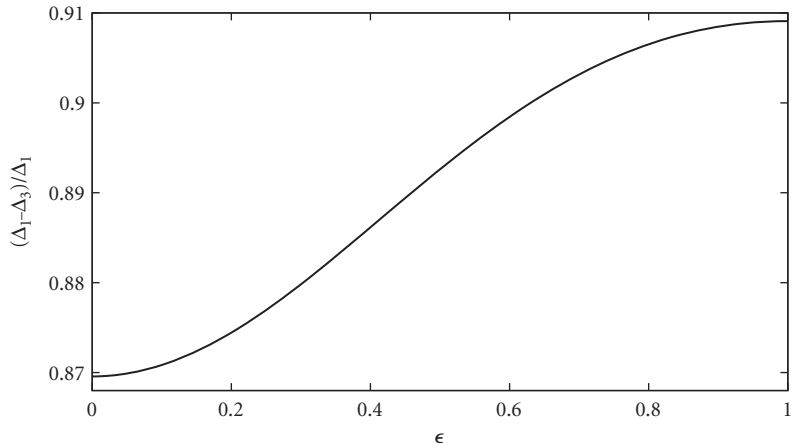


FIGURE 16 Relative tilt-corrected phase variance $[(\Delta_1(\epsilon) - \Delta_3(\epsilon))/\Delta_1(\epsilon)]$ as a function of the obscuration ratio ϵ .

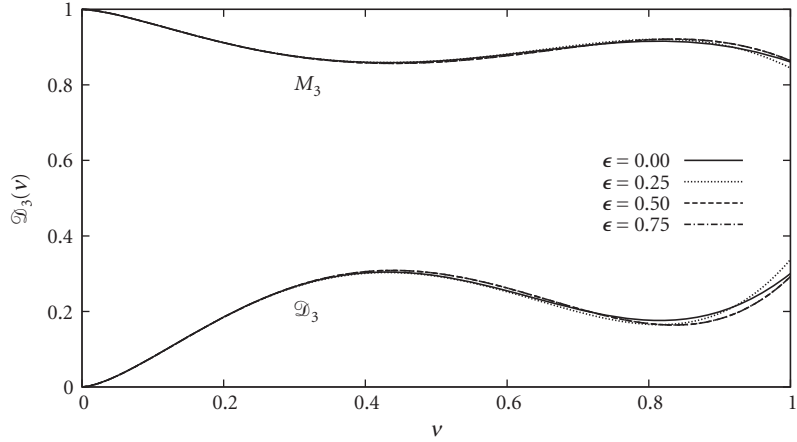


FIGURE 17 Tilt-corrected structure function $\mathcal{D}_3(v; \epsilon)$ and the corresponding mutual coherence function $M_3(v; \epsilon)$.

Using Eq. (92), the short-exposure MTF can be calculated as

$$\langle \tau(v; \epsilon; D/r_0) \rangle_3 = \tau(v; \epsilon) M_3(v; \epsilon; D/r_0) \quad (94)$$

where

$$M_3(v; \epsilon; D/r_0) = \exp[-(1/2)\mathcal{D}_3(v; \epsilon; D/r_0)] \quad (95)$$

is the tilt-corrected or short-exposure MCF or the MTF reduction factor associated with atmospheric turbulence. This factor does depend on the pupil diameter D , in contrast to the long-exposure MTF reduction factor given by Eq. (36), which is independent of D . As expected, it is smaller than the long-exposure reduction factor. Figure 17 shows how the tilt-corrected structure function and the corresponding MCF vary with v . The corresponding long-exposure functions are included for comparison.

The time-averaged Strehl ratio of the tilt-corrected image is given by

$$\langle S(\epsilon; D/r_0) \rangle_3 = \frac{8}{(1 - \epsilon^2)} \int_0^1 \langle \tau(v; \epsilon; D/r_0) \rangle_3 v dv \quad (96)$$

Replacing the long-exposure atmospheric MTF reduction factor in Eq. (47a) by the short-exposure MTF reduction factor, the short-exposure central irradiance η_{SE} may be written

$$\eta_3(\epsilon; D/r_0) = (1 - \epsilon^2) \left(\frac{D}{r_0} \right)^2 \langle S(\epsilon; D/r_0) \rangle_3 \quad (97)$$

The Strehl ratio $\langle S \rangle_3$ and the central irradiance η_3 are shown in Figs. 18 and 19, respectively. Some typical values of the tilt-corrected Strehl ratios are given in Table 1 for various values of D/r_0 and ϵ . For small values of D/r_0 , η_3 increases approximately as for a diffraction-limited system, since the aberration is small; reaches a maximum (e.g., 3.74 for $D/r_0 = 3.5$ for a circular pupil), and then decreases slowly but monotonically to unity. Since the image displacement due to wavefront tilt decreases as $D^{-1/3}$, the effect of tilt correction becomes negligible for large values of D/r_0 due to the large residual phase errors. As in the case of a circular pupil, if the covariance of the tilt coefficients with others is ignored, for example $\langle a_2(\epsilon) a_8(\epsilon) \rangle$ in Eq. (92) is neglected, an unrealistic overcorrection will result.^{11,26,27}

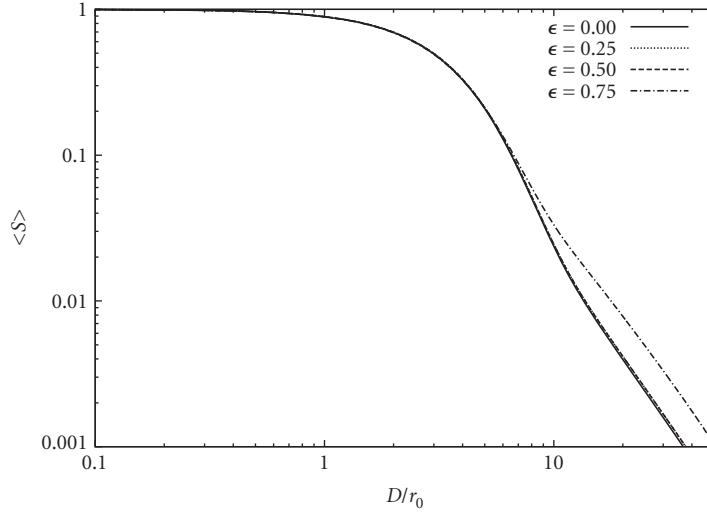


FIGURE 18 Variation of time-averaged short-exposure Strehl ratio $\langle S(\epsilon; D/r_0) \rangle_3$ with D/r_0 .

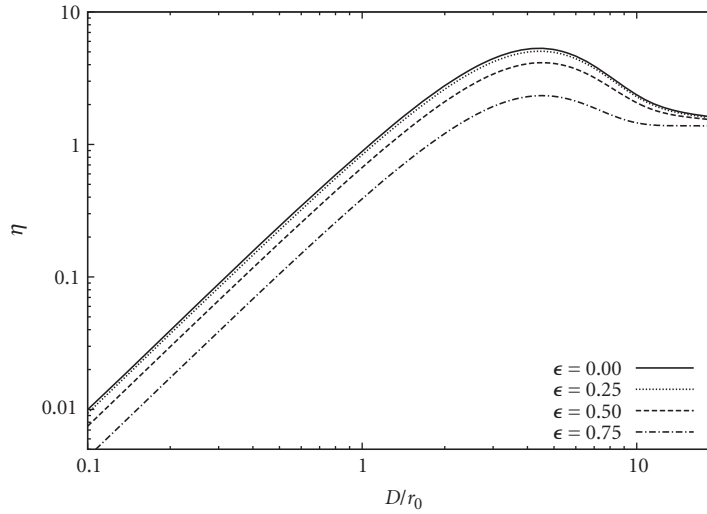


FIGURE 19 Variation of time-averaged short-exposure central irradiance $\langle \eta(\epsilon; D/r_0) \rangle_3$ with D/r_0 .

For a circular pupil, Fig. 20 shows how the tilt-corrected time-averaged Strehl ratio varies with D/r_0 . The uncorrected Strehl ratio is also shown to illustrate the improvement made by the tilt correction. An approximate value of the Strehl ratio given by $\exp(-\Delta_3)$, where

$$\Delta_3 = 0.134(D/r_0)^{5/3} \quad (98)$$

is the tilt-corrected time-averaged phase aberration variance, and shown by the dashed curve underestimates the true value $\langle S \rangle_3$.

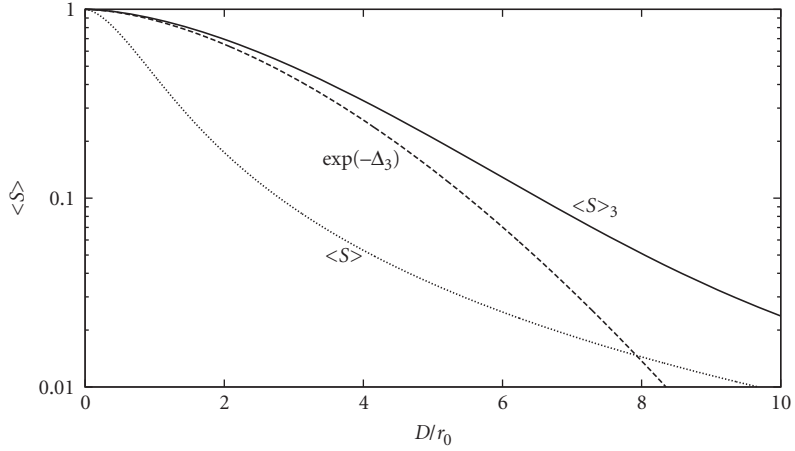


FIGURE 20 Short-exposure Strehl ratio $\langle S \rangle_3$ compared with its approximate value $\exp(-\Delta_3)$. The long-exposure Strehl ratio $\langle S \rangle$ is also shown to illustrate the improvement made by the tilt correction.

An example of an instantaneous short-exposure PSF illustrating the structure of the image of a star as seen by a ground-based telescope with a circular pupil and $D/r_0 = 10$ is shown in Fig. 21, where it is compared with the corresponding aberration-free image. On the average, the standard deviation $\sqrt{\Delta_3}$ of the instantaneous aberration is 2.5 radians or 0.4λ . We note that the image is broken up into small spots called *speckles*, which is a characteristic of large spatially random aberrations. The size of a speckle is determined by D , its angular radius being approximately equal to λ/D . The image size lies between its diffraction-limited value λ/D and its long-exposure value λ/r_0 (varying as $\lambda^{-0.2}$). The image becomes progressively worse as r_0 decreases, showing the effects of what astronomers call *seeing*. However, the size of a speckle decreases as D increases without affecting the long-exposure image size λ/r_0 . Thus, an increase in D does not significantly improve the resolution of the system (as determined by the overall size of the image). The angular spot radius in units of λ/D representing the point at which the irradiance drops to half of its central value is given in Table 7 for both the short- and long-exposure images for various values of D/r_0 and ϵ , and compared with the corresponding

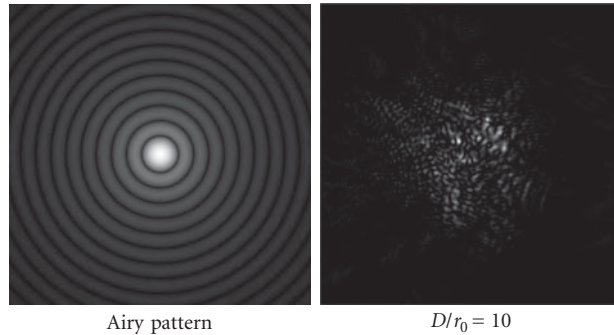


FIGURE 21 Instantaneous short-exposure PSF compared with the corresponding aberration-free PSF.

TABLE 7 Angular Spot Radius of Short-Exposure (SE), Long-Exposure (LE), and Diffraction-Limited (DL) Images in Units of λ/D for Various Values of ϵ and D/r_0

ϵ	$D/r_0 = 1$			$D/r_0 = 2$		$D/r_0 = 5$		$D/r_0 = 10$	
	DL	LE	SE	LE	SE	LE	SE	LE	SE
0	0.515	0.708	0.516	1.122	0.518	2.552	0.560	4.983	2.650
0.25	0.497	0.698	0.498	1.151	0.501	2.595	0.539	5.021	2.686
0.5	0.456	0.688	0.457	1.280	0.459	2.711	0.498	5.106	2.840
0.75	0.406	0.700	0.407	1.480	0.411	3.130	0.465	5.432	3.469

diffraction-limited value. We note that, whereas the LE radius is roughly D/r_0 times the DL radius, the SE radius for small values of D/r_0 , for example $D/r_0 = 5$, is roughly equal to the DL radius, but for large values of D/r_0 , for example $D/r_0 = 10$, it is about half the LE radius. Of course, as ϵ increases and more and more light goes into the diffraction rings, these factors change somewhat. For small values of D/r_0 , the diffraction ring structure of the DL image, smoothed out in the LE image, returns in the SE image. This is more and more evident for larger and larger values of ϵ .

If the image wander is corrected in real time and the image is averaged over time, the speckles disappear and the image becomes smooth. Such PSFs and the corresponding encircled power can be obtained from Eqs. (43) and (44), respectively, by substituting the short-exposure OTF given by Eq. (94). The MTF, PSF, and encircled power thus obtained are shown by the dotted curves in Figs. 5 and 8 for several values of ϵ . The short-exposure image when randomly displaced over time yields a smooth long-exposure image, as illustrated by the dashed curves in Fig. 8. Modal correction of a Gaussian beam (including one with a central obscuration) propagating through turbulence has also been discussed by Wang using the Zernike circle polynomials²⁸ though such polynomials are not orthogonal over the Gaussian weighted (or an annular) pupil.^{5,29}

4.12 ADAPTIVE OPTICS

The correction of wavefront errors in (near) real time is accomplished by using adaptive optics.^{1–3} In practice, a steering mirror with only three actuators is used to correct the large x and y wavefront tilts (also called tip and tilt). The residual aberration is corrected by a deformable mirror, which is deformed by an array of actuators attached to it. The signals for the actuators are determined either by sensing the wavefront errors with a wavefront sensor in a closed loop to minimize the variance of the residual errors, or the actuators are actuated to produce Zernike modes (e.g., focus, two modes of astigmatism, two modes of coma) iteratively until sharpness of the image is maximized.^{30–32} The signals are independent of the optical wavelength, provided atmospheric dispersion is negligible. The two approaches are referred to as *zonal* and *modal approaches*, respectively. The zonal approach has the advantage that the rate of correction is limited only by the rate at which the wavefront errors can be sensed and the actuators can be actuated. However, the amount of light that is used by the wavefront sensor is lost from the image. In practice, the image beam is split into two parts. The centroid of the image of one part is measured with a quad cell, and the tilt indicated by it is corrected with a steering mirror. The resulting tilt-corrected image of the other part with the residual aberration is corrected with a deformable mirror in a closed-loop manner.³³ In the modal approach, there is no loss of light, but the rate or the bandwidth of correction^{34–36} can be slow due to its iterative nature, especially when turbulence is severe and a large number of modes must be corrected. Moreover, for imaging an extended object, wavefront sensing requires a point source in its vicinity, but the modal approach is applicable to the extended object itself.

Of course, adaptive optics can improve the image quality only if the object lies within an isoplanatic angle of turbulence. In the case of a ground-to-space laser illuminating a satellite, the angular travel

(point-ahead angle) of the satellite during the round-trip time of the beam to the satellite must be less than the isoplanatic angle of turbulence. An estimate of the isoplanatic angle can be obtained from³⁷

$$\vartheta_0 = 0.058\lambda^{6/5}(\cos\vartheta)^{8/5} \left\{ \int_0^L C_n^2(z) z^{5/3} dz \right\}^{-3/5} \quad (99)$$

Substituting Eq. (31) into Eq. (99), we obtain $\vartheta_0 = 7 \mu\text{rad}$ for $\lambda = 0.5 \mu\text{m}$. Thus, the $H-V_{5/7}$ model for C_n^2 yields $r_0 = 5 \text{ cm}$ and $\vartheta_0 = 7 \mu\text{rad}$ for visible light; hence the 5/7 subscript.

4.13 SUMMARY

We have considered imaging through atmospheric turbulence by a system with an annular pupil, such as the Keck telescope. The results for a system with a circular pupil are obtained as a special case of those for the annular pupil. An atmospheric coherence length r_0 is defined such that the coherence of two points separated by this distance is 0.03. It is calculated using the well-known $H-V_{5/7}$ model for the refractive index structure parameter C_n^2 , assuming Kolmogorov turbulence, for looking up and down at a point source through the atmosphere as well as for plane wave propagation (see Fig. 2). It is shown that turbulence limits the resolution of a system to one with pupil diameter r_0 regardless of the actual pupil diameter. The two-point angular resolution on ground when observed from any point in space varies approximately as λ/r_0 , compared to the diffraction-limited resolution of λ/D . There is reciprocity in wave propagation through the turbulent atmosphere; otherwise adaptive optics couldn't have worked in ground-based astronomy. Accordingly, the PSF observed on an aircraft looking at a ground point object is the same as the irradiance distribution of a beam focused on ground from an aircraft. Similarly, the irradiance distribution of a beam focused in space from ground is the same as the PSF observed on ground looking at a space object. The effect of turbulence is shown in terms of the MTF, Strehl ratio, PSF, and encircled power for both long- and short-exposure images. How the image improves as more and more Zernike aberration modes are corrected is also discussed. The aberration function may also be expanded in terms of the Karhunen-Loève functions whose coefficients are statistically independent of each other.^{27,38} It is found though that the effect of correlation of Zernike coefficients is negligible for $D/r_0 \leq 4$.²⁶

The dependence of the various image-related quantities on the obscuration ratio ϵ may be summarized as follows:

1. Atmospheric turbulence dominates the degradation of MTF; the MTF gain at high frequencies due to the pupil obscuration is lost even for weak turbulence (compare Figs. 4 and 5). The Strehl ratio is similarly dominated by turbulence (see Fig. 6).
2. The piston-removed aberration variance in units of $(D/r_0)^{5/3}$ increases monotonically from a value of 1.0324 for a circular pupil ($\epsilon = 0$) to a value of 1.843 for an infinitesimally thin ring pupil ($\epsilon \rightarrow 1$) (see Fig. 13).
3. The fluctuations in the angle of arrival depend weakly on the pupil obscuration (see Figs. 12 and 13).
4. The time-averaged variance of some aberration coefficients increases with ϵ and decreases for others (Fig. 9). For example the variance of tilt and astigmatism increases, but that of defocus and spherical aberration decreases and approaches zero as $\epsilon \rightarrow 1$. The covariance follows a similar trend (Fig. 10).
5. When a certain number of Zernike aberration modes is corrected, for example with a deformable mirror, the image quality improves. The residual aberration variance when the first J modes are corrected is given in Table 6. If only tip and tilt modes are corrected, say with a steering mirror, the aberration variance is reduced by 87 percent for a circular pupil and slightly larger (up to 90 percent) for an annular pupil. Figure 19 illustrates the improvement in the Strehl ratio. The aberration-free Airy pattern representing the image of a star breaks up into speckles, as illustrated in Fig. 21. The time-averaged short-exposure OTFs are compared in Fig. 5 with the corresponding long-exposure results. The PSFs and encircled-power curves are similarly compared in Fig. 8.

4.14 ACKNOWLEDGMENTS

V. N. Mahajan gratefully acknowledges helpful discussions with H. T. Yura, who also read the manuscript and made useful comments.

4.15 REFERENCES

1. R. Fugate, "Adaptive Optics," *Handbook of Optics*, 3rd ed, M. Bass (ed.), Chap. 5, Vol. V, McGraw-Hill, New York, 2009.
2. J. W. Hardy, *Adaptive Optics for Astronomical Telescopes*, Oxford, New York, 1998.
3. R. K. Tyson, *Introduction to Adaptive Optics*, SPIE Press, Bellingham, Washington, 1999.
4. D. Fried, "Optical Resolution through a Randomly Inhomogeneous Medium for Very Long and Very Short Exposures," *J. Opt. Soc. Am.* **56**:1372–1379 (1966).
5. V. N. Mahajan, "Zernike Annular Polynomials for Imaging Systems with Annular Pupils," *J. Opt. Soc. Am.* **71**: 75–85 (1981).
6. V. N. Mahajan, "Zernike Annular Polynomials for Imaging Systems with Annular Pupils," *J. Opt. Soc. Am.* **71**: 1408 (1981).
7. V. N. Mahajan, "Zernike Annular Polynomials and Optical Aberrations of Systems with Annular Pupils," *Appl. Opt.* **33**:8125–8127 (1994).
8. V. N. Mahajan, "Orthogonal Polynomials in Wavefront Analysis," *Handbook of Optics*, 3rd ed, M. Bass (ed.), Chap. 11. Vol. II, McGraw-Hill, New York, 2009.
9. G.-M. Dai and V. N. Mahajan, "Zernike Annular Polynomials and Atmospheric Turbulence," *J. Opt. Soc. Am.* **A24**:139–155 (2007).
10. M. Born and E. Wolf, *Principles of Optics*, 7th ed., Oxford, New York, 1999.
11. V. N. Mahajan, *Optical Imaging and Aberrations*, Part II: *Wave Diffraction Optics*, SPIE Press, Bellingham, Washington (Second Printing 2004).
12. D. Fried, "Evaluation of r_0 for Propagation Down through the Atmosphere," *Appl. Opt.* **13**:2620–2622 (1974); errata 1, *Appl. Opt.* **14**:2567 (1975); errata 2, *Appl. Opt.* **16**:549 (1977).
13. R. R. Beland, "Propagation through Atmospheric Turbulence," *Atmospheric Propagation of Radiation, The Infrared & Electro-Optical Systems Handbook*, F. G. Smith (ed.), Springer, 1993. There is a typographical error in the power of 10 in the value of A given in this reference.
14. D. Fried, "Limiting Resolution Looking Down through the Atmosphere," *J. Opt. Soc. Am.* **A 56**:1380–1384 (1966).
15. D. L. Walters and L. W. Bradford, "Measurement of r_0 and θ_0 : Two Decades and 18 Sites," *Appl. Opt.* **36**: 7876–7886 (1997).
16. E. L. O'Neill, "Transfer Function for an Annular Aperture," *J. Opt. Soc. Am.* **46**:285–288 (1956). Note that a term of $-2\eta^2$ is missing in the second of O'Neill's Eq. (26).
17. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York, 1996.
18. V. N. Mahajan and B. K. C. Lum, "Imaging through Atmospheric Turbulence with Annular Pupils," *Appl. Opt.* **20**:3233–3237 (1981).
19. R. J. Noll, "Zernike Polynomials and Atmospheric Turbulence," *J. Opt. Soc. Am.* **66**:207–211 (1976).
20. R. J. Sasiela, *Electromagnetic Wave Propagation in Turbulence*, Springer-Verlag, New York, 1994.
21. V. N. Mahajan, *Optical Imaging and Aberrations*, Part I: *Ray Geometrical Optics*, SPIE Press, Bellingham, Washington (Second Printing 2001).
22. V. I. Tatarski, *The Effects of the Turbulent Atmosphere on Wave Propagation*, U. S. Department of Commerce, 1971.
23. V. N. Mahajan, "Line of Sight of an Aberrated Optical System," *J. Opt. Soc. Am.* **A2**:833–846 (1985).
24. W. Wetherell, "The Calculation of Image Quality," *Applied Optics and Applied Engineering*, R. R. Shannon and J. C. Wyant (eds.), Academic Press, 1980.
25. V. N. Mahajan, "Strehl Ratio for Primary Aberrations in Terms of Their Aberration Variance," *J. Opt. Soc. Am.* **73**:240–241 (1983).

26. J. Y. Wang, "Optical Resolution through a Turbulent Medium with Adaptive Phase Compensation," *J. Opt. Soc. Am.* **67**:383–390 (1977).
27. G.-M. Dai, "Modal Compensation of Atmospheric Turbulence with the Use of Zernike Polynomials and Karhunen Loève Functions," *J. Opt. Soc. Am.* **A12**:2182–2193 (1995).
28. J. Y. Wang, "Phase-Compensated Optical Beam Propagation through Atmospheric Turbulence," *Appl. Opt.* **17**:2580–2590 (1978).
29. V. N. Mahajan, "Zernike-Gauss Polynomials and Optical Aberrations of Systems with Gaussian Pupils," *Appl. Opt.* **34**:8057–8059 (1995).
30. R. A. Miller and A. Buffington, "Real-Time Wavefront Correction of Atmospherically Degraded Telescopic Images through Image Sharpening," *J. Opt. Soc. Am.* **61**:1200–1210 (1974).
31. A. Buffington, F. S. Crawford, R. A. Miller, A. J. Schwemin, and R. G. Smits, "Correction of Atmospheric Distortion with an Image-Sharpening Telescope," *J. Opt. Soc. Am.* **67**:298–305 (1977).
32. V. N. Mahajan, J. Govignon, and R. J. Morgan, "Adaptive Optics without Wavefront Sensors," *SPIE Proc.* **228**:63–69 (1980).
33. L. C. Roberts, Jr. and C. R. Neyman, "Characterization of the AEOS Adaptive Optics System," *Pub. Astro. Soc. Pacific* **114**:1260–1266 (2002).
34. J. Y. Wang, "Effect of Finite Bandwidth on Far-Field Performance of Modal Wavefront-Compensative Systems," *J. Opt. Soc. Am.* **69**:819–828 (1977).
35. C. B. Hogge and R. R. Butts, "Frequency Spectra for the Geometric Representation of Wavefront Distortions Due to Atmospheric Turbulence," *IEEE Trans. Antennas Propagation* **AP-24**:144–154 (1976).
36. D. P. Greenwood, "Bandwidth Specification of Adaptive Optics Systems," *J. Opt. Soc. Am.* **67**:390–393 (1977).
37. D. L. Fried, "Anisoplanatism in Adaptive Optics," *J. Opt. Soc. Am.* **72**:52–61 (1982).
38. J. Y. Wang and J. K. Markey, "Modal Compensation of Atmospheric Turbulence Phase Distortion," *J. Opt. Soc. Am.* **68**:78–87 (1978).

ADAPTIVE OPTICS

Robert Q. Fugate

Starfire Optical Range

Directed Energy Directorate

Air Force Research Laboratory

Kirtland Air Force Base, New Mexico

5.1 GLOSSARY

C_n^2	refractive index structure parameter
d_0	section size for laser beacon adaptive optics
$\mathcal{D}_n$	refractive index structure function
$\mathcal{D}_\phi$	phase structure function
E_f^2	mean square phase error due to fitting error
E_{FA}^2	mean square phase error due to focus anisoplanatism
E_n^2	mean square phase error due to sensor read noise
E_s^2	mean square phase error due to servo lag
f_G	Greenwood frequency
f_{T_G}	Tyler G-tilt tracking frequency
$F_\phi(f)$	phase power spectrum
$F_{\phi_G}(f)$	G-tilt power spectrum
$H(f, f_c)$	servo response function
N_{pde}	number of photo-detected electrons per wavefront subaperture
r_0	Fried's coherence length
SR_{HO}	Strehl ratio resulting from higher-order phase errors
SR_{tilt}	Strehl ratio due to full-aperture tilt
$Z_j(\rho, \theta)$	Zernike polynomials
σ_θ	rms full-aperture tracking error
σ_{θ_G}	rms full-aperture G-tilt induced by the atmosphere
θ_0	isoplanatic angle

5.2 INTRODUCTION

An Enabling Technology

Adaptive optics (AO) is *the* enabling technology for many applications requiring the real-time control of light propagating through an aberrating medium. Today, significant advances in AO system performance for scientific, commercial, and military applications are being made because of improvements in component technology, control algorithms, and signal processing. For example, the highest-resolution images of living human retinas ever made have been obtained using AO.^{1–3} Increased resolution of retinal features could provide ophthalmologists with a powerful tool for the early detection and treatment of eye diseases. Adaptive optics has brought new capabilities to the laser materials processing industry, including precision machining of microscopic holes and parts.^{4,5} Beaming power to space to raise satellites from low earth orbit to geosynchronous orbit, to maneuver satellites that are already in orbit, and to provide electricity by illuminating solar panels remains a topic of commercial interest.^{6,7} Recent work has shown that high-energy pulsed lasers, corrected for atmospheric distortions with AO, could be used to alter the orbits of space debris objects,⁸ causing them to reenter the atmosphere much sooner than natural atmospheric drag—creating an opportunity for the environmental restoration of space. High-speed optical communication between the ground, aircraft, and spacecraft (including very deep space probes) is a topic of continuing interest that would benefit from the use of AO.⁹ Adaptive optics offers significant benefit to a large variety of military applications, including the U.S. Air Force Airborne Laser program, in which a high-powered chemical oxygen-iodine laser will be mounted in an aircraft to engage boosting missiles at long ranges.¹⁰ Adaptive optics will be used to predistort the laser beam as it leaves the aircraft to compensate for the atmospheric distortions that the beam encounters as it propagates to the missile target.

Perhaps the most widely known and discussed application for AO is imaging through the atmosphere. Military sites are now using AO on medium-sized ground-based telescopes [of the order of 4 meters (m)] to inspect low-earth-orbiting satellites. However, the most dramatic use of AO is in astronomy. Astronomers have been plagued by the atmosphere for centuries—since the invention of the telescope. Indeed, distortion caused by turbulence is the principal motivation for launching large telescopes into space. However, if ground-based telescopes were able to achieve diffraction-limited resolution and high Strehl ratios, a significant obstacle to new discoveries and more productive research would be overcome. Adaptive optics, originally proposed by the astronomer Horace Babcock in 1953,¹¹ may enable that to happen, creating a revolution in ground-based optical astronomy. Adaptive optics is especially important in light of the large new telescope mirrors. Incredible advances in mirror technology have made 8-m-diameter monolithic mirrors almost commonplace (four have seen first light and five more are under construction). These are in addition to the two 10-m Keck telescopes and the 11-m Hobby-Eberly telescope, all of which use segmented primaries. The atmosphere limits the resolution of ground-based telescopes to an equivalent diameter that is equal to Fried's coherence length, r_0 (a few tens of centimeters at visible wavelengths at the best sites).¹² Adaptive optics offers the potential for telescopes to achieve diffraction-limited imaging and high throughput to spectroscopic instruments. Furthermore, interferometric arrays of large telescopes will enable unprecedented imaging resolution of faint objects, given that the unit telescopes can each produce nearly distortion-free wave fronts. Because imaging through the atmosphere is a topic of great interest, and because it embodies all aspects of AO technology, this discussion is oriented toward that application.

The chapter is organized as follows: Sec. 5.3 describes the basic concept of AO as applied to ground-based telescopes. Section 5.4 is a short summary of the classical description of atmospheric turbulence and parameters that are important for designing and evaluation of AO systems. Section 5.5 is a description of the hardware and the software that are needed for practical implementation of AO. This includes tracking, wavefront sensing, processors, and wavefront correctors. Section 5.6 is a discussion of design issues for top-level system performance trades using scaling laws and formulas.

5.3 THE ADAPTIVE OPTICS CONCEPT

This section provides a brief, qualitative overview of how AO imaging systems work. Figure 1 is a highly simplified diagram that illustrates the principles. In this figure, a conventional telescope is shown on the left and one that is equipped with AO is shown on the right. The science object

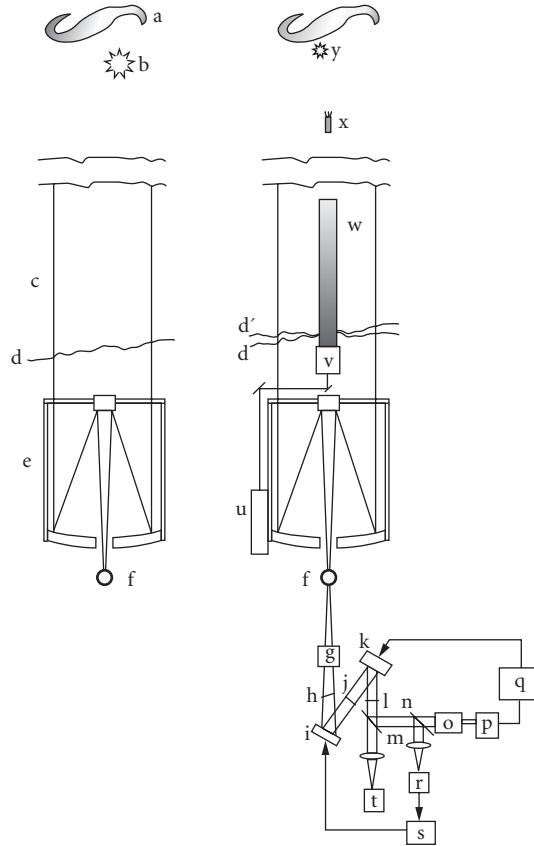


FIGURE 1 Conventional telescope with no AO (left) and additional components needed for AO (right). (a) Object of interest, (b) natural guide star, (c) atmospheric turbulence, (d) aberrated wavefront (including tilt and higher-order distortions) after passing through the turbulent atmosphere, (d') aberrated wavefront from the laser beacon with full-aperture tilt removed, (e) telescope, (f) aberrated image, (g) relay optics, (h) demagnified aberrated wavefront including full-aperture tilt, (i) fast-steering mirror, (j) tilt-removed wavefront with only higher-order aberrations remaining, (k) deformable mirror, (l) corrected wavefront, (m) aperture-sharing element, (n) tilt sensor pickoff beam splitter, (o) relay optics, (p) higher-order wavefront sensor, (q) electronic processor to compute deformable mirror commands, (r) full-aperture tilt sensor, (s) tilt mirror processor and controller, (t) science camera, (u) laser for generating laser beacons, (v) launch telescope for the beacon laser, (w) Rayleigh laser beacon, (x) mesospheric sodium laser beacon, and (y) faint natural guide star for full-aperture tilt sensing.

(a) and a natural guide star (b) are at the top of the figure. The turbulent atmosphere (c) creates higher-order phase distortion and an overall tilt on wavefronts reaching the telescope (d). The telescope (e) forms an aberrated image (f) at the location of the camera or spectrograph. The natural guide star is used, either manually or automatically, to correct pointing errors in the telescope mount and the overall wavefront tilt that are induced by atmospheric turbulence.

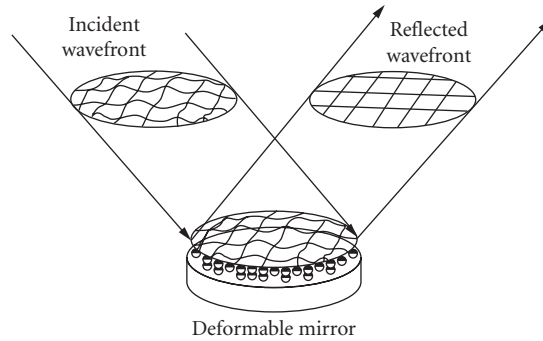


FIGURE 2 Deformable mirror concept in which phase aberrations on an incident wavefront are removed by setting the deformable mirror to the conjugate of the aberration.

The objectives of the AO system are to continuously remove the higher-order distortion and to stabilize the position of the image by removing the overall tilt. The components are shown on the right side of Fig. 1. One can think of the AO system as a dynamic optical system that simultaneously relays the original image to a new focal plane [the camera (*t*)] while removing both the higher-order aberrations (those distortions having spatial frequencies higher than tilt) with the deformable mirror (*k*) and the tilt aberrations with the fast-steering mirror (*i*), leaving only a minor amount of residual error in the wave front (*l*). The appropriate optical relay telescopes (*g* and *o*) are used as required by the particular implementation.

Figure 2 illustrates how a deformable mirror removes phase aberrations that are induced by atmospheric turbulence. The conjugate of the measured wavefront distortion is imposed on the surface of the deformable mirror so that on reflection, the distortions are removed. The AO system is never able to perfectly match the distortions in the wavefront because of a number of error sources that will be discussed in later sections.

To set the figure on the deformable mirror, the AO system must get information about the turbulent atmosphere by measuring its effect on the wavefronts from a beacon—a source of light such as a bright star (see Fig. 1, item *b*) located at or near the science object. The object itself may serve as the beacon if it is bright enough (either by self-emission, as in the case of a star or astronomical object, or by reflection of natural or man-made light, as in the case of artificial satellites). When a natural star does not exist or when the object itself is not bright enough, it may be possible to use an artificial beacon generated by a low-power laser (see Fig. 1, item *u*). A laser beacon can be created either by Rayleigh scattering up to a range of about 20 km, or by resonant scattering of atomic species such as sodium in the mesosphere, at an altitude of 90 km above Earth's surface. In the implementation that is shown in Fig. 1, a laser beam that is capable of exciting the D_2 line in atomic sodium in the mesosphere is projected from behind the secondary mirror of the telescope.

Control signals for the wavefront corrector mirrors are generated by a full-aperture tracking sensor (see Fig. 1, item *r*) and a wavefront sensor (*p*) by observing the residual error in the beacon wavefront (*l*). An aperture-sharing element (*m*) directs light from the beacon into the wavefront sensor, which is located at an image of the entrance pupil of the telescope. The wavefront sensor samples the slope of the wavefront over subaperture regions of the order of r_0 in size—for which the wavefront is essentially an undistorted, but tilted, plane wave—measuring the tilt and reporting that tilt as a wavefront phase gradient. The computer (*q*) combines the subaperture gradient measurements and “reconstructs” a best estimate of the residual phase error at specific points in the aperture, generally the locations of the actuators of the deformable mirror. Error signals derived from the reconstructed wavefront are sent to the deformable mirror to further reduce the residual error.

The tracking sensor measures the overall tilt of the beacon wavefront by computing the centroid of a focused image of the beacon that is formed by the full aperture of the telescope (or by other, more sophisticated means). The tracker processor (*s*) generates an error signal that controls the fast-steering mirror to keep the image of the beacon centered on the tracking sensor. It is also possible to derive full-aperture tilt information from the higher-order wavefront sensor, but optimum tracking performance usually requires a specialized sensor and processor. When a laser beacon is used to obtain higher-order wavefront information, it is still necessary to use a natural guide star to derive full-aperture tilt information: The laser beacon's position in the sky is not known with respect to an inertial reference (like a star), because its path on the upward propagation is random. To first order the beacon does not appear to move at all in the tracker due to near perfect reciprocity on the upward and downward paths, generating a return wavefront having no full aperture tilt as shown at (*d'*) in Fig. 1. The requirement to use a natural guide star for tracking is not as serious as it first sounds, however, because it is much more likely that a natural guide star can be found near the science object that satisfies tracking requirements but is still too faint for higher-order wavefront measurements. This is possible because the full aperture of the telescope and a small number of specialized detectors can be used for tracking so that stars fainter by a factor of $\sim (D/r_0)^2$ that are available to the wavefront sensor become available to the tracking sensor.

Figure 1 is highly simplified. The details of the optics have been left out to emphasize the main principles. For instance, it is customary to form an image of the entrance pupil of the telescope onto the fast-steering mirror, the deformable mirror, and again on the wavefront sensor. At least two powered optical elements are required to generate each of these images. Even more sophisticated approaches have been proposed in which multiple deformable mirrors are reimaged to layers in the turbulence. The nature of the aperture-sharing element will vary significantly depending on the adopted philosophy for the system. For example, if obtaining enough light for the wavefront sensor is a problem (as is usually the case in astronomy), it is prudent to send the entire spectrum over which the detectors are responsive to the wavefront sensor and divert only the absolute minimum needed part of the spectrum to the imaging sensor. This approach also means that the optics must be designed to work over a wide spectral range (e.g., from 450 to 1000 nm for a visible wavefront sensor) and also requires careful design of atmospheric dispersion correction. Furthermore, there are innovative approaches for implementing the wavefront corrector mirrors. The Steward Observatory group is developing a tilting, deformable secondary mirror for the new 6.5-m primary replacement at the Multiple Mirror Telescope (MMT) and 8.4-m Large Binocular Telescope (LBT) observatories.¹³ This arrangement requires only two reflecting surfaces between the sky and the camera.

5.4 THE NATURE OF TURBULENCE AND ADAPTIVE OPTICS REQUIREMENTS

Turbulence Generation and the Kolmogorov Model

This section summarizes the classical description of atmospheric turbulence and its effect on the propagation of light. The literature on this topic is enormous. Our purpose here is to introduce the principles and define a few important parameters that are relevant to the operation of adaptive optical systems. The references cited throughout provide more detail.

Atmospheric turbulence is generated by solar heating of the Earth's surface. Air at the surface of the Earth is warmed in the day and cooled at night. Temperature gradients develop, creating a convective flow of large masses of air. Turbulence develops as large-scale masses of air break up into smaller spatial scales and dissipate their energy to the surrounding air. The effects of turbulence on electromagnetic wave propagation is governed by the nature of the spatial and temporal fluctuations of the index of refraction of air. The refractive index of air at optical wavelengths obeys the formula¹⁴

$$n = 1 + 77.6 \left[1 + \frac{7.52 \cdot 10^{-3}}{\lambda^2} \right] \frac{P}{T} \cdot 10^{-6} \quad (1)$$

where λ is the wavelength of light in micrometers, P is the atmospheric pressure in millibars, and T is the atmospheric temperature in kelvins (K). The effect of atmospheric pressure changes on n are small and can, for problems of interest in this chapter, be neglected. The dominant influence on variations of n is the air temperature. At visible wavelengths, $dn/dT \sim 10^{-6}$ at standard pressure and temperature. This means that two 1-m-long columns of air having a temperature difference of 1 K create roughly one wave of optical path difference for visible light, a very significant effect. We can model the index of refraction as the sum of two parts,

$$n(\mathbf{r}, t, \lambda) = n_0(\mathbf{r}, \lambda) + n_1(\mathbf{r}, t) \quad (2)$$

where n_0 represents the deterministic, slowly changing contribution (such as variation with height above the ground), and n_1 represents the random fluctuations arising from turbulence. The wavelength dependence of n_1 is ignored. Furthermore, typical values of n_1 are several orders of magnitude smaller than unity.

During the 1940s, Andrey Nikolayevich Kolmogorov (1903–1987) developed theories¹⁵ describing how energy is dissipated in the atmosphere and modeled the spatial power spectrum of turbulent velocity fluctuations. V. I. Tatarskii postulated the applicability of the Kolmogorov spatial power spectrum to refractive index fluctuations and then solved the wave equation for the power spectrum of Kolmogorov's model and determined the effect on propagation through weak turbulence.¹⁶ David L. Fried subsequently extended Tatarskii's results to describe phase distortions of turbulence in terms of Zernike polynomials¹⁷ and derived his famous atmospheric coherence diameter parameter r_0 , as a measure of optical resolution.¹² Nearly all subsequent work in statistical atmospheric optics is based on the contributions of Kolmogorov, Tatarskii, and Fried. The effects of turbulence on wave propagation are critical because the spatial and temporal frequency distribution and optical depth of the phase aberrations drive the design requirements for beacon brightness, wavefront sensing, tracking, electronic processing, and wavefront corrector mirror subsystems in a real-time phase compensation system.

For refractive index fluctuations, the Kolmogorov spatial power spectral density, $\Phi_n(\kappa)$, is given by

$$\Phi_n(\kappa) = 0.033 C_n^2 \kappa^{-11/3} \quad (3)$$

where κ is the spatial wavenumber and C_n^2 is the refractive index structure constant, a measure of the optical strength of turbulence. The wavenumber κ is inversely related to the size of the turbulent eddies in the atmosphere. Equation (3) is valid between two limits of κ called the inertial subrange. The scale size of eddies for the smallest values of κ in the inertial subrange is called the outer scale, denoted $L_0 = 2\pi/\kappa_0$, and is of the order of meters to kilometers. The inner scale, $l_0 = 2\pi/\kappa_m$, or smallest eddies in the inertial subrange, is of the order of millimeters. For values of κ outside the inertial subrange, the von Kármán spectrum is often used in place of Eq. (3) to avoid mathematical complications as κ approaches zero. The von Kármán spectrum has the form

$$\Phi_n(\kappa) \equiv \frac{0.033 C_n^2}{(\kappa^2 + \kappa_0^2)^{11/6}} \exp\left(-\frac{\kappa^2}{\kappa_m^2}\right) \quad (4)$$

where κ_0 and κ_m are the spatial wavenumbers for the outer and inner scales.

The Variation of n and the C_n^2 Parameter

Tatarskii introduced structure functions to describe the effect of turbulence on wave propagation. Structure functions describe how a physical parameter is *different* between two points in space or time. The structure function of the index of refraction, $\mathcal{D}_n(\mathbf{r})$ is defined as

$$\mathcal{D}_n(\mathbf{r}) = \{[n_1(\mathbf{r}') - n_1(\mathbf{r}' - \mathbf{r})]^2\} \quad (5)$$

where $n_1(\mathbf{r})$ represents the fluctuating part of the index of refraction and $\{ \dots \}$ represents an ensemble average over the turbulent conditions. The autocorrelation function and power spectral density form a three-dimensional Fourier transform pair (a useful relation used many times in statistical optics), and using that relationship, it can be shown that¹⁸

$$\mathcal{D}_n(r) = C_n^2 r^{2/3} \quad (6)$$

where we have also assumed that the fluctuations are isotropic and dropped the vector dependence on r . Equation (6) is the defining equation for the refractive index structure constant, C_n^2 , in that the numerical coefficient on the right side of Eq. (6) is unity. Defined in this way, C_n^2 is a measure of the optical strength of turbulence.

A few meters above the ground, C_n^2 has an average value of the order of $10^{-14} \text{ m}^{-2/3}$, rapidly decreasing by three or four orders of magnitude at heights above 10 km. Several means have been developed to measure the C_n^2 profile, including in situ instruments on balloons as well as remote sensing optical and radar techniques.^{19,20} Several mathematical models of C_n^2 profiles have been developed based on experimental data. One of the most widely used was suggested by Hufnagel²¹ and modified to include boundary layer effects by Valley.²² The expression for the Hufnagel-Valley model for C_n^2 is

$$C_n^2(h) = 5.94 \cdot 10^{-23} h^{10} e^{-h} \left(\frac{W}{27} \right)^2 + 2.7 \cdot 10^{-16} e^{-2h/3} + A e^{-10h} \quad (7)$$

where h is the height above the site in kilometers, W is an adjustable wind correlating parameter, and A is a scaling constant, almost always taken to be $A = 1.7 \cdot 10^{-14}$. Winker²³ suggested $W = 21$ producing the $\text{HV}_{5/7}$ model named from the fact that the resulting profile yields a value of Fried's coherence diameter of $r_0 = 5 \text{ cm}$ and an isoplanatic angle of $\theta_0 = 7 \mu\text{rad}$ for zenith propagation at $0.5 \mu\text{m}$. (The parameters r_0 and θ_0 are defined and their relationships to C_n^2 are given in Sec. 5.4.) The $\text{HV}_{5/7}$ model has been widely used in the evaluation of AO system performance. Figure 3 shows how C_n^2 , computed with this model, varies with altitude.

The Hufnagel-Valley model is useful for continental sites or sites with well-developed boundary layers. The C_n^2 profile, which is associated with mountaintop sites that are surrounded by water

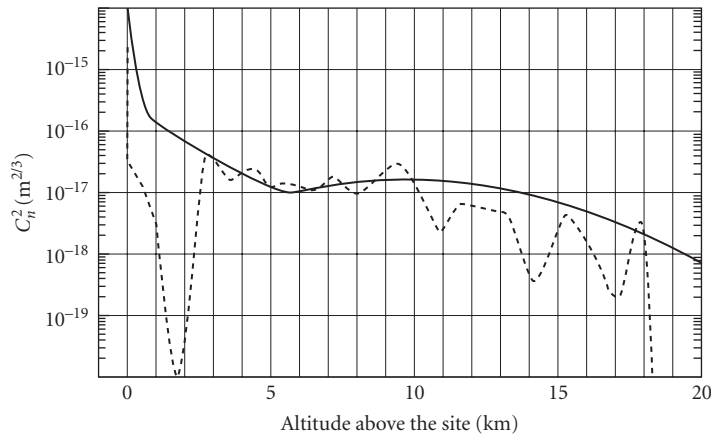


FIGURE 3 C_n^2 profiles for the $\text{HV}_{5/7}$ model (solid curve) and for average seeing conditions at Mauna Kea (dotted curve).

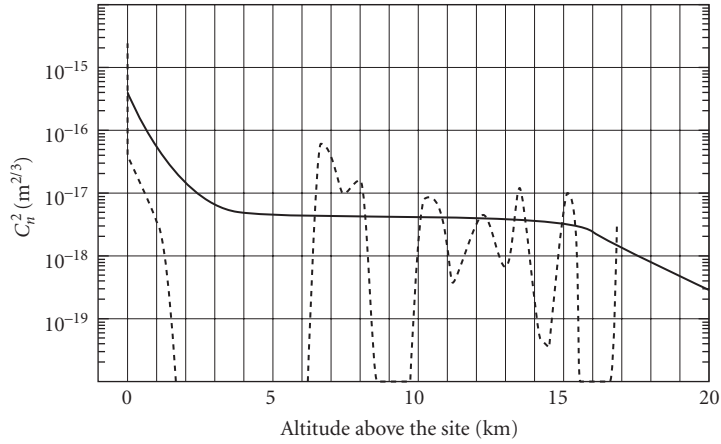


FIGURE 4 Average seeing C_n^2 profile analytical model for Paranal, Chile [site of European Southern Observatory's (ESO's) Very Large Telescope (VLT) (solid curve)], and a C_n^2 profile that is representative of best seeing conditions (occurring less than 10 percent of the time) at Mauna Kea (dotted curve) and represents 0.25 arcsec seeing. Note how the turbulence appears as layers and the lack of a boundary layer in the Mauna Kea profile. Similar conditions have also been observed at Paranal and La Palma.

(e.g., Mauna Kea or La Palma) or mountaintop sites that are close to a coastline [so as to be above the marine boundary layer (e.g., Paranal, Chile)], often exhibits distinct layers of turbulence, but little or no boundary layer in the first few kilometers above the ground. A C_n^2 profile that is representative of average seeing for Mauna Kea is also shown in Fig. 3 and yields an r_0 of 18 cm and θ_0 of 15 μ rad. Figure 4 shows a C_n^2 profile for Mauna Kea representative of the 90th-percentile best seeing (note the distinct layering), giving an r_0 of 45 cm, and an analytical model that is used for average seeing at Paranal,²⁴ giving an r_0 of 18 cm.

Wave Propagation in Turbulence

An electromagnetic wave propagating in the atmosphere must satisfy Maxwell's equations—in particular, the electric and magnetic fields must satisfy the time-independent Helmholtz equation:

$$\nabla^2 U(\mathbf{r}) + n^2 k^2 U(\mathbf{r}) = 0 \quad (8)$$

where $k = 2\pi/\lambda$ and n is the refractive index of air. $U(\mathbf{r})$ is the complex phasor representation of a spatial wave:

$$U(\mathbf{r}) = A(\mathbf{r})e^{i\phi(\mathbf{r})} \quad (9)$$

where the amplitude $A(\mathbf{r})$ and phase $\phi(\mathbf{r})$ are random variables governed by the statistics of the fluctuations of the refractive index, n . There is no closed-form solution for $A(\mathbf{r})$ and $\phi(\mathbf{r})$.

The surfaces in space that are defined by $\phi_0(\mathbf{r}) = \text{constant}$ are wavefronts. For starlight arriving at the top of the atmosphere, the wavefronts are planes of infinite extent. When a plane wave passes

through the atmosphere, whose refractive index varies randomly along each propagation path, the plane wave becomes distorted. The total phase change from that in a vacuum along a path z is

$$\Delta\phi = \phi_0 - k \int_0^z n_1(z) dz \quad (10)$$

where $k = 2\pi/\lambda$, and $n_1(z)$ is the fluctuating part of the index of refraction. If this equation could be evaluated along every direction, one could construct the three-dimensional shape of the wavefront at any position along its propagation path due to phase changes induced by atmospheric turbulence. This is, of course, an intractable problem in terms of a closed-form solution since the fluctuations of the index of refraction are a random process.

A phase deformation that propagates far enough becomes an amplitude fluctuation as different parts of the propagating wavefront combine to create regions of constructive and destructive interference. This effect is called scintillation—the twinkling of starlight is a famous example. Amplitude fluctuations are generally not severe for astronomical observing, and 70 to 80 percent or more of the distortions can be removed by phase-only compensation (by using a steering mirror and deformable mirror). However, both the amplitude and phase of the wave are random variables, and the most general representation of the wave is as given in Eq. (9).

The historically accepted approach to dealing with wave propagation in turbulence was introduced by Tatarskii by use of the Rytov transformation. The Rytov transformation defines a complex quantity ψ as the natural logarithm of U : $\psi = \ln U$. This substitution allows a solution to the wave equation in the form of a multiplicative perturbed version of the free space field:

$$U(\mathbf{r}) = A_0(\mathbf{r}) e^{i\phi_0(\mathbf{r})} e^{\ln[A(\mathbf{r})/A_0(\mathbf{r})] - i[\phi_0(\mathbf{r}) - \phi(\mathbf{r})]} \quad (11)$$

Since the field at any point is the superposition of many independent contributions of propagating waves in the turbulent medium, we expect the fluctuating parts of the field to obey Gaussian statistics (invoking the central limit theorem). This means if $\ln(A/A_0)$ obeys Gaussian statistics, we expect the amplitude fluctuations to be log-normally distributed. Most experimental evidence supports this for weak fluctuations. Tatarskii's results make it possible to compute important characteristic functions and to use the properties of Gaussian random processes to develop practical descriptions of the effects of turbulence.

Fried's Coherence Diameter, r_0 , and the Spatial Scale of Turbulence

Hufnagel and Stanley²⁵ extended Tatarskii's work and developed an expression for the modulation transfer function in terms of the mutual coherence function. Fried²⁶ developed an expression for *phase* structure function, $\mathcal{D}_\phi$, of a propagating electromagnetic plane wave showing that it is proportional to the 5/3 power of spatial separation, r , and is given by*

$$\mathcal{D}_\phi(r) = \{[\phi(r') - \phi(r' - r)]^2\} = \left[2.91 \left(\frac{2\pi}{\lambda} \right)^2 \int C_n^2(z) dz \right] r^{5/3} \quad (12)$$

where λ is the wavelength of propagation, $C_n^2(z)$ is the position-dependent refractive index structure constant and integration is over the optical path to the source.

*In the more general case where amplitude effects are significant, the *wave* structure function, $\mathcal{D}(r) = \mathcal{D}_\phi(r) + \mathcal{D}_\psi(r)$, should be used, but since it has exactly the same value as given by Eq. (12), and since conventional AO does not correct intensity fluctuations, the amplitude term will be dropped for the material that is presented here.

Fried²⁶ developed a very useful relationship between the phase structure function and a particular measure of optical resolution—the volume under the two-dimensional optical transfer function. Defined in this way, the seeing-limited resolving power asymptotically approaches a limiting value as the aperture size increases. The limiting value is set by the strength of the turbulence. Fried defined a quantity called the coherence diameter, r_0 , such that the limiting resolution that is obtained in the presence of atmospheric turbulence is the same as that that is obtained by a diffraction-limited lens of diameter r_0 in a vacuum. Its value can be computed from the definition¹⁷

$$r_0 = 0.185\lambda^{6/5}(\cos\psi)^{3/5} \left[\int_0^\infty C_n^2(h) dh \right]^{-3/5} \quad (13)$$

where λ is the wavelength, $C_n^2(h)$ is the vertical profile of the index of refraction structure constant, h is the height above the ground, ψ is the angle between the propagation direction and the zenith, and the integral is evaluated from the ground to an altitude at which $C_n^2(h)$ no longer contributes significantly (typically 20 km). This equation explicitly shows r_0 scales with wavelength as $\lambda^{6/5}$ and with zenith angle as $(\cos\psi)^{3/5}$. By convention, the numerical values of r_0 are usually expressed at 0.5 μm for zenith viewing. At the best astronomical sites, the median value of r_0 is from 15 to 25 cm, corresponding to seeing conditions of $\lambda/r_0 = 0.7$ to 0.4 arcsec. The best intracontinental mountain-top sites exhibit median values of r_0 between 10 and 20 cm, and average intracontinental sites have r_0 values between 5 and 10 cm.²⁰

Given the definition of r_0 in Eq. (13), the phase structure function becomes

$$\mathcal{D}_\phi(r) = 6.88 \left(\frac{r}{r_0} \right)^{5/3} \quad (14)$$

Fried defined r_0 so that the knee in the curve showing resolving power as a function of diameter occurs at $D/r_0 = 1$, where D is the aperture diameter of the imaging telescope. Most scaling laws of interest to AO are expressed in terms of the ratio D/r_0 .

Turbulence Noll²⁷ developed a description of the average spatial content of turbulence-induced wavefront distortion in terms of Zernike polynomials. The Zernike polynomials are often used to describe the classical aberrations of an optical system—tilt, defocus, astigmatism, coma, and spherical aberration. Noll expresses the phase of the distorted wavefront over a circular aperture of radius, R , as

$$\phi(\rho, \theta) = \sum_j a_j Z_j(\rho, \theta) \quad (15)$$

where Z_j is a modified set of the orthogonal Zernike polynomials defined over points in the aperture at reduced radius $\rho = r/R$ and azimuthal angle θ . The Z_j is given by

$$\left. \begin{aligned} Z_{\text{even}j} &= \sqrt{n+1} R_n^m(\rho) \sqrt{2} \cos m\theta \\ Z_{\text{odd}j} &= \sqrt{n+1} R_n^m(\rho) \sqrt{2} \sin m\theta \end{aligned} \right\} m \neq 0 \quad (16)$$

$$Z_j = \sqrt{n+1} R_n^0(\rho) \quad m = 0 \quad (17)$$

where

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s! [(n+m)/2 - s]! [(n-m)/2 - s]!} \rho^{n-2s} \quad (18)$$

The indices n and m are always postive integers and satisfy $m \leq n$ and $n - m = \text{even}$. The index, j , is a mode-ordering number and depends on n and m .

The a_j are coefficients having mean square values that accurately weight the particular Zernike mode to represent the Kolmogorov distribution of turbulence. Using this approach, Noll computed the values of a_j and showed that the piston-removed wavefront distortion for Kolmogorov turbulence averaged over an aperture of diameter, D , expressed as a mean square value in units of square radians (rad^2) of optical phase is

$$\langle \phi^2 \rangle = 1.0299 \left(\frac{D}{r_0} \right)^{5/3} \quad (19)$$

Table 1 lists the strengths of the first 21 Zernike modes, and the residual mean square distortion as each component is removed. Data in this table show that the two components of tilt (X and Y axes) make up 87 percent of the total wavefront distortion.

Noll's results are useful for estimating the performance of an AO system if one can estimate how many Zernike modes the system will correct (a rule of thumb is approximately one mode is corrected per deformable mirror actuator for systems having large numbers of actuators). Figure 5 shows the mean square residual phase error for a range of values of D/r_0 . Generally, the mean square residual error should be much less than 1 rad^2 . Even if all 21 modes listed in Table 1 are corrected, the mean square residual error for conditions in which D/r_0 is 20, for example, $0.0208(20)^{5/3} = 3.06 \text{ rad}^2$, is a very significant, generally unacceptable error.

When more than 21 modes are corrected, the residual mean square wavefront error is given by²⁷

$$\sigma_M^2 = 0.2944 M^{-(\sqrt{3/2})} (D/r_0)^{5/3} \quad (20)$$

where M is the number of modes corrected. Figure 6 shows the required number of Zernike modes to be removed as a function of D/r_0 for three image quality conditions. As an example, if r_0 is 12 cm at

TABLE 1 Mean Square Wavefront Distortion Contributions (in rad^2) from the First 21 Zernike Modes of Atmospheric Aberrations and Residual Error As Each Term Is Corrected

Aberration	Contribution	Distortion
All terms	$1.0299(D/r_0)^{5/3}$	$1.0299(D/r_0)^{5/3}$
Z_2 X-tilt	$0.4479(D/r_0)^{5/3}$	$0.582 (D/r_0)^{5/3}$
Z_3 Y-tilt	$0.4479(D/r_0)^{5/3}$	$0.134 (D/r_0)^{5/3}$
Z_4 Defocus	$0.0232(D/r_0)^{5/3}$	$0.111 (D/r_0)^{5/3}$
Z_5 Astigmatism at 45°	$0.0232(D/r_0)^{5/3}$	$0.0880 (D/r_0)^{5/3}$
Z_6 Astigmatism at 0°	$0.0232(D/r_0)^{5/3}$	$0.0649 (D/r_0)^{5/3}$
Z_7 X-coma	$0.0061(D/r_0)^{5/3}$	$0.0587 (D/r_0)^{5/3}$
Z_8 Y-coma	$0.0061(D/r_0)^{5/3}$	$0.0525 (D/r_0)^{5/3}$
Z_9	$0.0062(D/r_0)^{5/3}$	$0.0463 (D/r_0)^{5/3}$
Z_{10}	$0.0062(D/r_0)^{5/3}$	$0.0401 (D/r_0)^{5/3}$
Z_{11} Spherical aberration	$0.0024(D/r_0)^{5/3}$	$0.0377 (D/r_0)^{5/3}$
Z_{12}	$0.0025(D/r_0)^{5/3}$	$0.0352 (D/r_0)^{5/3}$
Z_{13}	$0.0024(D/r_0)^{5/3}$	$0.0328 (D/r_0)^{5/3}$
Z_{14}	$0.0024(D/r_0)^{5/3}$	$0.0304 (D/r_0)^{5/3}$
Z_{15}	$0.0025(D/r_0)^{5/3}$	$0.0279 (D/r_0)^{5/3}$
Z_{16}	$0.0012(D/r_0)^{5/3}$	$0.0267 (D/r_0)^{5/3}$
Z_{17}	$0.0012(D/r_0)^{5/3}$	$0.0255 (D/r_0)^{5/3}$
Z_{18}	$0.0012(D/r_0)^{5/3}$	$0.0243 (D/r_0)^{5/3}$
Z_{19}	$0.0011(D/r_0)^{5/3}$	$0.0232 (D/r_0)^{5/3}$
Z_{20}	$0.0012(D/r_0)^{5/3}$	$0.0220 (D/r_0)^{5/3}$
Z_{21}	$0.0012(D/r_0)^{5/3}$	$0.0208 (D/r_0)^{5/3}$

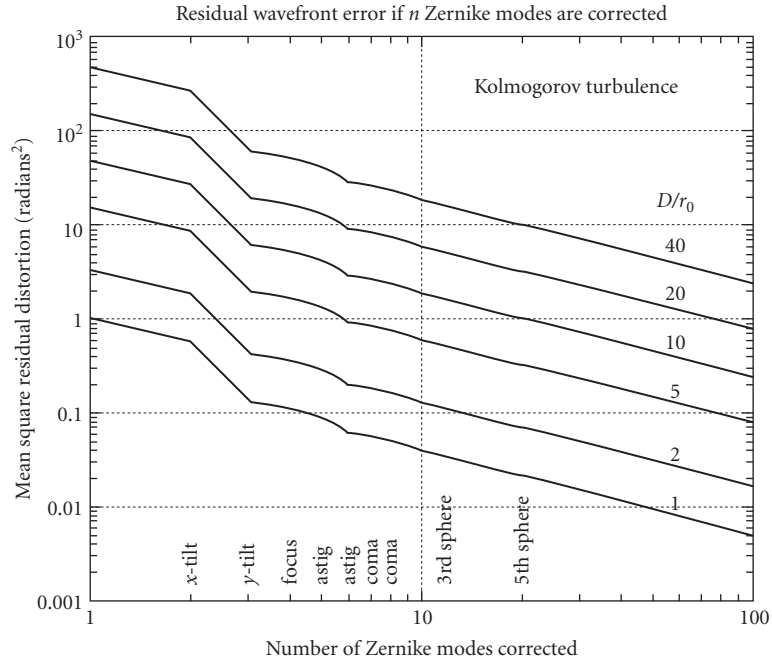


FIGURE 5 Residual mean square phase distortion when n Zernike modes of atmospheric turbulence are corrected for various values of D/r_0 .

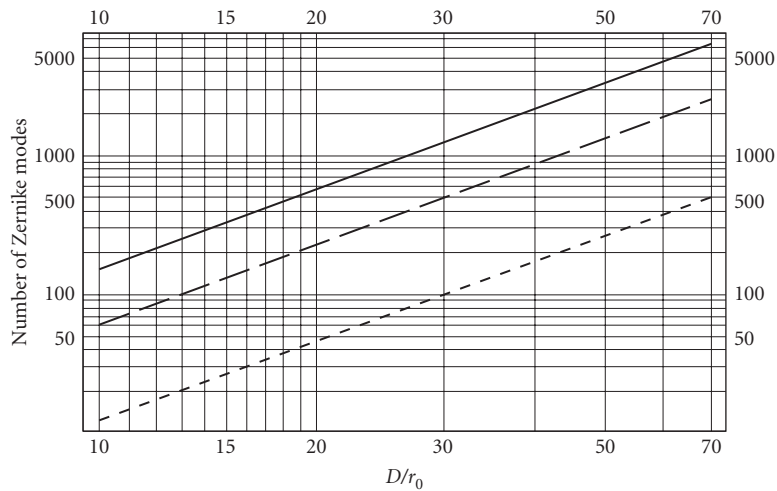


FIGURE 6 Required number of modes to be corrected (roughly the number of actuators or degrees of freedom in the deformable mirror) as a function of D/r_0 . Solid curve: $\lambda/15$ image quality; dashed curve: $\lambda/10$ image quality; dotted curve: $\lambda/5$ image quality. In these curves, r_0 is evaluated at the imaging wavelength.

$0.8\text{ }\mu\text{m}$ and the telescope aperture is 3.5 m , $D/r_0 = 29$, and more than 1000 modes must be removed to achieve $\lambda/15$ image quality, but only 100 modes need be removed if $\lambda/5$ image quality is sufficient. If r_0 is 60 cm at $1.6\text{ }\mu\text{m}$ and $D = 10\text{ m}$, we must remove nearly 400 modes to achieve $\lambda/15$ image quality. For imaging applications, several techniques for postprocessing are well established²⁸ and allow significant enhancements of images that are obtained in real time with AO. Postprocessing may be justification to allow the relaxation of wavefront quality requirements in an AO system. However, for spectroscopy (or other applications requiring high Strehl ratios in real time), there may be little choice but to design the AO system with the required number of degrees of freedom needed to achieve desired Strehl ratios.

The results in Table 1 can be used to estimate the maximum stroke requirement for the actuators in the deformable mirror. Assuming that tilt will be corrected with a dedicated two-axis beam-steering mirror, the root-mean-square (rms) higher-order distortion is (from Table 1) $0.366(D/r_0)^{5/6}$ rad of phase. A practical rule of thumb is that the deformable mirror surface should be able to correct five times the rms wavefront distortion to get 99 percent of the peak values. Since a reflected wavefront has twice the distortion of the mirror's surface (reducing the stroke requirement by a factor of 2), the maximum actuator stroke requirement for operation becomes

$$\sigma_m (\mu\text{m}) = 0.073 \left(\frac{D}{r_0} \right)^{5/6} \quad (21)$$

where σ_m is the maximum actuator stroke in μm , and r_0 is the value of Fried's coherence diameter at a wavelength of $0.5\text{ }\mu\text{m}$ along the line of sight corresponding to the maximum zenith angle of interest. Figure 7 shows the maximum stroke that is required for several aperture diameters as a function of seeing conditions. This figure assumes the outerscale is much larger than the aperture diameter and Kolmogorov turbulence. If the outerscale is of the order of the aperture size or even smaller, the stroke requirements will be reduced.

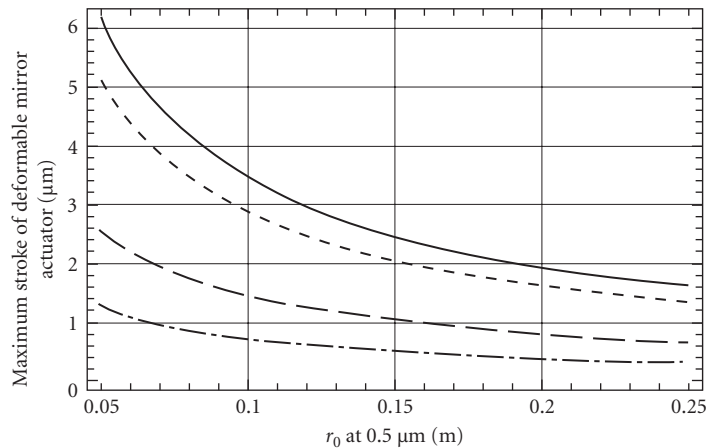


FIGURE 7 Maximum stroke requirement of actuators in the deformable mirror. The curves are for aperture diameters (from top to bottom) of 10, 8, 3.5, and 1.5 m. The value of r_0 is for an observing wavelength of $0.5\text{ }\mu\text{m}$ along the line of sight corresponding to the maximum zenith angle of interest, and the stroke is in micrometers. These curves are independent of wavelength since they do not include any dispersion effects.

Atmospheric Tilt and Its Effect on the Strehl Ratio

Tilt comprises 87 percent of the power in atmospheric turbulence-induced wavefront distortion (see the preceding section). The Strehl ratio of an image that is due to jitter alone is approximated to better than 10 percent by the equation^{29,30}

$$SR_{\text{tilt}} = \frac{1}{1 + \frac{\pi^2}{2} \left(\frac{\sigma_\theta}{\lambda/D} \right)^2} \quad (22)$$

where σ_θ is the one-axis rms angular jitter. Figure 8 shows how the one-axis rms jitter affects system Strehl ratio. Greenwood and Fried³¹ derived bandwidth requirements for AO systems and developed expressions for the variance of full-aperture tilt that is induced by turbulence. The one-axis angular jitter variance (units of angular rad²) is given by

$$\sigma_{\theta G}^2 = 0.170 \left(\frac{\lambda}{D} \right)^2 \left(\frac{D}{r_0} \right)^{5/3} \quad (23)$$

where $\sigma_{\theta G}$ is the rms G -tilt. G -tilt is the average gradient over the wavefront and is well approximated by most centroid trackers. For Z -tilt (the direction of the normal of the best-fit plane to the wavefront distortion), the coefficient in Eq. (23) is 0.184. Consequently, a tracking sensor that measures the centroid of the focused image improperly estimates the Z -tilt; this has been called centroid anisoplanatism.³⁰

Figure 9 shows the dependence of $\sigma_{\theta G}$ on aperture diameter for two seeing conditions, one representative of a continental site and one of a mountaintop island like Mauna Kea, Hawaii.

The effect of atmospheric tilt on the Strehl ratio can be seen by substituting Eq. (23) into Eq. (22). Figure 10 shows that the loss in the Strehl ratio due to jitter alone is significant. For $D/r_0 = 1$, the atmospherically induced jitter limits the Strehl ratio to 0.54. It is most important to effectively control image motion that is created by full-aperture atmospheric tilt.

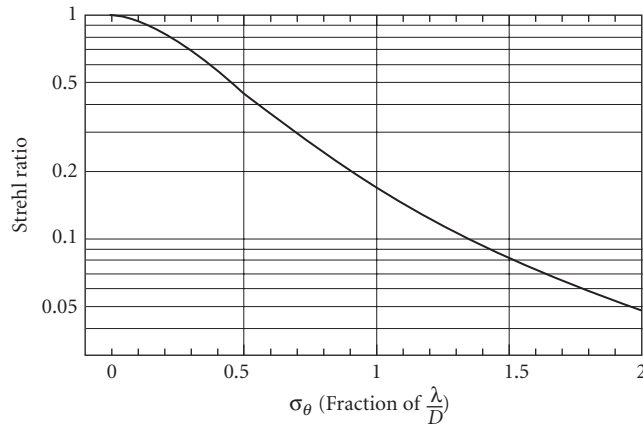


FIGURE 8 Strehl ratio versus one-axis rms jitter expressed as a fraction of λ/D .

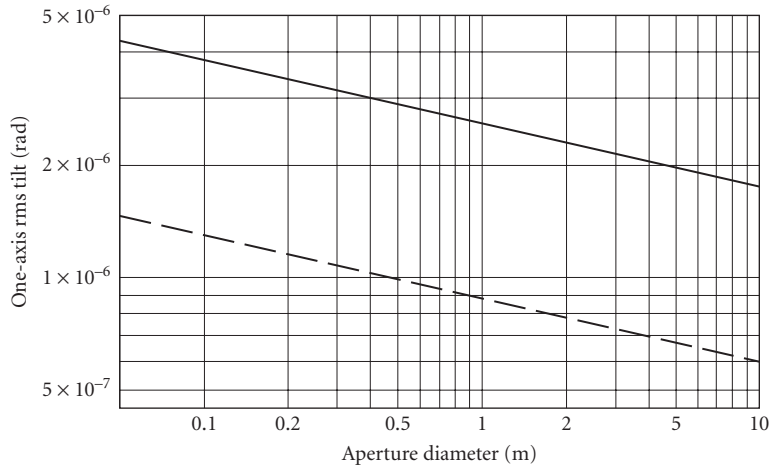


FIGURE 9 One-axis rms tilt for atmospheric turbulence. The upper line corresponds to an r_0 of 5 cm and the lower curve corresponds to an r_0 of 18 cm, both referenced to $0.5 \mu\text{m}$ and zenith. The value of full-aperture tilt is independent of wavelength for a given seeing condition.

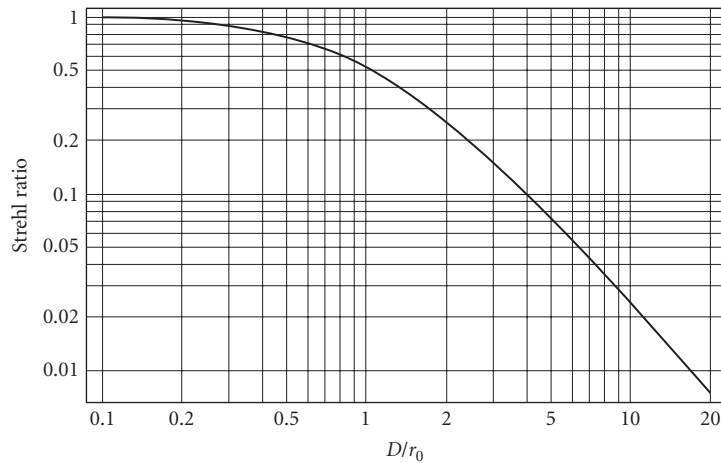


FIGURE 10 The effect of turbulence-induced jitter on system Strehl ratio.

Tracking: Bandwidth and Steering Mirror Stroke Requirements

Tyler³² has considered the problem of required tracking bandwidth by computing the tilt power spectrum and determining the error rejection achieved by a steering mirror under control of an RC-type servo. He considers both G-tilt and Z-tilt. Temporal fluctuations are derived by translating “frozen-turbulence” realizations past the telescope’s aperture. The hypothesis is that the internal structure of the turbulence changes more slowly than its mass transport by the wind (but this assumption may not suffice for today’s giant-aperture telescopes). The results for power spectra are

in integral form, requiring numerical evaluation. Sasiela³³ uses Mellin transforms to derive a solution in terms of a power series of ratios of gamma functions. The asymptotic limits for high and low frequencies are, however, tractable in a form that is easily evaluated. We consider here the results for only G -tilt since that is the most likely implementation in a real system (approximated by a centroid tracker). The low-frequency limit for G -tilt is

$$\lim_{f \rightarrow 0} F_{\phi G}(f) = 0.804 D^{-1/3} \sec \psi f^{-2/3} \int_0^\infty C_n^2(h) [v(h)/D]^{-1/3} dh \quad (24)$$

and the high-frequency limit is

$$\lim_{f \rightarrow \infty} F_{\phi G}(f) = 0.0110 D^{-1/3} \sec \psi f^{-11/3} \int_0^\infty C_n^2(h) [v(h)/D]^{8/3} dh \quad (25)$$

where $F_{\phi G}(f)$ is the one-axis G -tilt power spectrum in rad^2/Hz , D is the aperture diameter, ψ is the zenith angle, and $v(h)$ is the transverse wind component along a vertical profile. Note that the tilt spectrum goes as $f^{-2/3}$ at low frequencies and as $f^{-11/3}$ at high frequencies.

The disturbances described in Eqs. (24) and (25) can be partially corrected with a fast-steering mirror having an RC filter response, $H(f, f_c)$, of the form

$$H(f, f_c) = \frac{1}{1 + i \frac{f}{f_c}} \quad (26)$$

where f_c represents a characteristic frequency [usually the 3 decibel (dB) response]. The uncorrected power in the residual errors can be computed from control theory using the relation

$$\sigma_{\theta G}^2 = \int_0^\infty |1 - H(f, f_c)|^2 F_{\phi G}(f) df \quad (27)$$

where $F_{\phi G}(f)$ is the power spectrum (rad^2/Hz) of G -tilt induced by optical turbulence.

For a steering mirror response function given by Eq. (26), the G -tilt variance is given by

$$\sigma_{\theta G}^2 = \int_0^\infty \frac{(f/f_{3\text{dB}})^2}{1 + (f/f_{3\text{dB}})^2} F_{\phi G}(f) df \quad (28)$$

In the limit of a small (zero) servo bandwidth, this equation represents the tilt that is induced by atmospheric turbulence. Expressing the variance of the jitter in terms of the diffraction angle (λ/D) results in the same expression as Eq. (23).

What does the servo bandwidth need to be? Tyler defines a G -tilt tracking frequency characteristic of the atmospheric turbulence profile and wind velocity as

$$f_{T_G} = 0.331 D^{-1/6} \lambda^{-1} \sec^{1/2} \psi \left[\int_0^\infty C_n^2(h) v^2(h) dh \right]^{1/2} \quad (29)$$

such that the variance of the tilt can be expressed as

$$\sigma_\theta^2 = \left(\frac{f_{T_G}}{f_{3\text{dB}}} \right)^2 \left(\frac{\lambda}{D} \right)^2 \quad (30)$$

Three wind velocity profiles are shown in Fig. 11. The analytical model of Bufton³⁴ is given by

$$v_B(h) = 5 + 37 e^{-(h-12)^2/25} \quad (31)$$

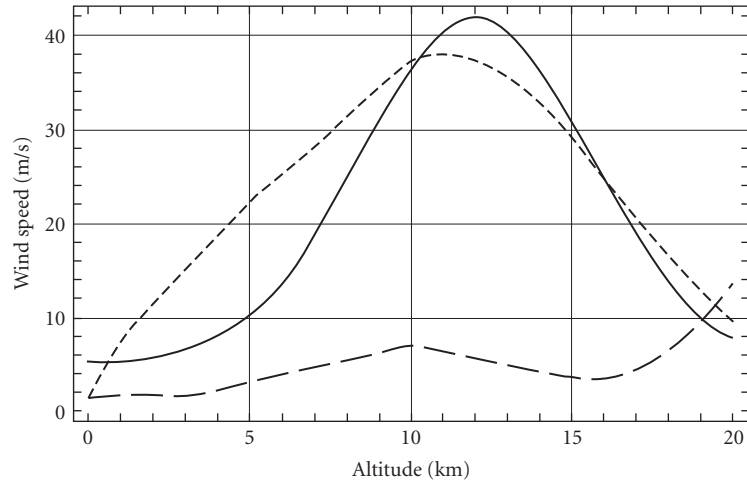


FIGURE 11 Wind profiles. The solid curve is an analytical model by Bufton. The dotted and dashed curves are based on two decades of rawinsonde measurements at Albuquerque, New Mexico, and are known as SOR winter and SOR summer models, respectively. These two models show clearly the absence of a jet stream in the summer months.

where h is the height above ground in km and v_B the wind speed in m/s. The other two curves in Fig. 11 are based on measurements made with National Weather Service sounding balloons launched in Albuquerque, New Mexico. They show a marked contrast between winter and summer months and the effect of the jet stream.

When the 3-dB closed-loop tracking servo frequency equals the G -tilt tracking frequency (the Tyler frequency), the one-axis rms jitter is equal to λ/D . The track bandwidth needs to be four times the Tyler frequency in order for $\sigma_{\theta G}$ to be $1/4(\lambda/D)$, a condition that provides a Strehl ratio due to tilt of greater than 80 percent. Figure 12 shows the required 3-dB tracking bandwidths to

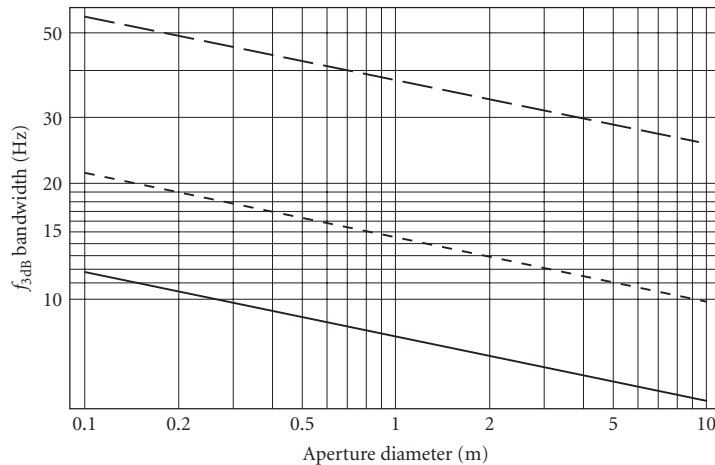


FIGURE 12 Required 3-dB tracking bandwidth to achieve Strehl ratio of 0.82. The dashed curve is for the SOR winter wind and $HV_{5/7}$ profiles; the dotted curve is for the Mauna Kea average C_n^2 profile and Bufton wind profile; the solid curve is for the SOR summer wind and $HV_{5/7} C_n^2$ profiles. All three curves are for $0.5 \mu\text{m}$ wavelength.

meet this criterion for three combinations of wind and C_n^2 profiles as a function of aperture diameter. Other combinations can be computed easily by using Eq. (29) and the wind profiles that are shown in Fig. 11.

The one-axis rms tilt that is induced by the atmosphere is given by Eq. (23). When specifying the maximum excursion of a tilt mirror, the mirror should be able to move five times $\sigma_{\theta G}$ in order to accommodate over 99 percent of the tilt spectrum. We also need to account for the optical magnification between the tilt mirror and the primary mirror of the telescope. If, for instance, the tilt mirror has a diameter of 10 cm, the telescope aperture is 10 m, and r_0 is 18 cm (dashed curve in Fig. 9), the required maximum stroke of the tilt mirror is $0.6 \mu\text{rad}$ (from Fig. 9) times 50 ($5\sigma_{\theta G} \times$ optical magnification of 10) or $30 \mu\text{rad}$.

Higher-Order Phase Fluctuations and Bandwidth Requirements

Time delays that are caused by detector readout and data processing create errors in the phase correction that is applied to the deformable mirror—the correction being applied is relevant to an error that is measured at an earlier time. The control system concepts discussed in the preceding two sections are also relevant to higher-order errors. The residual errors can be computed knowing the closed-loop servo response, $H(f, f_c)$, where f_c represents a characteristic frequency (usually the 3-dB response) and the disturbance. The uncorrected power is similar to Eq. (27) and is given by

$$\sigma_r^2 = \int_0^\infty |1 - H(f, f_c)|^2 F_\phi(f) df \quad (32)$$

where $F_\phi(f)$ is the power spectrum (rad^2/Hz) of the phase distortions that are induced by optical turbulence. At high frequencies, Greenwood³⁵ showed that $F_\phi(f)$ is proportional to $f^{-8/3}$ and is given by

$$\lim_{f \rightarrow \infty} F_\phi(f) = 0.0326 k^2 f^{-8/3} \int_0^\infty C_n^2(z) v^{5/3}(z) dz \quad (33)$$

where $k = 2\pi/\lambda$, f is the frequency in Hertz, and $v(z)$ is the transverse component of the wind profile. The closed-loop response of the AO servo can be modeled as a conventional resistorcapacitor (RC) filter making

$$H(f, f_c) = \frac{1}{1 + i \frac{f}{f_c}} \quad (34)$$

the characteristic frequency becomes

$$f_c = \left[0.102 \left(\frac{k}{\sigma_r} \right)^2 \int_0^\infty C_n^2(z) v^{5/3}(z) dz \right]^{3/5} \quad (35)$$

Figure 11 shows three wind profiles, one analytical and two based on measured data. For the $HV_{5/7} C_n^2$ profile and the SOR summer and winter wind models and $\sigma_r = 0.2\pi$, then $f_c = 23$ and 95 Hz , respectively.

The value of f_c for which $\sigma_r = 1$ is known as the Greenwood frequency and is given explicitly as

$$f_G = \frac{2.31}{\lambda^{6/5}} \left[\int_0^L C_n^2(z) v(z)^{5/3} dz \right]^{3/5} \quad (36)$$

Tyler³⁶ shows that the mean square residual phase in an AO control system having a -3 -dB closed-loop bandwidth of f_{3dB} Hz is

$$\sigma_{\phi_{\text{servo}}}^2 = (f_G / f_{3dB})^{5/3} \quad (37)$$

In a real system, this means that the closed-loop servo bandwidth should be several times the Greenwood frequency in order to “stay up with the turbulence” and keep the residual wavefront error to acceptable levels. Note that the Greenwood frequency scales as $\lambda^{-6/5}$ (the inverse of r_0 and θ_0 scaling) and that zenith angle scaling depends on the specific azimuthal wind direction.

Anisoplanatism

Consider two point sources in the sky (a binary star) and an instrument to measure the wavefront distortion from each. As the angular separation between the sources increases, the wavefront distortions from each source become decorrelated. The *isoplanatic angle* is that angle for which the difference in mean square wavefront distortion is 1 rad². Fried³⁷ defined the isoplanatic angle, θ_0 , as

$$\theta_0 = 0.058 \lambda^{6/5} (\sec \psi)^{-8/5} \left[\int_0^\infty C_n^2(h) h^{5/3} dh \right]^{-3/5} \quad (38)$$

where the integral is along a vertical path through the atmospheric turbulence. Note that θ_0 scales as $\lambda^{6/5}$ (the same as r_0), but as $(\cos \psi)^{8/5}$ with zenith angle. Fried has shown that if a very large diameter ($D/r_0 \gg 1$) imaging system viewing along one path is corrected by an AO system, the optical transfer function of the system viewing along a different path separated by an angle θ is reduced by a factor $\exp[-(\theta/\theta_0)^{5/3}]$. (For smaller diameters, there is not so much reduction.) The numerical value of the isoplanatic angle for zenith viewing at a wavelength of $0.5 \mu\text{m}$ is of the order of a few arc seconds for most sites. The isoplanatic angle is strongly influenced by high-altitude turbulence (note the $h^{5/3}$ weighting in the preceding definition). The small value of the isoplanatic angle can have very significant consequences for AO by limiting the number of natural stars suitable for use as reference beacons and by limiting the corrected field of view to only a few arc seconds.

Turbulence on Imaging and Spectroscopy

The optical transfer function (OTF) is one of the most useful performance measures for the design and analysis of AO imaging systems. The OTF is the Fourier transform of the optical system's point spread function. For an aberration-free circular aperture, the OTF for a spatial frequency f is well known³⁸ to be

$$H_0(f) = \frac{2}{\pi} \left[\arccos \left(\frac{f \bar{\lambda} F}{D} \right) - \frac{f \bar{\lambda} F}{D} \sqrt{1 - \left(\frac{f \bar{\lambda} F}{D} \right)^2} \right] \quad (39)$$

where D is the aperture diameter, F is the focal length of the system, and $\bar{\lambda}$ is the average imaging wavelength. Notice that the cutoff frequency (where the OTF of an aberration-free system reaches 0) is equal to $D/\bar{\lambda}F$.

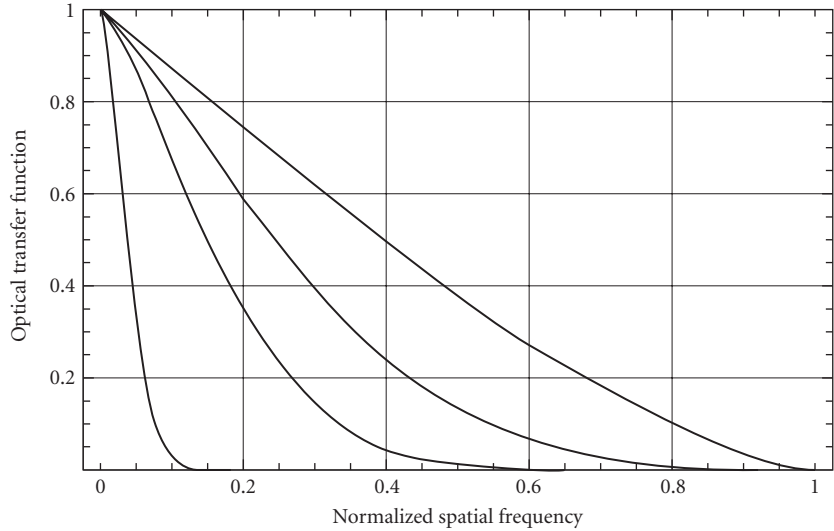


FIGURE 13 The OTF due to the atmosphere. Curves are shown for (top to bottom) $D/r_0 = 0.1, 1, 2$, and 10 . The cutoff frequency is defined when the normalized spatial frequency is 1.0 and has the value $D/\lambda F$, where F is the focal length of the telescope and D is its diameter.

Fried²⁶ showed that for a long-exposure image, the OTF is equal to $H_0(\mathbf{f})$ times the long-exposure OTF due to turbulence, given by

$$H_{LE}(\mathbf{f}) = \exp\left\{-\frac{1}{2}D(\bar{\lambda}F\mathbf{f})\right\} = \exp\left\{-\frac{1}{2}6.88\left(\frac{\bar{\lambda}F\mathbf{f}}{r_0}\right)^{5/3}\right\} \quad (40)$$

where D is the wave structure function and where we have substituted the phase structure function that is given by Eq. (14). Figure 13 shows the OTF along a radial direction for a circular aperture degraded by atmospheric turbulence for values of D/r_0 ranging from 0.1 to 10 . Notice the precipitous drop in the OTF for values of $D/r_0 > 2$. The objective of an AO system is, of course, to restore the high spatial frequencies that are lost due to turbulence.

Spectroscopy is a very important aspect of observational astronomy and is a major contributor to scientific results. The goals of AO for spectroscopy are somewhat different than for imaging. For imaging, it is important to stabilize the corrected point spread function in time and space so that postprocessing can be performed. For spectroscopy, high Strehl ratios are desired in real time. The goal is flux concentration and getting the largest percentage of the power collected by the telescope through the slit of the spectrometer. A 4-m telescope is typically limited to a resolving power of $R \sim 50,000$. Various schemes have been tried to improve resolution, but the instruments become large and complex.

However, by using AO, the corrected image size decreases linearly with aperture size, and very high resolution spectrographs are, in principle, possible without unreasonable-sized gratings. A resolution of $700,000$ was demonstrated on a 1.5-m telescope corrected with AO.³⁹ Tyler and Ellerbroek⁴⁰ have estimated the sky coverage at the galactic pole for the Gemini North 8-m telescope at Mauna Kea as a function of the slit power coupling percentage for a 0.1-arcsec slit width at J, H, and K bands in the near IR. Their results are shown in Fig. 14 for laser guide star (top curves) and natural guide star (lower curves) operation.

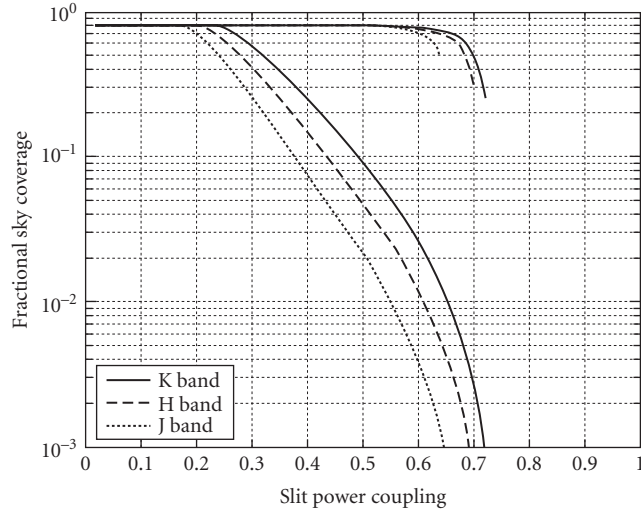


FIGURE 14 Spectrometer slit power sky coverage at Gemini North using NGS (lower curves) and LGS adaptive optics. See text for details.

The higher-order AO system that was analyzed⁴⁰ for the results in Fig. 14 employs a 12-by-12 subaperture Shack-Hartmann sensor for both natural guide star (NGS) and laser guide star (LGS) sensing. The spectral region for NGS is from 0.4 to 0.8 μm and the 0.589- μm LGS is produced by a 15-watt laser to excite mesospheric sodium at an altitude of 90 km. The wavefront sensor charge-coupled device (CCD) array has 3 electrons of read noise per pixel per sample and the deformable mirror is optically conjugate to a 6.5-km range from the telescope. The tracking is done with a 2-by-2-pixel sensor operating at J + H bands with 8 electrons of read noise. Ge et al.⁴¹ have reported similar results with between 40 and 60 percent coupling for LGSs and nearly 70 percent coupling for NGSs brighter than 13th magnitude.

5.5 AO HARDWARE AND SOFTWARE IMPLEMENTATION

Tracking

The wavefront tilt averaged over the full aperture of the telescope accounts for 87 percent of the power in turbulence-induced wavefront aberrations. Full-aperture tilt has the effect of blurring images and reducing the Strehl ratio of point sources. The Strehl ratio due to tilt alone is given by^{29, 30}

$$\text{SR}_{\text{tilt}} = \frac{1}{1 + \frac{\pi^2}{2} \left(\frac{\sigma_\theta}{\lambda/D} \right)^2} \quad (41)$$

where σ_θ is the one-axis rms full-aperture tilt error, $\bar{\lambda}$ is the imaging wavelength, and D is the aperture diameter. Figure 8 is a plot showing how the Strehl ratio drops as jitter increases. This figure

shows that in order to maintain a Strehl ratio of 0.8 due to tilt alone, the image must be stabilized to better than $0.25\lambda/D$. For an 8-m telescope imaging at $1.2\ \mu\text{m}$, $0.25\lambda/D$ is 7.75 milliarcsec.

Fortunately, there are several factors that make tilt sensing more feasible than higher-order sensing for faint guide stars that are available in any field. First, we can use the entire aperture, a light gain over higher-order sensing of roughly $(D/r_0)^2$. Second, the image of the guide star will be compensated well enough that its central core will have a width of approximately λ/D rather than λ/r_0 (assuming that tracking and imaging are near the same wavelength). Third, we can track with only four discrete detectors, making it possible to use photon-counting avalanche photodiodes (or other photon-counting sensors), which have essentially no noise at the short integration times required (~ 10 ms). Fourth, Tyler⁴² has shown that the fundamental frequency that determines the tracking bandwidth is considerably less (by as much as a factor of 9) than the Greenwood frequency, which is appropriate for setting the servo bandwidth of the deformable mirror control system. One must, however, include the vibrational disturbances that are induced into the line of sight by high-frequency jitter in the telescope mount, and it is dangerous to construct a simple rule of thumb comparing tracking bandwidth with higher-order bandwidth requirements.

The rms track error is given approximately by the expression

$$\sigma_\theta = 0.58 \frac{\text{angular image size}}{\text{SNR}_V} \quad (42)$$

where SNR_V is the voltage signal-to-noise ratio in the sensor. An error of $\sigma_\theta = \lambda/4D$ will provide a Strehl ratio of approximately 0.76. If the angular image size is λ/D , the SNR_V needs to be only 2. Since we can count on essentially shot-noise-limited performance, in theory we need only four detected photoelectrons per measurement under ideal conditions (in the real world, we should plan on needing twice this number). Further averaging will occur in the control system servo.

With these assumptions, it is straightforward to compute the required guide star brightness for tracking. Results are shown in Fig. 15 for 1.5-, 3.5-, 8-, and 10-m telescope apertures, assuming a 500-Hz sample rate and twice-diffraction-limited AO compensation of higher-order errors of the guide star at the track wavelength. These results are not inconsistent with high-performance tracking

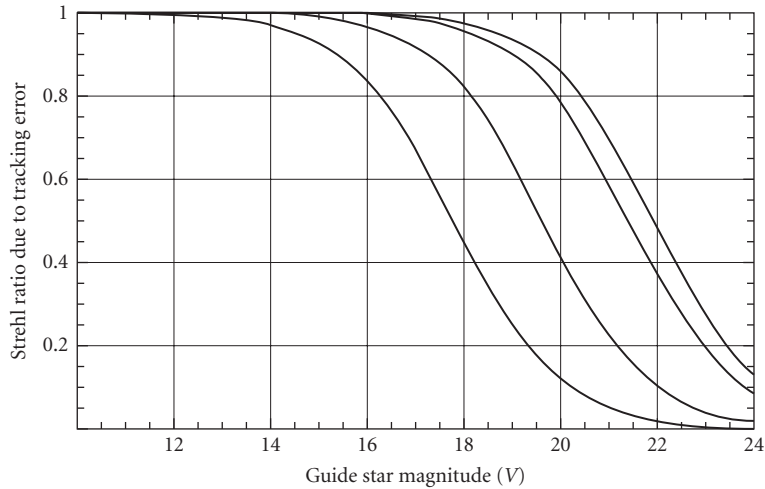


FIGURE 15 Strehl ratio due to tracking jitter. The curves are (top to bottom) for 10-, 8-, 3.5-, and 1.5-m telescopes. The assumptions are photon shot-noise-limited sensors, 25 percent throughput to the track sensor, 200-nm optical bandwidth, and 500-Hz sample rate. The track wavelength is $0.9\ \mu\text{m}$.

systems already in the field, such as the HRCam system that has been so successful on the Canada-France-Hawaii Telescope at Mauna Kea.⁴³

As mentioned previously, the track sensor can consist of just a few detectors implemented in a quadrant-cell algorithm or with a two-dimensional CCD or IR focal plane array. A popular approach for the quad-cell sensor is to use an optical pyramid that splits light in four directions to be detected by individual detectors. Avalanche photodiode modules equipped with photon counter electronics have been used very successfully for this application. These devices operate with such low dark current that they are essentially noise-free and performance is limited by shot noise in the signal and quantum efficiency.

For track sensors using focal plane arrays, different algorithms can be used depending on the tracked object. For unresolved objects, centroid algorithms are generally used. For resolved objects (e.g., planets, moons, asteroids, etc.), correlation algorithms may be more effective. In one instance at the Starfire Optical Range, the highest-quality images of Saturn were made by using an LGS to correct higher-order aberrations and a correlation tracker operating on the rings of the planet provided tilt correction to a small fraction of λ/D .⁴⁴

Higher-Order Wavefront Sensing and Reconstruction: Shack-Hartmann Technique

Higher-order wavefront sensors determine the *gradient*, or slope, of the wavefront measured over subapertures of the entrance pupil, and a dedicated controller maps the slope measurements into deformable mirror actuator voltages. The traditional approach to AO has been to perform these functions in physically different pieces of hardware—the wavefront sensor, the wavefront reconstructor and deformable mirror controller, and the deformable mirror. Over the years, several optical techniques (Shack-Hartmann, various forms of interferometry, curvature sensing, phase diversity, and many others) have been invented for wavefront sensing. A large number of wavefront reconstruction techniques, geometries, and predetermined or even adaptive algorithms have also been developed. A description of all these techniques is beyond the scope of this document, but the interested reader should review work by Wallner,⁴⁵ Fried,⁴⁶ Wild,⁴⁷ and Ellerbroek and Rhoadarmer.⁴⁸

A wavefront sensor configuration in wide use is the Shack-Hartmann sensor.⁴⁹ We will use it here to discuss wavefront sensing and reconstruction principles. Figure 16 illustrates the concept. An array of lenslets is positioned at a relayed image of the exit pupil of the telescope. Each lenslet represents a subaperture—in the ideal case, sized to be less than or equal to r_0 at the sensing wavelength.

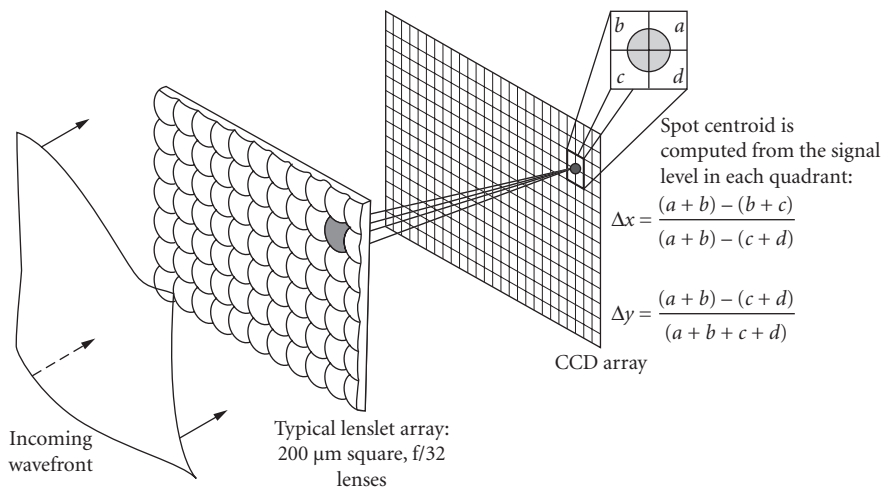


FIGURE 16 Geometry of a Shack-Hartmann sensor.

For subapertures that are roughly r_0 in size, the wavefront that is sampled by the subaperture is essentially flat but tilted, and the objective of the sensor is to measure the value of the subaperture tilt. Light collected by each lenslet is focused to a spot on a two-dimensional detector array. By tracking the position of the focused spot, we can determine the X - and Y -tilt of the wavefront averaged over the subaperture defined by the lenslet. By arranging the centers of the lenslets on a two-dimensional grid, we generate gradient values at points in the centers of the lenslets. Many other geometries are possible resulting in a large variety of patterns of gradient measurements.

Figure 17 is a small-scale example from which we can illustrate the basic equations. In this example, we want to estimate the phase at 16 points on the corners of the subapertures (the Fried geometry of the Shack-Hartmann sensor), and to designate these phases as $\phi_1, \phi_2, \phi_3, \dots, \phi_{16}$, using wavefront gradient measurements that are averaged in the centers of the subapertures, $S_{1x}, S_{1y}, S_{2x}, S_{2y}, \dots, S_{9x}, S_{9y}$. It can be seen by inspection that

$$\begin{aligned} S_{1x} &= \frac{(\phi_2 + \phi_6) - (\phi_1 + \phi_5)}{2d} \\ S_{1y} &= \frac{(\phi_5 + \phi_6) - (\phi_1 + \phi_2)}{2d} \\ &\dots \\ S_{9x} &= \frac{(\phi_{16} + \phi_{12}) - (\phi_{11} + \phi_{15})}{2d} \\ S_{9y} &= \frac{(\phi_{15} + \phi_{16}) - (\phi_{11} + \phi_{12})}{2d} \end{aligned}$$

This system of equations can be written in the form of a matrix as

$$\begin{bmatrix} S_{1x} \\ S_{2x} \\ S_{3x} \\ S_{5x} \\ S_{6x} \\ S_{7x} \\ S_{8x} \\ S_{9x} \\ S_{1y} \\ S_{2y} \\ S_{3y} \\ S_{4y} \\ S_{5y} \\ S_{6y} \\ S_{7y} \\ S_{8y} \\ S_{9y} \end{bmatrix} = \frac{1}{2d} \begin{bmatrix} -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 \\ -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \\ \phi_5 \\ \phi_6 \\ \phi_7 \\ \phi_8 \\ \phi_9 \\ \phi_{10} \\ \phi_{11} \\ \phi_{12} \\ \phi_{13} \\ \phi_{14} \\ \phi_{15} \\ \phi_{16} \end{bmatrix} \quad (43)$$

These equations express gradients in terms of phases,

$$S = H\Phi \quad (44)$$

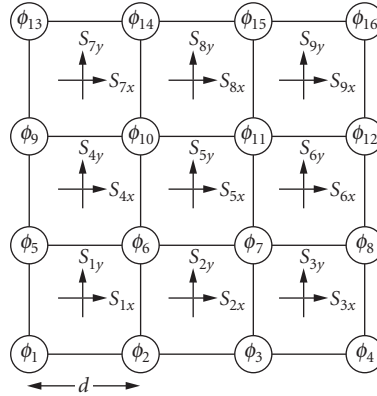


FIGURE 17 A simple Shack-Hartmann sensor in the Fried geometry.

where Φ is a vector of the desired phases, H is the measurement matrix, and S is a vector of the measured slopes.

However, in order to control the actuators in the deformable mirror, we need a control matrix, M , that maps subaperture slope measurements into deformable mirror actuator control commands. In essence, we need to invert Eq. (44) to make it of the form

$$\Phi = MS \quad (45)$$

where M is the desired control matrix. The most straightforward method to derive the control matrix, M , is to minimize the difference in the measured wavefront slopes and the actual slopes on the deformable mirror. We can do this by a maximum a posteriori method accounting for actuator influence functions in the deformable mirror, errors in the wavefront slope measurements due to noise, and statistics of the atmospheric phase distortions. If we do not account for any effects except the geometry of the actuators and the wavefront subapertures, the solution is a least-squares estimate (the most widely implemented to date). It has the form

$$\Phi = [H^T H]^{-1} H^T S \quad (46)$$

and is the *pseudoinverse* of H . (For our simple geometry shown in Fig. 17, the pseudoinverse solution is shown in Fig. 18.) Even this simple form is often problematical since the matrix $[H^T H]$ is often singular or acts as a singular matrix from computational roundoff error and cannot be inverted.

However, in these instances, singular-value-decomposition (SVD) algorithms can be used to directly compute a solution for the inverse of H . Singular value decomposition decomposes an $m \times n$ matrix into the product of an $m \times n$ matrix (U), an $n \times n$ diagonal matrix (D), and an $n \times n$ square matrix (V). So that

$$H = UDV^T \quad (47)$$

and H^{-1} is then

$$H^{-1} = VD^{-1}U^T \quad (48)$$

$-\frac{7}{16}$	$-\frac{9}{80}$	$-\frac{1}{20}$	$\frac{9}{80}$	$-\frac{3}{20}$	$-\frac{1}{80}$	$-\frac{1}{20}$	$\frac{1}{80}$	$-\frac{1}{16}$	$-\frac{7}{16}$	$\frac{9}{80}$	$-\frac{1}{20}$	$-\frac{9}{80}$	$-\frac{3}{20}$	$\frac{1}{80}$	$-\frac{1}{20}$	$-\frac{1}{80}$	$-\frac{1}{16}$
$\frac{11}{60}$	$-\frac{61}{240}$	$-\frac{1}{16}$	$-\frac{29}{240}$	$\frac{1}{20}$	$-\frac{19}{240}$	$\frac{1}{16}$	$-\frac{11}{240}$	$\frac{1}{60}$	$-\frac{11}{60}$	$-\frac{61}{240}$	$\frac{1}{16}$	$-\frac{29}{240}$	$-\frac{1}{20}$	$-\frac{19}{240}$	$-\frac{1}{16}$	$-\frac{11}{240}$	$-\frac{1}{60}$
$\frac{1}{16}$	$\frac{61}{240}$	$-\frac{11}{60}$	$\frac{19}{240}$	$-\frac{1}{20}$	$\frac{29}{240}$	$-\frac{1}{60}$	$\frac{11}{240}$	$-\frac{1}{16}$	$\frac{1}{16}$	$-\frac{61}{240}$	$-\frac{11}{60}$	$-\frac{19}{240}$	$-\frac{1}{20}$	$-\frac{29}{240}$	$-\frac{1}{60}$	$-\frac{11}{240}$	$-\frac{1}{16}$
$\frac{1}{20}$	$\frac{9}{80}$	$\frac{7}{16}$	$\frac{1}{80}$	$\frac{3}{20}$	$-\frac{9}{80}$	$\frac{1}{16}$	$-\frac{1}{80}$	$\frac{1}{20}$	$-\frac{1}{20}$	$\frac{9}{80}$	$-\frac{7}{16}$	$\frac{1}{80}$	$-\frac{3}{20}$	$-\frac{9}{80}$	$-\frac{1}{16}$	$-\frac{1}{80}$	$-\frac{1}{20}$
$-\frac{11}{60}$	$-\frac{29}{240}$	$-\frac{1}{16}$	$-\frac{61}{240}$	$-\frac{1}{20}$	$-\frac{11}{240}$	$\frac{1}{16}$	$-\frac{19}{240}$	$-\frac{1}{60}$	$\frac{11}{60}$	$-\frac{29}{240}$	$\frac{1}{16}$	$-\frac{61}{240}$	$\frac{1}{20}$	$-\frac{11}{240}$	$-\frac{1}{16}$	$-\frac{19}{240}$	$\frac{1}{60}$
$\frac{1}{16}$	$-\frac{9}{80}$	$-\frac{1}{20}$	$\frac{9}{80}$	$-\frac{3}{20}$	$-\frac{1}{80}$	$-\frac{1}{20}$	$\frac{1}{80}$	$-\frac{1}{16}$	$\frac{1}{16}$	$\frac{9}{80}$	$-\frac{1}{20}$	$-\frac{9}{80}$	$-\frac{3}{20}$	$\frac{1}{80}$	$-\frac{1}{20}$	$-\frac{1}{80}$	$-\frac{1}{16}$
$\frac{1}{20}$	$\frac{9}{80}$	$-\frac{1}{16}$	$\frac{1}{80}$	$\frac{3}{20}$	$-\frac{9}{80}$	$\frac{1}{16}$	$-\frac{1}{80}$	$\frac{1}{20}$	$-\frac{1}{20}$	$\frac{9}{80}$	$\frac{1}{16}$	$\frac{1}{80}$	$-\frac{3}{20}$	$-\frac{9}{80}$	$-\frac{1}{16}$	$-\frac{1}{80}$	$-\frac{1}{20}$
$\frac{1}{16}$	$\frac{29}{240}$	$\frac{11}{60}$	$\frac{11}{240}$	$\frac{1}{20}$	$\frac{61}{240}$	$\frac{1}{60}$	$\frac{19}{240}$	$-\frac{1}{16}$	$\frac{1}{16}$	$-\frac{29}{240}$	$\frac{11}{60}$	$-\frac{11}{240}$	$\frac{1}{20}$	$-\frac{61}{240}$	$\frac{1}{60}$	$-\frac{19}{240}$	$-\frac{1}{16}$
$\frac{1}{16}$	$-\frac{19}{240}$	$-\frac{1}{60}$	$-\frac{61}{240}$	$-\frac{1}{20}$	$-\frac{11}{240}$	$-\frac{11}{60}$	$-\frac{29}{240}$	$-\frac{1}{16}$	$\frac{1}{16}$	$\frac{19}{240}$	$-\frac{1}{60}$	$\frac{61}{240}$	$-\frac{1}{20}$	$\frac{11}{240}$	$-\frac{11}{60}$	$\frac{29}{240}$	$-\frac{1}{16}$
$-\frac{1}{20}$	$\frac{1}{80}$	$-\frac{1}{16}$	$\frac{9}{80}$	$-\frac{3}{20}$	$-\frac{1}{80}$	$\frac{1}{16}$	$-\frac{9}{80}$	$-\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{80}$	$\frac{1}{16}$	$\frac{9}{80}$	$\frac{3}{20}$	$-\frac{1}{80}$	$-\frac{1}{16}$	$-\frac{9}{80}$	$\frac{1}{20}$
$\frac{1}{16}$	$-\frac{1}{80}$	$\frac{1}{20}$	$\frac{1}{80}$	$\frac{3}{20}$	$-\frac{9}{80}$	$\frac{1}{20}$	$\frac{9}{80}$	$-\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{80}$	$\frac{1}{20}$	$-\frac{1}{80}$	$\frac{3}{20}$	$\frac{9}{80}$	$\frac{1}{20}$	$-\frac{9}{80}$	$-\frac{1}{16}$
$\frac{1}{60}$	$\frac{19}{240}$	$-\frac{1}{16}$	$\frac{11}{240}$	$\frac{1}{20}$	$\frac{61}{240}$	$\frac{1}{16}$	$\frac{29}{240}$	$\frac{11}{60}$	$-\frac{1}{60}$	$\frac{19}{240}$	$\frac{1}{16}$	$\frac{11}{240}$	$-\frac{1}{20}$	$\frac{61}{240}$	$-\frac{1}{16}$	$\frac{29}{240}$	$-\frac{11}{60}$
$-\frac{1}{20}$	$\frac{1}{80}$	$-\frac{1}{16}$	$\frac{9}{80}$	$-\frac{3}{20}$	$-\frac{1}{80}$	$-\frac{7}{16}$	$-\frac{9}{80}$	$-\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{80}$	$\frac{1}{16}$	$\frac{9}{80}$	$\frac{3}{20}$	$-\frac{1}{80}$	$\frac{7}{16}$	$-\frac{9}{80}$	$\frac{1}{20}$
$\frac{1}{16}$	$-\frac{11}{240}$	$\frac{1}{60}$	$-\frac{29}{240}$	$\frac{1}{20}$	$-\frac{19}{240}$	$\frac{11}{60}$	$-\frac{61}{240}$	$-\frac{1}{16}$	$\frac{1}{16}$	$\frac{11}{240}$	$\frac{1}{60}$	$\frac{29}{240}$	$\frac{1}{20}$	$\frac{19}{240}$	$\frac{11}{60}$	$\frac{61}{240}$	$-\frac{1}{16}$
$\frac{1}{60}$	$\frac{11}{240}$	$\frac{1}{16}$	$\frac{19}{240}$	$\frac{1}{20}$	$\frac{29}{240}$	$\frac{1}{16}$	$\frac{61}{240}$	$-\frac{11}{60}$	$\frac{1}{60}$	$\frac{11}{240}$	$\frac{1}{16}$	$\frac{19}{240}$	$\frac{1}{20}$	$\frac{29}{240}$	$-\frac{1}{16}$	$\frac{61}{240}$	$\frac{11}{60}$
1	$-\frac{1}{80}$	$\frac{1}{20}$	$\frac{1}{80}$	$\frac{3}{20}$	$-\frac{9}{80}$	$\frac{1}{20}$	$\frac{9}{80}$	$\frac{7}{16}$	$\frac{1}{16}$	$\frac{1}{80}$	$\frac{1}{20}$	$-\frac{1}{80}$	$\frac{3}{20}$	$\frac{9}{80}$	$\frac{1}{20}$	$-\frac{9}{80}$	$\frac{7}{16}$

FIGURE 18 Least-squares reconstructor matrix for the geometry shown in Fig. 17.

If H is singular, some of the diagonal elements of D will be zero and D^{-1} cannot be defined. However, this method allows us to obtain the closest possible solution in a least-squares sense by zeroing the elements in the diagonal of D^{-1} that come from zero elements in the matrix D . We can arrive at a solution that discards only those equations that generated the problem in the first place. In addition to straightforward SVD, more general techniques have been proposed involving iterative solutions for the phases.^{46, 50, 51}

In addition, several other “tricks” have been developed to alleviate the singularity problem. For instance, piston error is normally ignored, and this contributes to the singularity since there are then an infinite number of solutions that could give the same slope measurements. Adding a row of 1s to the measurement matrix H and setting a corresponding value of 0 in the slope vector for piston allows inversion.⁵²

In the implementation of Shack-Hartmann sensors, a reference wavefront must be provided to calibrate imperfections in the lenslet array and distortions that are introduced by any relay optics to match the pitch of the lenslet array to the pitch of the pixels in the detector array. It is general practice to inject a “perfect” plane wave into the optical train just in front of the lenslet array and to record the positions of all of the Shack-Hartmann spots. During normal operation, the wavefront sensor gradient processor then computes the difference between the spot position of the residual error wavefront and the reference wavefront.

Laser Beacons (Laser Guide Stars)

Adaptive optical systems require a beacon or a source of light to sense turbulence-induced wave distortions. From an anisoplanatic point of view, the ideal beacon is the object being imaged. However, most objects of interest to astronomers are not bright enough to serve as beacons.

It is possible to create artificial beacons (also referred to as synthetic beacons) that are suitable for wavefront sensing with lasers as first demonstrated by Fugate et al.⁵³ and Primmerman et al.⁵⁴ Laser beacons can be created by Rayleigh scattering of focused beams at ranges between 15 and 20 km or by resonant scattering from a layer of sodium atoms in the mesosphere at an altitude of between 90 and 100 km. Examples of Rayleigh beacon AO systems are described by Fugate et al.⁵⁵ for the SOR 1.5-m telescope and by Thompson et al.⁵⁶ for the Mt. Wilson 100-in telescope; examples of sodium beacon AO systems are described by Olivier et al.⁵⁷ for the Lick 3-m Shane telescope and by Butler et al.⁵⁸ for the 3.5-m Calar Alto telescope. Researchers at the W. M. Keck Observatory at Mauna Kea are in the process of installing a sodium dye laser to augment their NGS AO system on Keck II.

The laser beacon concept was first conceived and demonstrated within the U.S. Department of Defense during the early 1980s (see Fugate⁵⁹ for a short summary of the history). The information developed under this program was not declassified until May 1991, but much of the early work was published subsequently.⁶⁰ The laser beacon concept was first published openly by Foy and Labeyrie⁶¹ in 1985 and has been of interest in the astronomy community since.

Even though laser beacons solve a significant problem, they also introduce new problems and have two significant limitations compared with bright NGSS. The new problems include potential light contamination in science cameras and tracking sensors, cost of ownership and operation, and observing complexities that are associated with propagating lasers through the navigable airspace and near-earth orbital space, the home of thousands of space payloads. The technical limitations are that laser beacons provide no information on full-aperture tilt, and that a “cone effect” results from the finite altitude of the beacon, which contributes an additional error called focus (or focal) anisoplanatism. Focus anisoplanatism can be partially alleviated by using a higher-altitude beacon, such as a sodium guide star, as discussed in the following.

Focus Anisoplanatism The mean square wavefront error due to the finite altitude of the laser beacon is given by

$$\sigma_{FA}^2 = (D/d_0)^{5/3} \quad (49)$$

where d_0 is an effective aperture size corrected by the laser beacon that depends on the height of the laser beacon, the C_n^2 profile, the zenith angle, and the imaging wavelength. This parameter was defined by Fried and Belsher.⁶² Tyler⁶³ developed a method to rapidly evaluate d_0 for arbitrary C_n^2 profiles given by the expression

$$d_0 = \lambda^{6/5} \cos^{3/5}(\psi) \left[\int C_n^2(z) F(z/H) dz \right]^{-3/5} \quad (50)$$

where H is the vertical height of the laser beacon and the function $F(z/H)$ is given by

$$\begin{aligned}
 F(z/H) = & 16.71371210(1.032421640 - 0.8977579487u) \\
 & \times [1 + (1 - z/H)^{5/3}] - 2.168285442 \\
 & \times \left\{ \frac{6}{11} {}_2F_1 \left[-\frac{11}{6}, -\frac{5}{6}; 2; \left(1 - \frac{z}{H}\right)^2 \right] \right. \\
 & \left. - \frac{6}{11} (z/H)^{5/3} - u \frac{10}{11} \left(1 - \frac{z}{H}\right) \right. \\
 & \left. \times {}_2F_1 \left[-\frac{11}{6}, \frac{1}{6}; 3; \left(1 - \frac{z}{H}\right)^2 \right] \right\} \quad (51)
 \end{aligned}$$

for $z < H$ and

$$F(z/H) = 16.71371210(1.032421640 - 0.8977579487u) \quad (52)$$

for $z > H$. In these equations, z is the vertical height above the ground, H is the height of the laser beacon, u is a parameter that is equal to zero when only piston is removed and that is equal to unity when piston and tilt are removed, and ${}_2F_1[a, b; c; z]$ is the hypergeometric function.

Equations (50) and (51) are easily evaluated on a programmable calculator or a personal computer. They are very useful for quickly establishing the expected performance of laser beacons for a particular C_n^2 profile, imaging wavelength, and zenith angle view. Figures 19 and 20 are plots of d_0 representing the best and average seeing at Mauna Kea and a site described by the HV_{57} turbulence profile. Since d_0 scales as $\lambda^{6/5}$, values at $2.2 \mu\text{m}$ are 5.9 times larger, as shown in the plots.

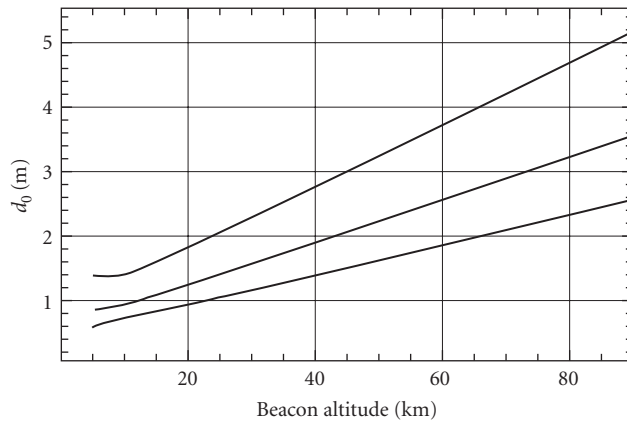


FIGURE 19 Values of d_0 versus laser beacon altitude for zenith imaging at $0.5 \mu\text{m}$ and (top to bottom) best Mauna Kea seeing, average Mauna Kea seeing, and the HV_{57} turbulence profile.

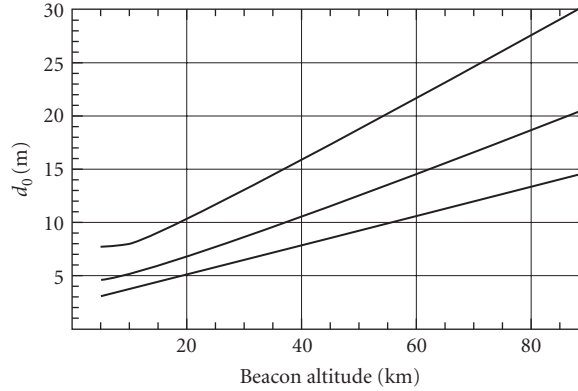


FIGURE 20 Values of d_0 versus laser beacon altitude for zenith imaging at $2.2\ \mu\text{m}$ and (top to bottom) best Mauna Kea seeing, average Mauna Kea seeing, and the $HV_{5/7}$ turbulence profile.

It is straightforward to compute the Strehl ratio due only to focus anisoplanatism using the approximation $SR = e^{-\sigma_{FA}^2}$, where $\sigma_{FA}^2 = (D/d_0)^{5/3}$. There are many possible combinations of aperture sizes, imaging wavelength, and zenith angle, but to illustrate the effect for 3.5- and 10-m apertures, Figs. 21 and 22 show the focus anisoplanatism Strehl ratio as a function of wavelength for 15- and 90-km beacon altitudes and for three seeing conditions. As these plots show, the effectiveness of laser beacons is very sensitive to the aperture diameter, seeing conditions, and beacon altitude. A single Rayleigh beacon at an altitude of 15 km is essentially useless on a 10-m aperture in $HV_{5/7}$ seeing, but it is probably useful at a Mauna Kea—like site. One needs to keep in mind that the curves in Figs. 21 and 22 are the upper limits of performance since other effects will further reduce the Strehl ratio.

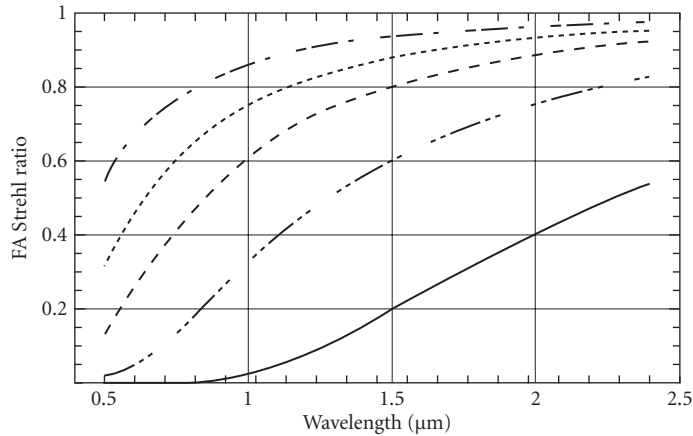


FIGURE 21 The telescope Strehl ratio due to focus anisoplanatism only. Conditions are for a 3.5-m telescope viewing at 30° zenith angle. Curves are (top to bottom): best seeing at Mauna Kea, 90-km beacon; average seeing at Mauna Kea, 90-km beacon; $HV_{5/7}$, 90-km beacon; best seeing at Mauna Kea, 15-km beacon; and $HV_{5/7}$, 15-km beacon.

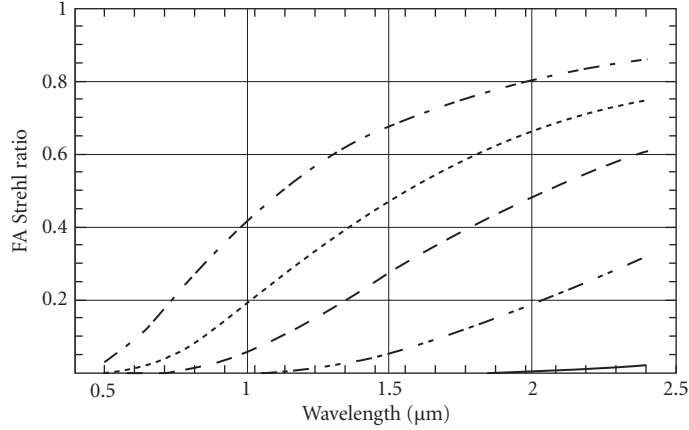


FIGURE 22 The telescope Strehl ratio due to focus anisoplanatism only. Conditions are for a 10.0-m telescope viewing at 30° zenith angle. Curves are (top to bottom): best seeing at Mauna Kea, 90-km beacon; average seeing at Mauna Kea, 90-km beacon; HV_{5/7}, 90-km beacon; best seeing at Mauna Kea, 15-km beacon; and HV_{5/7}, 15-km beacon.

Generation of Rayleigh Laser Beacons For Rayleigh scattering, the number of photo-detected electrons (pdes) per subaperture in the wavefront sensor, N_{pde} , can be computed by using a lidar equation of the form

$$N_{\text{pde}} = \eta_{\text{QE}} T_t T_r T_{\text{atm}}^2 \frac{A_{\text{sub}} \beta_{\text{BS}} \Delta l}{R^2} \frac{E_p \lambda}{hc} \quad (53)$$

where η_{QE} = quantum efficiency of the wavefront sensor

T_t = laser transmitter optical transmission

T_r = optical transmission of the wavefront sensor

T_{atm} = one-way transmission of the atmosphere

A_{sub} = area of a wavefront sensor subaperture

β_{BS} = fraction of incident laser photons backscattered per meter of scattering volume [steradian (sr)⁻¹m⁻¹]

R = range to the midpoint of the scattering volume

Δl = length of the scattering volume—the range gate

E_p = energy per pulse

λ = laser wavelength

h = Planck's constant

c = speed of light

The laser beam is focused at range R , and the wavefront sensor is gated on and off to exclude back-scattered photons from the beam outside a range gate of length Δl centered on range R . The volume-scattering coefficient is proportional to the atmospheric pressure and is inversely proportional to the temperature and the fourth power of the wavelength. Penndorf⁶⁴ developed the details of this relationship, which can be reduced to

$$\beta_{\text{BS}} = 4.26 \times 10^{-7} \frac{P(h)}{T(h)} \text{sr}^{-1} \text{m}^{-1} \quad (54)$$

where values of number density, pressure, and temperature at sea level (needed in Penndorf's equations) have been obtained from the U.S. Standard Atmosphere.⁶⁵ At an altitude of 10 km, $\beta_{BS} = 5.1 \times 10^{-7} \text{ sr}^{-1} \text{ m}^{-1}$ and a 1-km-long volume of the laser beam scatters only 0.05 percent of the incident photons per steradian. Increasing the length of the range gate would increase the total signal that is received by the wavefront sensor; however, we should limit the range gate so that subapertures at the edge of the telescope are not able to resolve the projected length of the scattered light. Range gates that are longer than this criterion increase the size of the beacon's image in a subaperture and increase the rms value of the measurement error in each subaperture. A simple geometric analysis leads to an expression for the maximum range gate length:

$$\Delta L = 2 \frac{\lambda R_b^2}{D r_0} \quad (55)$$

where R_b is the range to the center of the beacon and D is the aperture diameter of the telescope. Figure 23 shows the computed signal in detected electrons per subaperture as a function of altitude for a 100-watt (W) average power laser operating at 1000 pulses per second at either 351 nm (upper curve) or 532 nm (lower curve). Other parameters used to compute these curves are listed in the figure caption. When all first-order effects are accounted for, notice that (for high-altitude beacons) there is little benefit to using an ultraviolet wavelength laser over a green laser—even though the scattering goes as λ^{-4} . With modern low-noise detector arrays, signal levels as low as 50 pdes are very usable; above 200 pdes, the sensor generally no longer dominates the performance.

Equations (53) and (55) accurately predict what has been realized in practice. A pulsed copper-vapor laser, having an effective pulse energy of 180 millijoules (mJ) per wavefront sample, produced 190 pdes per subaperture on the Starfire Optical Range 1.5-m telescope at a backscatter range of 10 km, range gate of 2.4 km, and subapertures of 9.2 cm. The total round-trip optical and quantum efficiency was only 0.4 percent.⁵⁵

There are many practical matters on how to design the optical system to project the laser beam. If the beam shares any part of the optical train of the telescope, then it is important to inject the beam

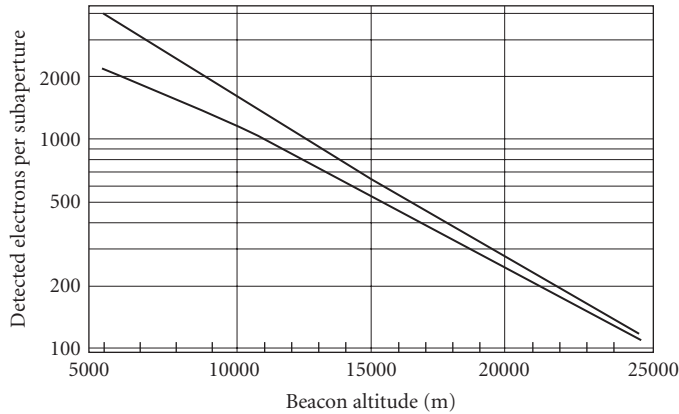


FIGURE 23 Rayleigh laser beacon signal versus altitude of the range gate for two laser wavelengths: 351 nm (top curve) and 532 nm (bottom curve). Assumptions for each curve: range gate length $= 2\lambda R_b^2 / (D r_0)$, $D = 1.5$ m, $r_0 = 5$ cm at 532 nm and 3.3 cm at 351 nm, $T_i = 0.30$, $T_r = 0.25$, $\eta_{QE} = 90$ percent at 532 nm and 70 percent at 351 nm, $E_p = 0.10$ J, subaperture size = 10 cm², one-way atmospheric transmission computed as a function of altitude and wavelength.

as close to the output of the telescope as feasible. The best arrangement is to temporally share the aperture with a component that is designed to be out of the beam train when sensors are using the telescope (e.g., a rotating reflecting wheel as described by Thompson et al.⁵⁶). If the laser is injected optically close to wavefront sensors or trackers, there is the additional potential of phosphorescence in mirror substrates and coatings that can emit photons over a continuum of wavelengths longer than the laser fundamental. The decay time of these processes can be longer than the interpulse interval of the laser, presenting a pseudo-continuous background signal that can interfere with faint objects of interest. This problem was fundamental in limiting the magnitude of natural guide star tracking with photon counting avalanche photo diodes at the SOR 1.5-m telescope during operations with the copper-vapor laser.⁶⁶

Generation of Mesospheric Sodium Laser Beacons As the curves in Figs. 21 and 22 show, performance is enhanced considerably when the beacon is at high altitude (90 versus 15 km). The signal from Rayleigh scattering is 50,000 times weaker for a beacon at 90 km compared with one at 15 km. Happer et al.⁶⁷ suggested laser excitation of mesospheric sodium in 1982; however, the generation of a beacon that is suitable for use with AO has turned out to be a very challenging problem. The physics of laser excitation is complex and the development of an optimum laser stresses modern materials science and engineering. This section only addresses how the temporal and spectral format of the laser affects the signal return. The engineering issues of building a laser are beyond the present scope.

The sodium layer in the mesosphere is believed to arise from meteor ablation. The average height of the layer is approximately 95 km above sea level and it is 10 km thick. The column density is only $2\text{--}5 \cdot 10^9$ atoms/cm², or roughly 10^3 atoms/cm³. The temperature of the layer is approximately 200 K, resulting in a Doppler broadened absorption profile having a full width half maximum (FWHM) of about 3 gigahertz (GHz), which is split into two broad resonance peaks that are separated by 1772 MHz, caused by splitting of the $3S_{1/2}$ ground state. The natural lifetime of an excited state is only 16 ns. At high laser intensities (roughly 6 mW/cm² for lasers with a natural line width of 10 MHz or 5 W/cm² for lasers having spectral content that covers the entire Doppler broadened spectrum), saturation occurs and the return signal does not increase linearly with increasing laser power.

The three types of lasers that have been used to date for generating beacons are the continuous-wave (CW) dye laser, the pulsed-dye laser, and a solid-state sum-frequency laser. Continuous-wave dye lasers are available commercially and generally provide from 2 to 4 W of power. A specialized pulsed-dye laser installed at Lick Observatory's 3-m Shane Telescope, built at Lawrence Livermore National Laboratory by Friedman et al.,⁶⁸ produces an average power of 20 W consisting of 100- to 150-ns-long pulses at an 11-kHz pulse rate. The sum-frequency laser concept relies on the sum-frequency mixing of the 1.064- and 1.319- μ m lines of the Nd:YAG laser in a nonlinear crystal to produce the required 0.589- μ m wavelength for the spectroscopic D_2 line. The first experimental devices were built by Jeys at MIT Lincoln Laboratory.^{69,70} The pulse format is one of an envelope of macropulses, lasting of the order of 100 μ s, containing mode-locked pulses at roughly a 100-MHz repetition rate, and a duration of from 400 to 700 ps. The sum-frequency lasers built to date have had average powers of from 8 to 20 W and have been used in field experiments at SOR and Apache Point Observatory.^{71,72}

A comprehensive study of the physics governing the signal that is generated by laser excitation of mesospheric sodium has been presented by Milonni et al.^{73,74} for short, intermediate, and long pulses, and CW, corresponding to the lasers described previously. Results are obtained by numerical computations and involve a full-density matrix treatment of the sodium D_2 line. In some specific cases, it has been possible to approximate the results with analytical models that can be used to make rough estimates of signal strengths.

Figure 24 shows results of computations by Milonni et al. for short- and long-pulse formats and an analytical extension of numerical results for a high-power CW laser. Results are presented as the number of photons per square centimeter per millisecond received at the primary mirror of the telescope versus average power of the laser. The curves correspond to (top to bottom) the sum-frequency laser; a CW laser whose total power is spread over six narrow lines that are distributed across the Doppler profile; a CW laser having only one narrow line, and the pulsed-dye laser. The specifics are given in the caption for Fig. 24. Figure 25 extends these results to 200-W average power and shows

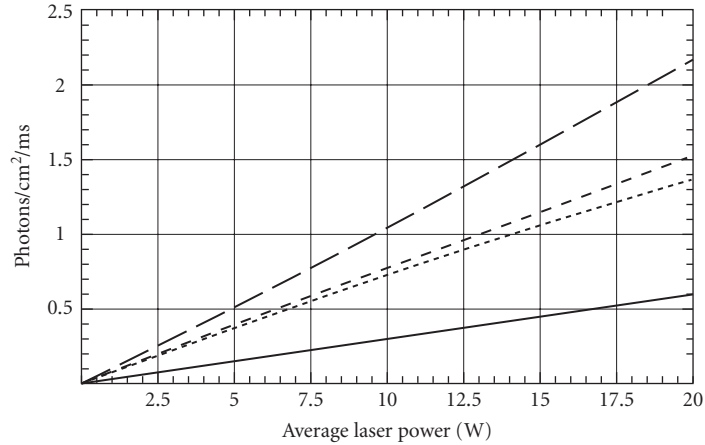


FIGURE 24 Sodium laser beacon signal versus average power of the pump laser. The lasers are (top to bottom) sum-frequency laser with a micropulse-macropulse format (150- μ s macropulses filled with 700-ps micropulses at 100 MHz), a CW laser having six 10-MHz-linewidth lines, a CW laser having a single 10-MHz linewidth, and a pulsed-dye laser having 150-ns pulses at a 30-kHz repetition rate. The sodium layer is assumed to be 90 km from the telescope, the atmospheric transmission is 0.7, and the spot size in the mesosphere is 1.2 arcsec.

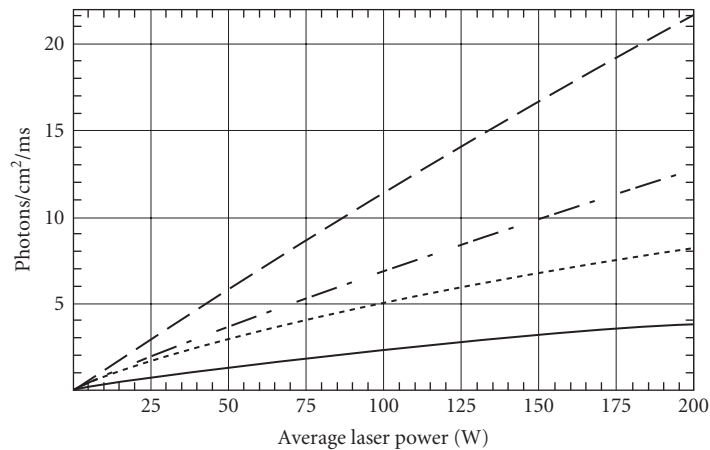


FIGURE 25 Sodium laser beacon signal versus average power of the pump laser. The lasers are (top to bottom) sum-frequency laser with a micropulse-macropulse format (150- μ s macropulses filled with 700-ps micropulses at 100 MHz), a CW laser having six 10-MHz-linewidth lines, a CW laser having a single 10-MHz linewidth, and a pulsed-dye laser having 150-ns pulses at a 30-kHz repetition rate. The sodium layer is assumed to be 90 km from the telescope, the atmospheric transmission is 0.7, and the spot size in the mesosphere is 1.2 arcsec. Saturation of the return signal is very evident for the temporal format of the pulsed-dye laser and for the single-line CW laser. Note, however, that the saturation is mostly eliminated (the line is nearly straight) in the single-line laser by spreading the power over six lines. There appears to be very little saturation in the micropulse-macropulse format.

significant saturation for the pulsed-dye laser format and the single-frequency CW laser as well as some nonlinear behavior for the sum-frequency laser. If the solid-state laser technology community can support high-power, narrow-line CW lasers, this chart says that saturation effects at high power can be ameliorated by spreading the power over different velocity classes of the Doppler profile. These curves can be used to do first-order estimates of signals that are available from laser beacons that are generated in mesospheric sodium with lasers having these temporal formats.

Real-Time Processors

The mathematical process described by Eq. (45) is usually implemented in a dedicated processor that is optimized to perform the matrix multiplication operation. Other wavefront reconstruction approaches that are not implemented by matrix multiplication routines are also possible, but they are not discussed here.

Matrix multiplication lends itself to parallel operations and the ultimate design to maximize speed is to dedicate a processor to each actuator in the deformable mirror. That is, each central processing unit (CPU) is responsible for multiplying and accumulating the sum of each element of a row in the M matrix with each element of the slope vector S . The values of the slope vector should be broadcast to all processors simultaneously to reap the greatest benefit. Data flow and throughput is generally a more difficult problem than raw computing power. It is possible to buy very powerful commercial off-the-shelf processing engines, but getting the data into and out of the engines usually reduces their ultimate performance. A custom design is required to take full advantage of component technology that is available today.

An example of a custom-designed system is the SOR 3.5-m telescope AO system⁷⁵ containing 1024 digital signal processors running at 20 MHz, making a 20-billion-operations-per-second system. This system can perform a $(2048 \times 1024) \times (2048)$ matrix multiply to 16-bit precision (40-bit accumulation), low-pass-filter the data, and provide diagnostic data collection in less than 24 μs . The system throughput exceeds 400 megabytes per second (MB/s). Most astronomy applications have less demanding latency and throughput requirements that can be met with commercial off-the-shelf hardware and software. The importance of latency is illustrated in Fig. 26. This figure shows results

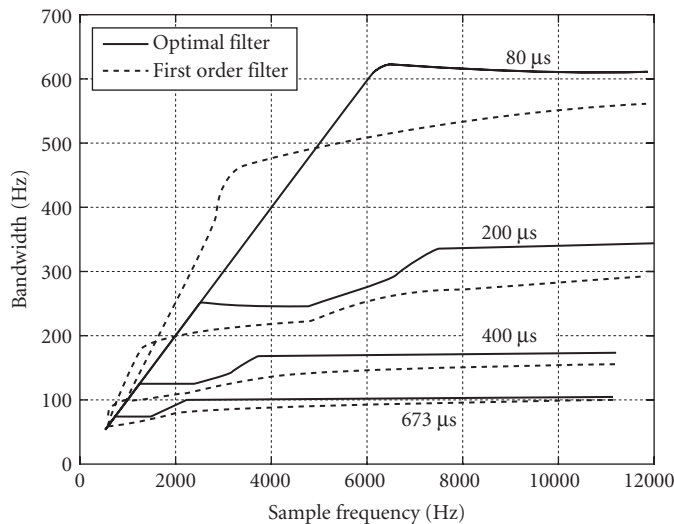


FIGURE 26 Effect of data latency (varying from 80 to 673 μs in this plot) on the control loop bandwidth of the AQ system. In this plot, the latency includes only the sensor readout and waverfront processing time.

of analytical predictions showing the relationship between control loop bandwidth of the AO system and the wavefront sensor frame rate for different data latencies.⁷⁶ These curves show that having a high-frame-rate wavefront sensor camera is not a sufficient condition to achieve high control loop bandwidth. The age of the data is of utmost importance. As Fig. 26 shows, the optimum benefit in control bandwidth occurs when the latency is significantly less than a frame time ($\sim 1/2$).

Ensemble-averaged atmospheric turbulence conditions are very dynamic and change on the scale of minutes. Optimum performance of an AO system cannot be achieved if the system is operating on phase estimation algorithms that are based on inaccurate atmospheric information. Optimum performance requires changing the modes of operation in near-real time. One of the first implementations of adaptive control was the system ADONIS, which was implemented on the ESO 3.6-m telescope at La Silla, Chile.⁷⁷ ADONIS employs an artificial intelligence control system that controls which spatial modes are applied to the deformable mirror, depending on the brightness of the AO beacon and the seeing conditions.

A more complex technique has been proposed⁴⁸ that would allow continuous updating of the wavefront estimation algorithm. The concept is to use a recursive least-squares adaptive algorithm to track the temporal and spatial correlations of the distorted wavefronts. The algorithm uses current and recent past information that is available to the servo system to predict the wavefront for a short time in the future and to make the appropriate adjustments to the deformable mirror. A sample scenario has been examined in a detailed simulation, and the system Strehl ratio achieved with the recursive least-squares adaptive algorithm is essentially the same as an optimal reconstructor with *a priori* knowledge of the wind and turbulence profiles. Sample results of this simulation are shown in Fig. 27. The requirements for implementation of this algorithm in a real-time hardware processor have not been worked out in detail. However, it is clear that they are considerable, perhaps greater by an order of magnitude than the requirements for an ordinary AO system.

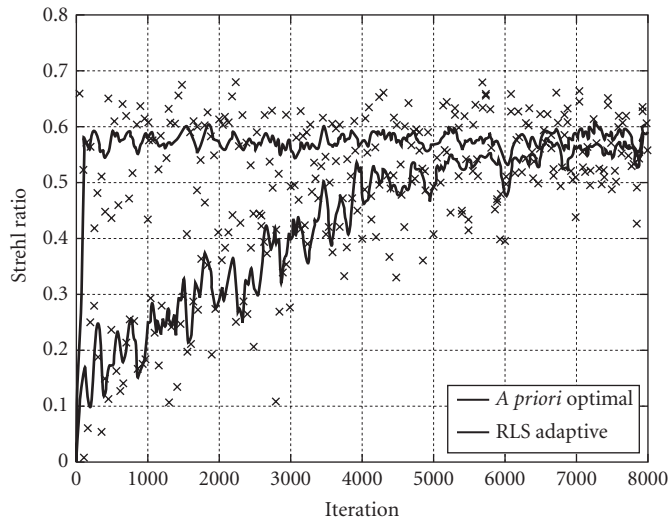


FIGURE 27 Results of a simulation of the Strehl ratio versus time for a recursive least-squares adaptive estimator (lower curve) compared with an optimal estimator having *a priori* knowledge of turbulence conditions (upper curve). The $\times$'s represent the instantaneous Strehl ratio computed by the simulation, and the lines represent the average values.

Other Higher-Order Wavefront Sensing Techniques

Various forms of shearing interferometers have been successfully used to implement wavefront sensors for atmospheric turbulence compensation.^{78, 79} The basic principle is to split the wavefront into two copies, translate one laterally with respect to the other, and then interfere with them. The bright and dark regions of the resulting fringe pattern are proportional to the slope of the wavefront. Furthermore, a lateral shearing interferometer is self-referencing—it does not need a plane wave reference like the Shack-Hartmann sensor. Shearing interferometers are not in widespread use today, but they have been implemented in real systems in the past.⁷⁸

Roddier^{80, 81} introduced a new concept for wavefront sensing based on measuring local wavefront curvature (the second derivative of the phase). The concept can be implemented by differencing the irradiance distributions from two locations on either side of the focal plane of a telescope. If the two locations are displaced a distance, l , from the focus of the telescope and the spatial irradiance distribution is given by $I_1(\mathbf{r})$ and $I_2(\mathbf{r})$ at the two locations, then the relationship between the irradiance and the phase is given by

$$\frac{I_1(\mathbf{r}) - I_2(-\mathbf{r})}{I_1(\mathbf{r}) + I_2(-\mathbf{r})} = \frac{\lambda F^2 (F - l)}{2\pi l^2} \left[\frac{\partial \phi}{\partial n}(\mathbf{r}) \delta_c - \nabla^2 \phi(\mathbf{r}) \right] \quad (56)$$

where F is the focal length of the telescope, and the Dirac delta δ_c represents the outward-pointing normal derivative on the edge of the phase pattern. Equation (56) is valid in the geometrical optics approximation. The distance, l , must be chosen such that the validity of the geometrical optics approximation is ensured, requiring that the blur at the position of the defocused pupil image is small compared with the size of the wavefront aberrations desired to be measured. These considerations lead to a condition on l that

$$l \geq \theta_b \frac{F^2}{d} \quad (57)$$

where θ_b is the blur angle of the object that is produced at the positions of the defocused pupil, and d is the size of the subaperture determined by the size of the detector. For point sources and when $d > r_0$, $\theta_b = \lambda/r_0$ and $l \geq \lambda F^2/r_0 d$. For extended sources of angular size $\theta > \lambda/r_0$, l must be chosen such that $l \geq \theta F^2/d$. Since increasing l decreases the sensitivity of the curvature sensor, in normal operations l is set to satisfy the condition $l \geq \lambda F^2/r_0 d$, but once the loop is closed and low-order aberrations are reduced, the sensitivity of the sensor can be increased by making l smaller. To perform wavefront reconstruction, an iterative procedure can be used to solve Poisson's equation. The appeal of this approach is that certain deformable mirrors such as piezoelectric bimorphs deform locally as nearly spherical shapes, and can be driven directly with no intermediate mathematical wavefront reconstruction step. This has not been completely realized in practice, but excellent results have been obtained with two systems deployed at Mauna Kea.⁸² All implementations of these wavefront sensors to date have employed single-element avalanche photodiodes operating in the photon-counting mode for each subaperture. Since these devices remain quite expensive, it may be cost prohibitive to scale the curvature-sensing technique to very high density actuator systems. An additional consideration is that the noise gain goes up linearly with the number of actuators, not logarithmically as with the Shack-Hartmann or shearing interferometer.

The problem of deriving phase from intensity measurements has been studied extensively.^{83–86} A particular implementation of multiple-intensity measurements to derive phase data has become known as phase diversity.⁸⁷ In this approach, one camera is placed at the focus of the telescope and another in a defocused plane with a known amount of defocus. Intensity gathered simultaneously from both cameras can be processed to recover the phase in the pupil of the telescope. The algorithms needed to perform this operation are complex and require iteration.⁸⁸ Such a technique does not presently lend itself to real-time estimation of phase errors in an adaptive optical system, but it could in the future.

Artificial neural networks have been used to estimate phase from intensity as well. The concept is similar to other methods using multiple-intensity measurements. Two focal planes are set up, one at the focus of the telescope and one near focus. The pixels from each focal plane are fed into the nodes of an artificial neural network. The output of the network can be set up to provide almost any desired

information from Zernike decomposition elements to direct-drive signals to actuators of a zonal, deformable mirror. The network must be trained using known distortions. This can be accomplished by using a deterministic wavefront sensor to measure the aberrations and using that information to adjust the weights and coefficients in the neural network processor. This concept has been demonstrated on real telescopes in atmospheric turbulence.⁸⁹ The concept works for low-order distortions, but it appears to have limited usefulness for large, high-density actuator adaptive optical systems.

Wavefront Correctors

There are three major classes of wavefront correctors in use today: segmented mirrors, bimorph mirrors, and stacked-actuator continuous-facesheet mirrors. Figure 28 shows the concept for each of these mirrors.

The segmented mirror can have piston-only or piston-and-tilt actuators. Since the individual segments are completely independent, it is possible for significant phase errors to develop between the segments. It is therefore important that these errors be controlled by real-time interferometry or by strain gauges or other position-measuring devices on the individual actuators. Some large segmented mirrors have been built⁹⁰ for DOD applications and several are in use today for astronomy.^{91, 92}

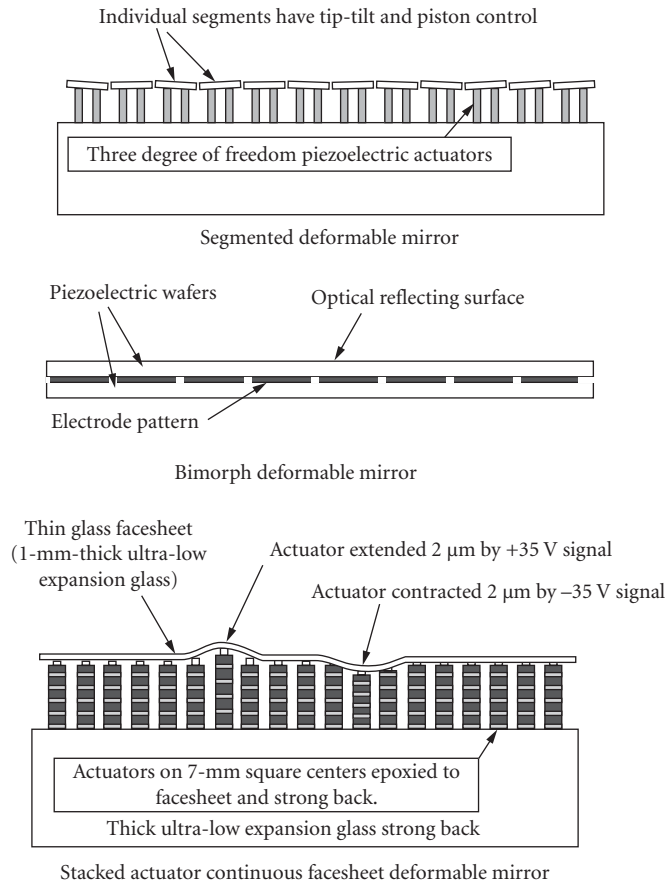


FIGURE 28 Cross sections of three deformable mirror designs.

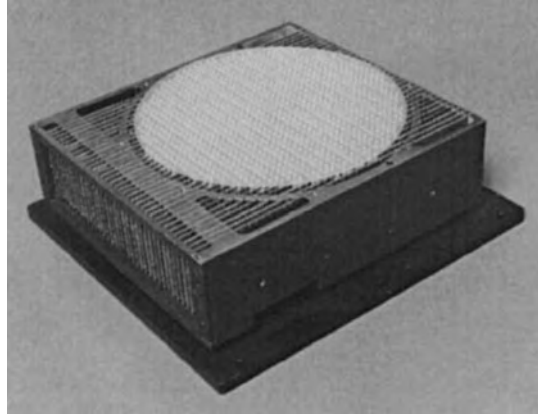


FIGURE 29 The 941-actuator deformable mirror built by Xinetics, Inc., for the SOR 3.5-m telescope.

The stacked actuator deformable mirror is probably the most widely used wavefront corrector. The modern versions are made with lead-magnesium-niobate—a ceramic-like electrostrictive material producing $4\text{ }\mu\text{m}$ of stroke for 70 V of drive.⁹³ The facesheet of these mirrors is typically 1 mm thick. Mirrors with as many as 2200 actuators have been built. These devices are very stiff structures with first resonant frequencies as high as 25 kHz. They are typically built with actuators spaced as closely as 7 mm. One disadvantage with this design is that it becomes essentially impossible to repair individual actuators once the mirror structure is epoxied together. Actuator failures are becoming less likely with today's refined technology, but a very large mirror may have 1000 actuators, which increases its chances for failures over the small mirrors of times past. Figure 29 is a photograph of the 941-actuator deformable mirror in use at the 3.5-m telescope at the SOR.

New Wavefront Corrector Technologies

Our progress toward good performance at visible wavelengths will depend critically on the technology that is available for high-density actuator wavefront correctors. There is promise in the areas of liquid crystals, MEM devices, and nonlinear-optics processes. However, at least for the next few years, it seems that we will have to rely on conventional mirrors with piezoelectric-type actuators or bimorph mirrors made from sandwiched piezoelectric layers. Nevertheless, there is a significant development in the conventional mirror area that has the potential for making mirrors with very large numbers of actuators that are smaller, more reliable, and much less expensive.

5.6 HOW TO DESIGN AN ADAPTIVE OPTICAL SYSTEM

Adaptive optical systems are complex and their performance is governed by many parameters, some controlled by the user and some controlled by nature. The system designer is faced with selecting parameter values to meet performance requirements. Where does he or she begin? One approach is presented here and consists of the following six steps:

1. Determine the average seeing conditions (the mean value of r_0 and f_G) for the site.
2. Determine the most important range of wavelengths of operation.
3. Decide on the minimum Strehl ratio that is acceptable to the users for these seeing conditions and operating wavelengths.
4. Determine the brightness of available beacons.
5. Given the above requirements, determine the optimum values of the subaperture size and the servo bandwidth to minimize the residual wavefront error at the most important wavelength and minimum beacon brightness. (The most difficult parameters to change after the system is built are the wavefront sensor subaperture and deformable mirror actuator geometries. These parameters need to be chosen carefully to address the highest-priority requirements in terms of seeing conditions, beacon brightness, operating wavelength, and required Strehl ratio.) Determine if the associated Strehl ratio is acceptable.
6. Evaluate the Strehl ratio for other values of the imaging wavelength, beacon brightness, and seeing conditions. If these are unsatisfactory, vary the wavefront sensor and track parameters until an acceptable compromise is reached.

Our objective in this section is to develop some practical formulas that can be implemented and evaluated quickly on desktop or laptop computers using programs like Mathematica™ or MatLab™ that will allow one to iterate the six aforementioned steps to investigate the top-level trade space and optimize system performance for the task at hand.

The most important parameters that the designer has some control over are the subaperture size, the wavefront sensor and track sensor integration times, the latency of data in the higher-order and tracker control loops, and the wavelength of operation. One can find optimum values for these parameters since changing them can either increase or decrease the system Strehl ratio. The parameters that the designer has little or no control over are those that are associated with atmospheric turbulence. On the other hand, there are a few parameters that always make performance better the larger (or smaller) we make the parameter: the quantum efficiency of the wavefront and track sensors, the brightness of the beacon(s), the optical system throughput, and read noise of sensors. We can never make a mistake by making the quantum efficiency as large as physics will allow and the read noise as low as physics will allow. We can never have a beacon too bright, nor can we have too much optical transmission (optical attenuators are easy to install). Unfortunately, it is not practical with current technology to optimize the AO system's performance for changing turbulence conditions, for different spectral regions of operation, for different elevation angle, and for variable target brightness by changing the mechanical and optical design from minute to minute. We must be prepared for some compromises based on our initial design choices.

We will use two system examples to illustrate the process that is outlined in the six aforementioned steps: (1) a 3.5-m telescope at an intracontinental site that is used for visible imaging of low-earth-orbiting artificial satellites, and (2) a 10-m telescope that operates at a site of excellent seeing for near-infrared astronomy.

Establish the Requirements

Table 2 lists a set of requirements that we shall try to meet by trading system design parameters. The values in Table 2 were chosen to highlight how requirements can result in significantly different system designs.

In the 3.5-m telescope example, the seeing is bad, the control bandwidths will be high, the imaging wavelength is short, and the required Strehl ratio is significant. Fortunately, the objects (artificial earth satellites) are bright. The 10-m telescope application is much more forgiving with respect to the seeing and imaging wavelengths, but the beacon brightness is four magnitudes fainter and a high Strehl ratio is still required at the imaging wavelength. We now investigate how these requirements determine an optimum choice for the subaperture size.

TABLE 2 Requirements for Two System Examples

Parameter	3.5-m Requirement	10-m Requirement
C_n^2 profile	Modified HV ₅₇	Average Mauna Kea
Wind profile	Slew dominated	Buften
Average r_0 (0.5 μm)	10 cm	18 cm
Average f_G (0.5 μm)	150 Hz	48 Hz
Imaging wavelength	0.85 μm	1.2–2.2 μm
Elevation angle	45°	45°
Minimum Strehl ratio	0.5	0.5
Beacon brightness	$m_v = 6$	$m_v = 10$

Selecting a Subaperture Size

As was mentioned previously, choosing the subaperture size should be done with care, because once a system is built it is not easily changed. The approach is to develop a mathematical expression for the Strehl ratio as a function of system design parameters and seeing conditions and then to maximize the Strehl ratio by varying the subaperture size for the required operating conditions that are listed in Table 2.

The total-system Strehl ratio is the product of the higher-order and full-aperture tilt Strehl ratios:

$$\text{SR}_{\text{sys}} = \text{SR}_{\text{HO}} \cdot \text{SR}_{\text{tilt}} \quad (58)$$

The higher-order Strehl ratio can be estimated as

$$\text{SR}_{\text{HO}} = e^{-[E_n^2 + E_f^2 + E_s^2 + E_{\text{FA}}^2]} \quad (59)$$

where E_n^2 , E_f^2 , E_s^2 , and E_{FA}^2 are the mean square phase errors due to wavefront measurement and reconstruction noise, fitting error, servo lag error, and focus anisoplanatism (if a laser beacon is used), respectively. Equation (59) does not properly account for interactions between these effects and could be too conservative in estimating performance. Ultimately, a system design should be evaluated with a detailed computer simulation, which should include wave-optics atmospheric propagations, diffraction in wavefront sensor optics, details of hardware, processing algorithms and time delays, and many other aspects of the system's engineering design. A high-fidelity simulation will properly account for the interaction of all processes and provide a realistic estimate of system performance. For our purposes here, however, we will treat the errors as independent in order to illustrate the processes that are involved in system design and to show when a particular parameter is no longer the dominant contributor to the total error. We will consider tracking effects later [see below Eq. (66)], but for now we need expressions for components of the higher-order errors so that we may determine an optimum subaperture size.

Wavefront Measurement Error, E_n^2 We will consider a Shack-Hartmann sensor for the discussion that follows. The wavefront measurement error contribution, E_n^2 , can be computed from the equation⁷³

$$E_n^2 = \alpha \left[0.09 \ln(n_{\text{sa}}) \sigma_{\theta_{\text{sa}}}^2 d_s^2 \left(\frac{2\pi}{\lambda_{\text{img}}} \right)^2 \right] \quad (60)$$

where α accounts for control loop averaging and is described below, n_{sa} is the number of subapertures in the wavefront sensor, $\sigma_{\theta_{sa}}^2$ is the angular measurement error over a subaperture, d_s is the length of a side of a square subaperture, and λ_{img} is the imaging wavelength. The angular measurement error in a subaperture is proportional to the angular size of the beacon [normally limited by the seeing for unresolved objects but, in some cases, by the beacon itself (e.g., laser beacons or large astronomical targets)] and is inversely proportional to the signal-to-noise ratio in the wavefront sensor. The value of $\sigma_{\theta_{sa}}$ is given by⁹⁴

$$\sigma_{\theta_{sa}} = \pi \left(\frac{\lambda_b}{d_s} \right) \left[\left(\frac{3}{16} \right)^2 + \left(\frac{\theta_b/2}{4\lambda_b/d_s} \right)^2 \right]^{1/2} \left(\frac{1}{n_s} + \frac{4n_e^2}{n_s^2} \right)^{1/2} \quad (61)$$

where λ_b is the center wavelength of the beacon signal, θ_b is the angular size of the beacon, r_0 is the Fried seeing parameter at 0.5 μm at zenith, ψ is the zenith angle of the observing direction, n_s is the number of photodetected electrons per subaperture per sample, and n_e is the read noise in electrons per pixel. We have used the wavelength and zenith scaling laws for r_0 to account for the angular size of the beacon at the observing conditions.

The number of photodetected electrons per subaperture per millisecond, per square meter, per nanometer of spectral width is given by

$$n_s = \frac{(101500)^{\sec\psi}}{(2.51)^{m_v}} d_s^2 T_{ro} n_{QE} t_s \Delta\lambda \quad (62)$$

where m_v is the equivalent visual magnitude of the beacon at zenith, d_s is the subaperture size in meters, T_{ro} is the transmissivity of the telescope and wavefront sensor optics, n_{QE} is the quantum efficiency of the wavefront sensor, t_s is the integration time per frame of the wavefront sensor, and $\Delta\lambda$ is the wavefront sensor spectral bandwidth in nanometers.

The parameter α is a factor that comes from the filtering process in the control loop and can be considered the mean square gain of the loop. Ellerbroek^{73, 95} has derived an expression (discussed by Milonni et al.⁷³ and by Ellerbroek⁹⁵) for α using the Z-transform method and a simplified control system model. His result is $\alpha = g/(2-g)$, where $g = 2\pi f_{3dB} t_s$ is the gain, f_{3dB} is the control bandwidth, and t_s is the sensor integration time. For a typical control system, we sample at a rate that is 10 times the control bandwidth, making $t_s = 1/(10f_{3dB})$, $g = 0.628$, and $\alpha = 0.458$.

Fitting Error, E_f^2 The wavefront sensor has finite-sized subapertures, and the deformable mirror has a finite number of actuators. There is a limit, therefore, to the spatial resolution to which the system can “fit” a distorted wavefront. The mean square phase error due to fitting is proportional to the $-5/6$ th power of the number of actuators and $(D/r_0)^{5/3}$ and is given by

$$E_f^2 = C_f \left(\frac{D}{r_0 \left(\frac{\lambda_{img}}{0.5 \mu\text{m}} \right)^{6/5} (\cos\psi)^{3/5}} \right)^{5/3} n_a^{-5/6} \quad (63)$$

where C_f is the fitting error coefficient that is determined by the influence function of the deformable mirror and has a value of approximately 0.28 for thin, continuous-facesheet mirrors.

Servo Lag, E_s^2 The mean square wavefront error due to a finite control bandwidth has been described earlier by Eq. (37), which is recast here as

$$E_s^2 = \left(\frac{f_G \left(\frac{\lambda_{img}}{0.5 \mu\text{m}} \right)^{-6/5} (\cos\psi)^{-3/5}}{f_{3dB}} \right)^{5/3} \quad (64)$$

where f_G is the Greenwood frequency scaled for the imaging wavelength and a worst-case wind direction for the zenith angle, and f_{3dB} is the control bandwidth -3 -dB error rejection frequency. Barchers⁷⁶ has developed an expression from which one can determine f_{3dB} for a conventional proportional integral controller, given the sensor sample time (t_s), the additional latency from readout and data-processing time (δ_0), and the design gain margin [G_M (a number between 1 and ∞ that represents the minimum gain that will drive the loop unstable)].⁹⁶ The phase crossover frequency of a proportional integral controller with a fixed latency is $\omega_{cp} = \pi/(2\delta)$ rad/s, where $\delta = t_s + \delta_0$ is the total latency, which is made up of the sample period and the sensor readout and processing time, δ_0 . The loop gain that achieves the design gain margin is $K = \omega_{cp}/G_M$ where G_M is the design gain margin. Barchers shows that f_{3dB} can be found by determining the frequency for which the modulus of the error rejection function is equal to 0.707 or when

$$|S(i\omega_{3dB})| = \left| \frac{i\omega_{3dB}}{i\omega_{3dB} + \frac{\pi/2(t_s + \delta_0)}{G_M} e^{-i(t_s + \delta_0)\omega_{3dB}}} \right| = 0.707 \quad (65)$$

where $f_{3dB} = \omega_{3dB}/2\pi$. This equation can be solved graphically or with dedicated iterative routines that find roots, such as Mathematica. Figure 30 shows the error rejection curves for four combinations of t_s , δ_0 , and G_M , which are detailed in the caption. Note in particular that decreasing the latency δ_0 from 2 ms to 0.5 ms increased f_{3dB} from 14 to 30 Hz (compare two curves on the left), illustrating the sensitivity to readout and processing time.

Focus Anisoplanatism, E_{FA}^2 If laser beacons are being used, the effects of focus anisoplanatism, which are discussed in Sec. 5.5, must be considered since this error often dominates the wavefront sensor error budget. Focus anisoplanatism error is given by

$$E_{FA}^2 = \left(\frac{D}{d_0 \left(\frac{\lambda_{img}}{0.5 \mu m} \right)^{6/5} (\cos \psi)^{3/5}} \right)^{5/3} \quad (66)$$

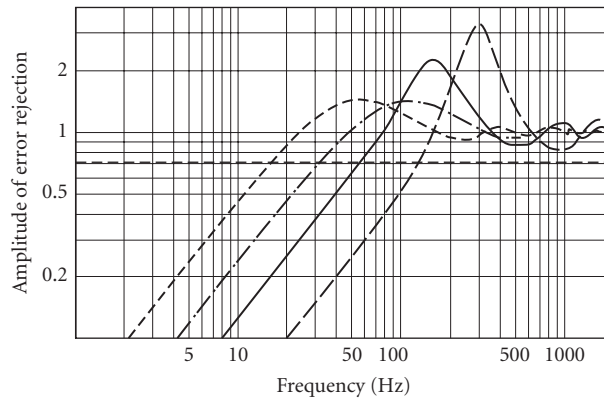


FIGURE 30 Control loop error rejection curves for a proportional integral controller. The curves (left to right) represent the following parameters: dotted ($t_s = 1$ ms, $\delta_0 = 2$ ms, $G_M = 4$); dash-dot ($t_s = 1$ ms, $\delta_0 = 0.5$ ms, $G_M = 4$); solid ($t_s = 667$ μ s, $\delta_0 = 640$ μ s, $G_M = 2$); dashed ($t_s = 400$ μ s, $\delta_0 = 360$ μ s, $G_M = 1.5$). f_{3dB} is determined by the intersection of the horizontal dotted line with each of the three curves and has values of 14, 30, 55, and 130 Hz, respectively.

where d_0 is the section size given by Eqs. (50) and (51), scaled for the imaging wavelength and the zenith angle of observation.

Tracking The tracking system is also very important to the performance of an AO system and should not be overlooked as an insignificant problem. In most instances, a system will contain tilt disturbances that are not easily modeled or analyzed arising most commonly from the telescope mount and its movement and base motions coupled into the telescope from the building and its machinery or other seismic disturbances. As described earlier in Eq. (22), the Strehl ratio due to full-aperture tilt variance, σ_θ^2 , is

$$\text{SR}_{\text{tilt}} = \frac{1}{1 + \frac{\pi^2}{2} \left(\frac{\sigma_\theta}{\lambda/D} \right)^2} \quad (67)$$

As mentioned previously, the total system Strehl ratio is then the product of these and is given by

$$\text{SR}_{\text{sys}} = \text{SR}_{\text{HO}} \cdot \text{SR}_{\text{tilt}} \quad (68)$$

Results for 3.5-m Telescope AO System

When the preceding equations are evaluated, we can determine the dependence of the system Strehl ratio on d_s for targets of different brightness. Figure 31 shows the Strehl ratio versus subaperture size for four values of target brightness corresponding to (top to bottom) $m_v = 5, 6, 7$, and 8 and for other system parameters as shown in the figure caption.

Figure 31 shows that we can achieve the required Strehl ratio of 0.5 (see Table 2) for a $m_v = 6$ target by using a subaperture size of about 11 to 12 cm. We could get slightly more performance ($\text{SR} = 0.52$) by reducing the subaperture size to 8 cm, but at significant cost since we would have nearly twice as many subapertures and deformable mirror actuators. Notice also that for fainter

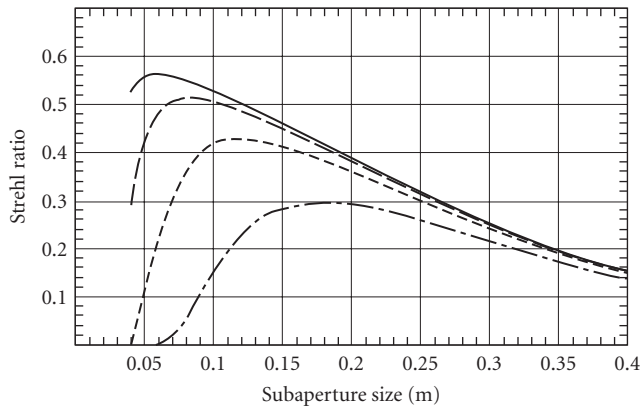


FIGURE 31 System Strehl ratio as a function of subaperture size for the 3.5-m telescope example. Values of m_v are (top to bottom) 5, 6, 7, and 8. Other parameters are: $r_0 = 10$ cm, $f_G = 150$ Hz, $t_s = 400$ μ s, $\delta_0 = 360$ μ s, $G_M = 1.5$, $f_{3dB} = 127$ Hz, $\lambda_{\text{img}} = 0.85$ μ m, $\psi = 45^\circ$, $T_{ro} = 0.25$, $\eta_{QE} = 0.90$, $\Delta\lambda = 400$ nm.

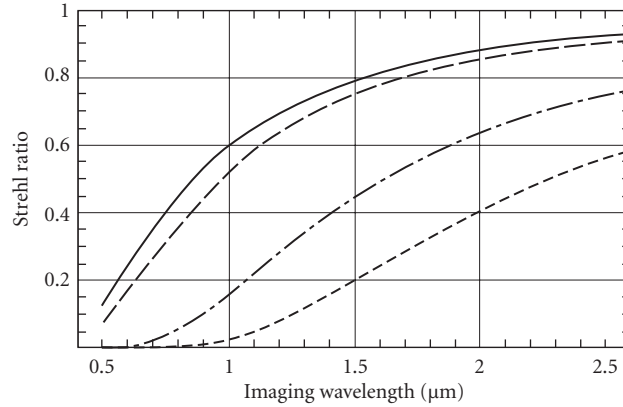


FIGURE 32 System Strehl ratio as a function of imaging wavelength for the 3.5-m telescope example. Curves are for (top to bottom) $m_V = 6$, $d_s = 12$ cm; $m_V = 6$, $d_s = 18$ cm; $m_V = 8$, $d_s = 18$ cm; $m_V = 8$, $d_s = 12$ cm. Other parameters are: $r_0 = 10$ cm, $f_G = 150$ Hz, $t_s = 400$ μ s, $\delta_0 = 360$ μ s, $G_M = 1.5$, $f_{3dB} = 127$ Hz, $\psi = 45^\circ$, $T_{ro} = 0.25$, $\eta_{QE} = 0.90$, $\Delta\lambda = 400$ nm.

objects, a larger subaperture (18 cm) would be optimum for an $m_V = 8$ target, but the SR would be down to 0.4 for the $m_V = 6$ target and would not meet the requirement.

Figure 32 shows the performance of the system for sixth- and eighth-magnitude targets as a function of wavelength. The top two curves in this figure are for 12-cm subapertures (solid curve) and 18-cm subapertures (dashed curve) for the $m_V = 6$ target. Notice that the 12-cm subapertures have better performance at all wavelengths with the biggest difference in the visible. The bottom two curves are for 18-cm subapertures (dash-dot curve) and 12-cm subapertures (dotted curve) for the $m_V = 8$ target. These curves show quantitatively the trade-off between two subaperture sizes and target brightness.

Figure 33 shows how the system will perform for different seeing conditions (values of r_0) and as a function of target brightness. These curves are for a subaperture choice of 12 cm. The curves in

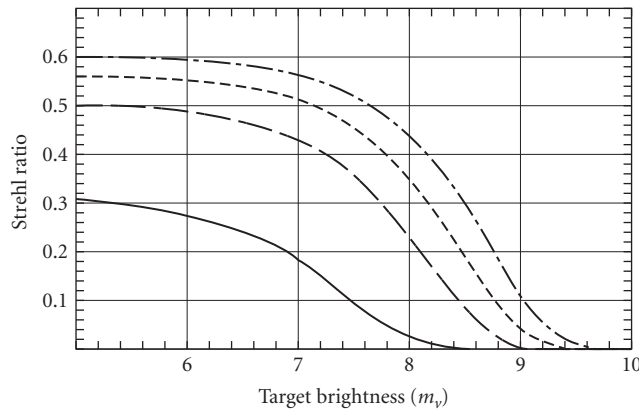


FIGURE 33 System Strehl ratio as a function of target brightness for the 3.5-m telescope example. Curves are for values of r_0 of (top to bottom) 25, 15, 10, and 5 cm. Other parameters are: $d_s = 12$ cm, $f_G = 150$ Hz, $t_s = 400$ μ s, $\delta_0 = 360$ μ s, $G_M = 1.5$, $f_{3dB} = 127$ Hz, $\lambda_{img} = 0.85$ μ m, $\psi = 45^\circ$, $T_{ro} = 0.25$, $\eta_{QE} = 0.90$, $\Delta\lambda = 400$ nm.

Figs. 31 through 33 give designers a good feel of what to expect and how to do top-level system trades for those conditions and parameters for which only they and the users can set the priorities. These curves are easily and quickly generated with modern mathematics packages.

Results for the 10-m AO Telescope System

Similar design considerations for the 10-m telescope example lead to Figs. 34 through 36. Figure 34 shows the Strehl ratio for an imaging wavelength of $1.2\ \mu\text{m}$ versus the subaperture size for four

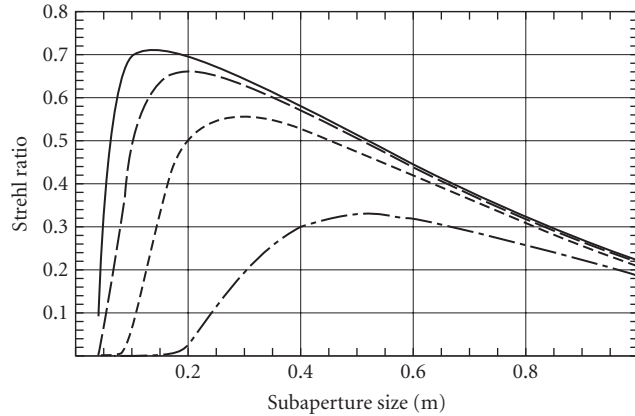


FIGURE 34 System Strehl ratio as a function of subaperture size for the 10-m telescope example. Values of m_v are (top of bottom) 8, 9, 10, and 11. Other parameters are: $r_0 = 18\text{ cm}$, $f_G = 48\text{ Hz}$, $t_s = 1000\ \mu\text{s}$, $\delta_0 = 1000\ \mu\text{s}$, $G_M = 2$, $f_{3\text{dB}} = 39\text{ Hz}$, $\lambda_{\text{img}} = 1.2\ \mu\text{m}$, $\psi = 45^\circ$, $T_{\text{ro}} = 0.25$, $\eta_{\text{QE}} = 0.90$, $\Delta\lambda = 400\text{ nm}$.

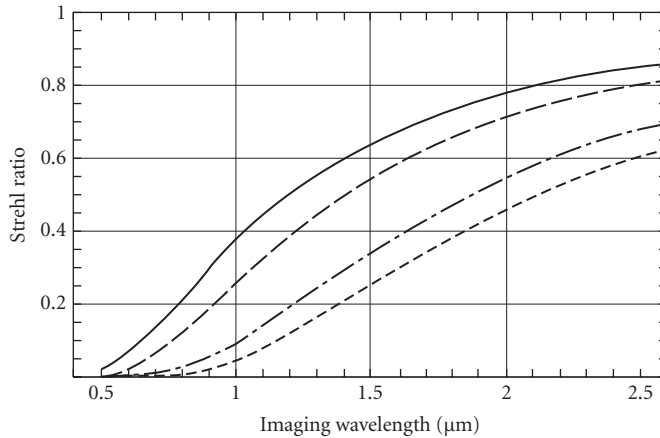


FIGURE 35 System Strehl ratio as a function of imaging wavelength for the 10-m telescope example. Curves are for (top to bottom) $m_v = 10$, $d_s = 45\text{ cm}$; $m_v = 10$, $d_s = 65\text{ cm}$; $m_v = 12$, $d_s = 65\text{ cm}$; $m_v = 12$, $d_s = 45\text{ cm}$. Other parameters are: $r_0 = 18\text{ cm}$, $f_G = 48\text{ Hz}$, $t_s = 1000\ \mu\text{s}$, $\delta_0 = 1000\ \mu\text{s}$, $G_M = 2$, $f_{3\text{dB}} = 39\text{ Hz}$, $\psi = 45^\circ$, $T_{\text{ro}} = 0.25$, $\eta_{\text{QE}} = 0.90$, $\Delta\lambda = 400\text{ nm}$.

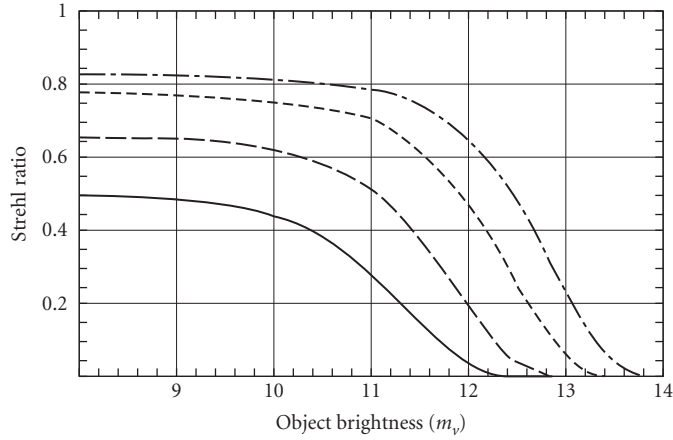


FIGURE 36 System Strehl ratio as a function of target brightness for the 10-m telescope example. Curves are for values of r_0 of (top to bottom) 40, 25, 15, and 10 cm. Other parameters are: $d_s = 45$ cm, $f_G = 48$ Hz, $t_s = 1000$ μ s, $\delta_0 = 1000$ μ s, $G_M = 2$, $f_{3dB} = 39$ Hz, $\lambda_{img} = 1.6$ μ m, $\psi = 45^\circ$, $T_{ro} = 0.25$, $\eta_{QE} = 0.90$, $\Delta\lambda = 400$ nm.

values of beacon brightness corresponding to (top to bottom) $m_v = 8, 9, 10$, and 11 . Other system parameters are listed in the figure caption. These curves show that we can achieve the required Strehl ratio of 0.5 for the $m_v = 10$ beacon with subapertures in the range of 20 to 45 cm. Optimum performance at 1.2 μ m produces a Strehl ratio of 0.56, with a subaperture size of 30 cm. Nearly 400 actuators are needed in the deformable mirror for 45-cm subapertures, and nearly 900 actuators are needed for 30-cm subapertures. Furthermore, Fig. 34 shows that 45 cm is a better choice than 30 cm in the sense that it provides near-optimal performance for the $m_v = 11$ beacon, whereas the Strehl ratio is down to 0.2 for the 30-cm subapertures.

Figure 35 shows system performance as a function of imaging wavelength. The top two curves are for the $m_v = 10$ beacon, with $d_s = 45$ and 65 cm, respectively. Note that the 45-cm subaperture gives excellent performance with a Strehl ratio of 0.8 at 2.2 μ m (the upper end of the required spectral range) and a very useful Strehl ratio of 0.3 at 0.8 μ m. A choice of $d_s = 65$ cm provides poorer performance for $m_v = 10$ (in fact, it does not satisfy the requirement) than does $d_s = 45$ cm due to fitting error, whereas a choice of $d_s = 65$ cm provides better performance for $m_v = 12$ due to improved wavefront sensor signal-to-noise ratio.

Figure 36 shows system performance in different seeing conditions as a function of beacon brightness for an intermediate imaging wavelength of 1.6 μ m. These curves are for a subaperture size of 45 cm. This chart predicts that in exceptional seeing, a Strehl ratio of 0.5 μ m can be achieved using an $m_v = 12.5$ beacon.

As in the 3.5-m telescope case, these curves and others like them with different parameters can be useful in performing top-level design trades and in selecting a short list of design candidates for further detailed analysis and simulation.

5.7 ACKNOWLEDGMENTS

I would like to thank David L. Fried, Earl Spillar, Jeff Barchers, John Anderson, Bill Lowrey, and Greg Peisert for reading the manuscript and making constructive suggestions that improved the content and style of this chapter. I would especially like to acknowledge the efforts of my editor, Bill Wolfe, who relentlessly kept me on course from the first pitiful draft.

5.8 REFERENCES

1. D. R. Williams, J. Liang, and D. T. Miller, "Adaptive Optics for the Human Eye," *OSA Technical Digest 1996* **13**:145–147 (1996).
2. J. Liang, D. Williams, and D. Miller, "Supernormal Vision and High-Resolution Retinal Imaging through Adaptive Optics," *JOSA A* **14**:2884–2892 (1997).
3. A. Roorda and D. Williams, "The Arrangement of the Three Cone Classes in the Living Human Eye," *Nature* **397**:520–522 (1999).
4. K. Bar, B. Freisleben, C. Kozlik, and R. Schmiedl, "Adaptive Optics for Industrial CO₂-Laser Systems," *Lasers in Engineering*, vol. 4, no. 3, unknown publisher, 1961.
5. M. Huonker, G. Waibel, A. Giesen, and H. Hugel, "Fast and Compact Adaptive Mirror for Laser Materials Processing," *Proc. SPIE* **3097**:310–319 (1997).
6. R. Q. Fugate, "Laser Beacon Adaptive Optics for Power Beaming Applications," *Proc SPIE* **2121**:68–76 (1994).
7. H. E. Bennett, J. D. G. Rather, and E. E. Montgomery, "Free-Electron Laser Power Beaming to Satellites at China Lake, California," *Proc SPIE* **2121**:182–202 (1994).
8. C. R. Phipps, G. Albrecht, H. Friedman, D. Gavel, E. V. George, J. Murray, C. Ho, W. Priedhorsky, M. M. Michaels, and J. P. Reilly, "ORION: Clearing Near-Earth Space Debris Using a 20-kW, 530-nm, Earth-Based Repetitively Pulsed Laser," *Laser and Particle Beams* **14**:1–44 (1996).
9. K. Wilson, J. Lesh, K. Araki, and Y. Arimoto, "Overview of the Ground to Orbit Lasercom Demonstration," *Space Communications* **15**:89–95 (1998).
10. R. Szeto and R. Butts, "Atmospheric Characterization in the Presence of Strong Additive Measurement Noise," *JOSA A* **15**:1698–1707 (1998).
11. H. Babcock, "The Possibility of Compensating Astronomical Seeing," *Publications of the Astronomical Society of the Pacific* **65**:229–236 (October 1953).
12. D. Fried, "Optical Resolution through a Randomly Inhomogeneous Medium for Very Long and Very Short Exposures," *J. Opt. Soc. Am.* **56**:1372–1379 (October 1966).
13. R. P. Angel, "Development of a Deformable Secondary Mirror," *Proceedings of the SPIE* **1400**:341–351 (1997).
14. S. F. Clifford, "The Classical Theory of Wave Propagation in a Turbulent Medium," *Laser Beam Propagation in the Atmosphere*, Springer-Verlag, New York, 1978.
15. A. N. Kolmogorov, "The Local Structure of Turbulence in Incompressible Viscous Fluids for Very Large Reynolds' Numbers," *Turbulence, Classic Papers on Statistical Theory*, Wiley-Interscience, New York, 1961.
16. V. I. Tatarski, *Wave Propagation in a Turbulent Medium*, McGraw-Hill, New York, 1961.
17. D. Fried, "Statistics of a Geometrical Representation of Wavefront Distortion," *J. Opt. Soc. Am.* **55**:1427–1435 (November 1965).
18. J. W. Goodman, *Statistical Optics*, Wiley-Interscience, New York, 1985.
19. F. D. Eaton, W. A. Peterson, J. R. Hines, K. R. Peterman, R. E. Good, R. R. Beland, and J. H. Brown, "Comparisons of vhf Radar, Optical, and Temperature Fluctuation Measurements of C_n^2 , r_0 , and θ_0 ," *Theoretical and Applied Climatology* **39**:17–29 (1988).
20. D. L. Walters and L. Bradford, "Measurements of r_0 and θ_0 : 2 Decades and 18 Sites," *Applied Optics* **36**:7876–7886 (1997).
21. R. E. Hufnagel, *Proc. Topical Mtg. on Optical Propagation through Turbulence*, Boulder, CO **1** (1974).
22. G. Valley, "Isoplanatic Degradation of Tilt Correction and Short-Term Imaging Systems," *Applied Optics* **19**:574–577 (February 1980).
23. D. Winker, "Unpublished Air Force Weapons Lab Memo, 1986," U.S. Air Force, 1986.
24. E. Marchetti and D. Bonaccini, "Does the Outer Scale Help Adaptive Optics or Is Kolmogorov Gentler?" *Proc. SPIE* **3353**:1100–1108 (1998).
25. R. E. Hufnagel and N. R. Stanley, "Modulation Transfer Function Associated with Image Transmission through Turbulent Media," *J. Opt. Soc. Am.* **54**:52–61 (January 1964).
26. D. Fried, "Limiting Resolution Looking down through the Atmosphere," *J. Opt. Soc. Am.* **56**:1380–1384 (October 1966).

27. R. Noll, "Zernike Polynomials and Atmospheric Turbulence," *J. Opt. Soc. Am.* **66**:207–211 (March 1976).
28. J. Christou, "Deconvolution of Adaptive Optics Images," *Proceedings, ESO/OSA Topical Meeting on Astronomy with Adaptive Optics: Present Results and Future Programs* **56**:99–108 (1998).
29. G. A. Tyler, *Reduction in Antenna Gain Due to Random Jitter*, The Optical Sciences Company, Anaheim, CA, 1983.
30. H. T. Yura and M. T. Tavis, "Centroid Anisoplanatism," *JOSA A* **2**:765–773 (1985).
31. D. P. Greenwood and D. L. Fried, "Power Spectra Requirements for Wave-Front-Compensative Systems," *J. Opt. Soc. Am.* **66**:193–206 (March 1976).
32. G. A. Tyler, "Bandwidth Considerations for Tracking through Turbulence," *JOSA A* **11**:358–367 (1994).
33. R. J. Sasiela, *Electromagnetic Wave Propagation in Turbulence*, Springer-Verlag, New York, 1994.
34. J. Bufton, "Comparison of Vertical Profile Turbulence Structure with Stellar Observations," *Appl. Opt.* **12**:1785 (1973).
35. D. Greenwood, "Tracking Turbulence-Induced Tilt Errors with Shared and Adjacent Apertures," *J. Opt. Soc. Am.* **67**:282–290 (March 1977).
36. G. A. Tyler, "Turbulence-Induced Adaptive-Optics Performance Degradation: Evaluation in the Time Domain," *JOSA A* **1**:358 (1984).
37. D. Fried, "Anisoplanatism in Adaptive Optics," *J. Opt. Soc. Am.* **72**:52–61 (January 1982).
38. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, San Francisco, 1968.
39. J. Ge, "Adaptive Optics," *OSA Technical Digest Series* **13**:122 (1996).
40. D. W. Tyler and B. L. Ellerbroek, "Sky Coverage Calculations for Spectrometer Slit Power Coupling with Adaptive Optics Compensation," *Proc. SPIE* **3353**:201–209 (1998).
41. J. Ge, R. Angel, D. Sandler, C. Shelton, D. McCarthy, and J. Burge, "Adaptive Optics Spectroscopy: Preliminary Theoretical Results," *Proc. SPIE* **3126**:343–354 (1997).
42. G. Tyler, "Rapid Evaluation of d0," Tech. Rep. TR-1159, The Optical Sciences Company, Placentia, CA, 1991.
43. R. Racine and R. McClure, "An Image Stabilization Experiment at the Canada-France-Hawaii Telescope," *Publications of the Astronomical Society of the Pacific* **101**:731–736 (August 1989).
44. R. Q. Fugate, J. F. Riker, J. T. Roark, S. Stogsdill, and B. D. O'Neil, "Laser Beacon Compensated Images of Saturn Using a High-Speed Near-Infrared Correlation Tracker," *Proc. Top. Mtg. on Adaptive Optics, ESO Conf. and Workshop Proc.* **56**:287 (1996).
45. E. Wallner, "Optimal Wave-Front Correction Using Slope Measurements," *J. Opt. Soc. Am.* **73**:1771–1776 (December 1983).
46. D. L. Fried, "Least-Squares Fitting a Wave-Front Distortion Estimate to an Array of Phase-Difference Measurements," *J. Opt. Soc. Am.* **67**:370–375 (1977).
47. W. J. Wild, "Innovative Wavefront Estimators for Zonal Adaptive Optics Systems, ii," *Proc. SPIE* **3353**:1164–1173 (1998).
48. B. L. Ellerbroek and T. A. Rhoadarmer, "Real-Time Adaptive Optimization of Wave-Front Reconstruction Algorithms for Closed Loop Adaptive-Optical Systems," *Proc. SPIE* **3353**:1174–1185 (1998).
49. R. B. Shack and B. C. Platt, "Production and Use of a Lenticular Hartman Screen," *J. Opt. Soc. Am.* **61**:656 (1971).
50. B. R. Hunt, "Matrix Formulation of the Reconstruction of Phase Values from Phase Differences," *J. Opt. Soc. Am.* **69**:393 (1979).
51. R. H. Hudgin, "Optimal Wave-Front Estimation," *J. Opt. Soc. Am.* **67**:378–382 (1977).
52. R. J. Sasiela and J. G. Mooney, "An Optical Phase Reconstructor Based on Using a Multiplier-Accumulator Approach," *Proc. SPIE* **551**:170 (1985).
53. R. Q. Fugate, D. L. Fried, G. A. Ameer, B. R. Boeke, S. L. Browne, P. H. Roberts, R. E. Ruane, G. A. Tyler, and L. M. Wopat, "Measurement of Atmospheric Wavefront Distortion Using Scattered Light from a Laser Guide-Star," *Nature* **353**:144–146 (September 1991).
54. C. A. Primmerman, D. V. Murphy, D. A. Page, B. G. Zollars, and H. T. Barclay, "Compensation of Atmospheric Optical Distortion Using a Synthetic Beacon," *Nature* **353**:140–141 (1991).
55. R. Q. Fugate, B. L. Ellerbroek, C. H. Higgins, M. P. Jelonek, W. J. Lange, A. C. Slavin, W. J. Wild, D. M. Winker, J. M. Wynia, J. M. Spinhirne, B. R. Boeke, R. E. Ruane, J. F. Moroney, M. D. Olier,

- D. W. Swindle, and R. A. Cleis, "Two Generations of Laser-Guide-Star Adaptive Optics Experiments at the Starfire Optical Range," *JOSA A* **11**:310–324 (1994).
56. L. A. Thompson, R. M. Castle, S. W. Teare, P. R. McCullough, and S. Crawford, "Unisic: A Laser Guided Adaptive Optics System for the Mt. Wilson 2.5-m Telescope," *Proc. SPIE* **3353**:282–289 (1998).
 57. S. S. Olivier, D. T. Gavel, H. W. Friedman, C. E. Max, J. R. An, K. Avicola, B. J. Bauman, J. M. Brase, E. W. Campbell, C. Carrano, J. B. Cooke, G. J. Freeze, E. L. Gates, V. K. Kanz, T. C. Kuklo, B. A. Macintosh, M. J. Newman, E. L. Pierce, K. E. Waltjen, and J. A. Watson, "Improved Performance of the Laser Guide Star Adaptive Optics System at Lick Observatory," *Proc. SPIE* **3762**:2–7 (1999).
 58. D. J. Butler, R. I. Davies, H. Fews, W. Hackenburg, S. Rabien, T. Ott, A. Eckart, and M. Kasper, "Calar Alto Affa and the Sodium Laser Guide Star in Astronomy," *Proc. SPIE* **3762**:184–193 (1999).
 59. R. Q. Fugate, "Laser Guide Star Adaptive Optics for Compensated Imaging," *The Infrared and Electro-Optical Systems Handbook*, S. R. Robinson, (ed.), vol. 8, 1993.
 60. See the special edition of *JOSA A* on Atmospheric-Compensation Technology (January-February 1994).
 61. R. Foy and A. Labeyrie, "Feasibility of Adaptive Telescope with Laser Probe," *Astronomy and Astrophysics* **152**: L29–L31 (1985).
 62. D. L. Fried and J. F. Belsher, "Analysis of Fundamental Limits to Artificial-Guide-Star Adaptive-Optics-System Performance for Astronomical Imaging," *JOSA A* **11**:277–287 (1994).
 63. G. A. Tyler, "Rapid Evaluation of d_0 : The Effective Diameter of a Laser-Guide-Star Adaptive-Optics System," *JOSA A* **11**:325–338 (1994).
 64. R. Penndorf, "Tables of the Refractive Index for Standard Air and the Rayleigh Scattering Coefficient for the Spectral Region Between 0.2 and 20.0 μm and Their Application to Atmospheric Optics," *J. Opt. Soc. Am.* **47**:176–182 (1957).
 65. *U.S. Standard Atmosphere*, National Oceanic and Atmospheric Administration, Washington, D.C., 1976.
 66. R. Q. Fugate, "Observations of Faint Objects with Laser Beacon Adaptive Optics," *Proceedings of the SPIE* **2201**:10–21 (1994).
 67. W. Happer, G. J. MacDonald, C. E. Max, and F. J. Dyson, "Atmospheric Turbulence Compensation by Resonant Optical Backscattering from the Sodium Layer in the Upper Atmosphere," *JOSA A* **11**:263–276 (1994).
 68. H. Friedman, G. Erbert, T. Kuklo, T. Salmon, D. Smauley, G. Thompson, J. Malik, N. Wong, K. Kanz, and K. Neeb, "Sodium Beacon Laser System for the Lick Observatory," *Proceedings of the SPIE* **2534**:150–160 (1995).
 69. T. H. Jeys, "Development of a Mesospheric Sodium Laser Beacon for Atmospheric Adaptive Optics," *The Lincoln Laboratory Journal* **4**:133–150 (1991).
 70. T. H. Jeys, A. A. Brailove, and A. Mooradian, "Sum Frequency Generation of Sodium Resonance Radiation," *Applied Optics* **28**:2588–2591 (1991).
 71. M. P. Jelonek, R. Q. Fugate, W. J. Lange, A. C. Slavin, R. E. Ruane, and R. A. Cleis, "Characterization of artificial guide stars generated in the mesospheric sodium layer with a sum-frequency laser," *JOSA A* **11**:806–812 (1994).
 72. E. J. Kibblewhite, R. Vuilleumier, B. Carter, W. J. Wild, and T. H. Jeys, "Implementation of CW and Pulsed Laser Beacons for Astronomical Adaptive Optics," *Proceedings of the SPIE* **2201**:272–283 (1994).
 73. P. W. Milonni, R. Q. Fugate, and J. M. Telle, "Analysis of Measured Photon Returns from Sodium Beacons," *JOSA A* **15**:217–233 (1998).
 74. P. W. Milonni, H. Fern, J. M. Telle, and R. Q. Fugate, "Theory of Continuous-Wave Excitation of the Sodium Beacon," *JOSA A* **16**:2555–2566 (1999).
 75. R. J. Eager, "Application of a Massively Parallel DSP System Architecture to Perform Wavefront Reconstruction for a 941 Channel Adaptive Optics System," *Proceedings of the ICSPAT* **2**:1499–1503 (1977).
 76. J. Barchers, Air Force Research Laboratory/DES, Starfire Optical Range, Kirtland AFB, NM, private communication, 1999.
 77. E. Gendron and P. Lena, "Astronomical Adaptive Optics in Modal Control Optimization," *Astron. Astrophys.* **291**:337–347 (1994).
 78. J. W. Hardy, J. E. Lefebvre, and C. L. Koliopoulos, "Real-Time Atmospheric Compensation," *J. Opt. Soc. Am.* **67**:360–367 (1977); and J. W. Hardy, *Adaptive Optics for Astronomical Telescopes*, Oxford University Press, Oxford, 1998.

79. J. Wyant, "Use of an AC Heterodyne Lateral Shear Interferometer with Real-Time Wavefront Correction Systems," *Applied Optics* **14**:2622–2626 (November 1975).
80. F. Roddier, "Curvature Sensing and Compensation: A New Concept in Adaptive Optics," *Applied Optics* **27**:1223–1225 (April 1988).
81. F. Roddier, *Adaptive Optics in Astronomy*, Cambridge University Press, Cambridge, England, 1999.
82. F. Roddier and F. Rigault, "The VH-CFHT Systems," *Adaptive Optics in Astronomy*, ch. 9, F. Roddier, (ed.), Cambridge University Press, Cambridge, England, 1999.
83. J. R. Fienup, "Phase Retrieval Algorithms: A Comparison," *Appl. Opt.* **21**:2758 (1982).
84. R. A. Gonsalves, "Fundamentals of wavefront sensing by phase retrieval," *Proc. SPIE* **351**, p. 56, 1982.
85. J. T. Foley and M. A. A. Jalil, "Role of Diffraction in Phase Retrieval from Intensity Measurements," *Proc. SPIE* **351**:80 (1982).
86. S. R. Robinson, "On the Problem of Phase from Intensity Measurements," *J. Opt. Soc. Am.* **68**:87 (1978).
87. R. G. Paxman, T. J. Schultz, and J. R. Fienup, "Joint Estimation of Object and Aberrations by Using Phase Diversity," *JOSA A* **9**:1072–1085 (1992).
88. R. L. Kendrick, D. S. Acton, and A. L. Duncan, "Phase Diversity Wave-Front Sensor for Imaging Systems," *Appl. Opt.* **33**:6533–6546 (1994).
89. D. G. Sandler, T. Barrett, D. Palmer, R. Fugate, and W. Wild, "Use of a Neural Network to Control an Adaptive Optics System for an Astronomical Telescope," *Nature* **351**:300–302 (May 1991).
90. B. Hulburd and D. Sandler, "Segmented Mirrors for Atmospheric Compensation," *Optical Engineering* **29**:1186–1190 (1990).
91. D. S. Acton, "Status of the Lockheed 19-Segment Solar Adaptive Optics System," *Real Time and Post Facto Solar Image Correction*, Proc. Thirteenth National Solar Observatory, Sacramento Peak, Summer Shop Series 13, 1992.
92. D. F. Busher, A. P. Doel, N. Andrews, C. Dunlop, P. W. Morris, and R. M. Myers, "Novel Adaptive Optics with the Durham University Electra System," *Adaptive Optics*, Proc. OSA/ESO Conference Tech Digest, Series 23, 1995.
93. M. A. Ealey and P. A. Davis, "Standard Select Electrostrictive PMN Actuators for Active and Adaptive Components," *Optical Engineering* **29**:1373–1382 (1990).
94. G. A. Tyler and D. L. Fried, "Image-Position Error Associated with a Quadrant Detector," *J. Opt. Soc. Am.* **72**:804–808 (1982).
95. B. L. Ellerbroek, Gemini Telescopes Project, Hilo, Hawaii, private communication, 1998.
96. C. L. Phillips and H. T. Nagle, *Digital Control System: Analysis and Design*, Prentice Hall, Upper Saddle River, NJ, 1990.

PART

3

MODULATORS

This page intentionally left blank

ACOUSTO-OPTIC DEVICES

I-Cheng Chang

*Accord Optics
Sunnyvale, California*

6.1 GLOSSARY

$\delta\theta_o, \delta\theta_a$	divergence: optical, acoustic
ΔB_m	impermeability tensor
$\Delta f, \Delta F$	bandwidth, normalized bandwidth
Δn	birefringence
$\Delta\theta$	deflection angle
λ_o, λ	optical wavelength (in vacuum/medium)
Λ	acoustic wavelength
ρ	density
τ	acoustic transit time
ψ	phase mismatch function
A	optical to acoustic divergence ratio
a	optical to acoustic wavelength ratio
D	optical aperture
E_i, E_d	electric field, incident, diffracted light
f, F	acoustic frequency, normalized acoustic frequency
H	acoustic beam height
k_i, k_o, k_a	wavevector: incident, diffracted light, acoustic wave
L, l	interaction length, normalized interaction length
L_o	characteristic length
M	figure of merit
n_o, n_e	refractive index: ordinary, extraordinary
P_o, P_d	acoustic power, acoustic power density
P, P_{mn}, P_{ijkl}	elasto-optic coefficient

S, S_i	strain, strain tensor components
t_r, T	rise time scan time
V	acoustic velocity
W	bandpass function

6.2 INTRODUCTION

When an acoustic wave propagates in an optically transparent medium, it produces a periodic modulation of the index of refraction via the elasto-optical effect. This provides a moving phase grating which may diffract portions of an incident light into one or more directions. This phenomenon, known as the acousto-optic (AO) diffraction, has led to a variety of optical devices that can be broadly grouped into AO deflectors, modulators, and tunable filters to perform spatial, temporal, and spectral modulations of light. These devices have been used in optical systems for light-beam control, optical signal processing, and optical spectrometry applications.

Historically, the diffraction of light by acoustic waves was first predicted by Brillouin¹ in 1921. Nearly a decade later, Debye and Sears² and Lucas and Biquard³ experimentally observed the effect. In contrast to Brillouin's prediction of a single diffraction order, a large number of diffraction orders were observed. This discrepancy was later explained by the theoretical work of Raman and Nath.⁴ They derived a set of coupled wave equations that fully described the AO diffraction in unbounded isotropic media. The theory predicts two diffraction regimes; the Raman-Nath regime, characterized by the multiple of diffraction orders, and the Bragg regime, characterized by a single diffraction order. Discussion of the early work on AO diffraction can be found in Ref. 5.

The earlier theoretically work tend to treat AO diffraction from a mathematical point of view, and for decades, solving the multiple-order Raman-Nath diffraction has been the primary interest on acousto-optics research. As such, the early development did not lead to any AO devices for practical applications prior to the invention of the laser. It was the need of optical devices for laser beam modulation and deflection that stimulated extensive research on the theory and practice of AO devices. Significant progress has been made in the decade from 1966 to 1976, due to the development of superior AO materials and efficient broadband ultrasonic transducers. During this period several important research results of AO devices and techniques were reported. These include the works of Gordon⁶ on the theory of AO diffraction in finite interaction geometry, by Korpel et al. on the use of acoustic beam steering,⁷ the study of AO interaction in anisotropic media by Dixon;⁸ and the invention of AO tunable filter by Harris and Wallace⁹ and Chang.¹⁰ As a result of these basic theoretical works, various AO devices were developed and demonstrated its use for laser beam control and optical spectrometer applications. Several review papers during this period are listed in Refs. 11 to 14.

Intensive research programs in the 1980s and early 1990s further advanced the AO technology in order to explore the unique potential as real-time spatial light modulators (SLMs) for optical signal processing and remote sensing applications. By 1995, the technology had matured, and a wide range of high performance AO devices operating from UV to IR spectral regions had become commercially available. These AO devices have been integrated with other photonic components and deployed into optical systems with electronic technology in diverse applications.

It is the purpose of this chapter to review the theory and practice of bulk-wave AO devices and their applications. In addition to bulk AO, there have also been studies based on the interaction of optical guided waves and surface acoustic waves (SAW). Since the basic AO interaction structure and fabrication process is significantly different from that of the bulk acousto-optics, this subject is treated separately in Chap. 7.

This chapter is organized as follows: Section 6.3 discusses the theory of acousto-optic interaction. It provides the necessary background for the design of acousto-optic devices. The subject of acousto-optic materials is discussed in Sec. 6.4. The next three sections deal with the three basic types of acousto-optic devices. Detailed discussion of AO deflectors, modulators, and tunable filters are presented in Section 6.5, 6.6, and 6.7, respectively.

6.3 THEORY OF ACOUSTO-OPTIC INTERACTION

Elasto-Optic Effect

The elasto-optic effect is the basic mechanism responsible for the AO interaction. It describes the change of refractive index of an optical medium due to the presence of an acoustic wave. To describe the effect in crystals, we need to introduce the elasto-optic tensor based on Pockels' phenomenological theory.¹⁵

An elastic wave propagating in a crystalline medium is generally described by the strain tensor S , which is defined as the symmetric part of the deformation gradient

$$S_{ij} = \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) / 2 \quad i, j = 1 \text{ to } 3 \quad (1)$$

where u_i is the displacement. Since the strain tensor is symmetric, there are only six independent components. It is customary to express the strain tensor in the contracted notation

$$S_1 = S_{11} \quad S_2 = S_{22} \quad S_3 = S_{33} \quad S_4 = S_{23} \quad S_5 = S_{13} \quad S_6 = S_{12} \quad (2)$$

The conventional elasto-optic effect introduced by Pockels states that the change of the impermeability tensor ΔB_{ij} is linearly proportional to the symmetric strain tensor.

$$\Delta B_{ij} = p_{ijkl} S_{kl} \quad (3)$$

where p_{ijkl} is the elasto-optic tensor. In the contracted notation

$$\Delta B_m = p_{mn} S_n \quad m, n = 1 \text{ to } 6 \quad (4)$$

Most generally, there are 36 components. For the more common crystals of higher symmetry, only a few of the elasto-optic tensor components are nonzero.

In the above classical Pockels' theory, the elasto-optic effect is defined in terms of the change of the impermeability tensor ΔB_{ij} . In the more recent theoretical work on AO interactions, analysis of the elasto-optic effect has been more convenient in terms of the nonlinear polarization resulting from the change of dielectric tensor $\Delta \epsilon_{ij}$. We need to derive the proper relationship that connects the two formulations.

Given the inverse relationship of ϵ_{ij} and B_{ij} in a principal axis system $\Delta \epsilon_{ij}$ is

$$\Delta \epsilon_{ij} = -\epsilon_{ii} \Delta B_{ij} \epsilon_{jj} = -n_i^2 n_j^2 \Delta B_{ij} \quad (5)$$

where n_i is the refractive index. Substituting Eq. (3) into Eq. (5), we can write

$$\Delta \epsilon_{ij} = \chi_{ijkl} S_{kl} \quad (6)$$

where we have introduced the elasto-optic susceptibility tensor

$$\chi_{ijkl} = -n_i^2 n_j^2 p_{ijkl} \quad (7)$$

For completeness, two additional modifications of the basic elasto-optic effect are discussed as follows.

Roto-Optic Effect Nelson and Lax¹⁶ discovered that the classical formulation of elasto-optic effect was inadequate for birefringent crystals. They pointed out that there exists an additional roto-optic susceptibility due to the antisymmetric rotation part of the deformation gradient.

$$\Delta B'_{ij} = p'_{ijkl} R_{kl} \quad (8)$$

where $R_{ij} = (S_{ij} - S_{ji})/2$.

It turns out that the roto-optic tensor components can be predicted analytically. The coefficient of p'_{ijkl} is antisymmetric in kl and vanishes except for shear waves in birefringent crystals. In a uniaxial crystal the only nonvanishing components are $p_{2323} = p_{2313} = (n_o^{-2} - n_e^{-2})/2$, where n_o and n_e are the principal refractive indices for the ordinary and extraordinary wave, respectively. Thus, the roto-optic effect can be ignored except when the birefringence is large.

Indirect Elasto-Optic Effect In the piezoelectric crystal, an indirect elasto-optic effect occurs as the result of the piezoelectric effect and electro-optic effect in succession. The effective elasto-optic tensor for the indirect elasto-optic effect is given by¹⁷

$$p_{ij}^* = p_{ij} - \frac{r_{im} S_m e_{jn} S_n}{\epsilon_{mn} S_m S_n} \quad (9)$$

where p_{ij} is the direct elasto-optic tensor, r_{im} is the electro-optic tensor, e_{jn} is the piezoelectric tensor, ϵ_{mn} is the dielectric tensor, and S_m is the unit acoustic wavevector. The effective elasto-optic tensor thus depends on the direction of the acoustic mode. In most crystals the indirect effect is negligible. A notable exception is LiNbO_3 . For instance, along the z axis, $r_{33} = 31 \times 10^{-12}$ m/v, $e_{33} = 1.3$ c/m², $E_{33}^s = 29$, thus $p^* = 0.088$, which differs notably from the contribution $p_{33} = 0.248$.

Plane Wave Analysis of Acousto-Optic Interaction

We now consider the diffraction of light by acoustic waves in an optically transparent medium. As pointed out before, in the early development, the AO diffraction in isotropic media was described by a set of coupled wave equations known as the Raman-Nath equations.⁴ In this model, the incident light is assumed to be a plane wave of infinite extent. It is diffracted by a rectangular sound column into a number of plane waves propagating along different directions. Solution of the Raman-Nath equations gives the amplitudes of these various orders of diffracted optical waves.

In general, the Raman-Nath equations can be solved only numerically and judicious approximations are required to obtain analytic solutions. Using a numerical procedure computation Klein and Cook¹⁸ calculated the diffracted light intensities for various diffraction orders in this regime. Depending on the interaction length L relative to a characteristic length $L_o = n\Lambda^2/\lambda_o$, where n is the refractive index and Λ and λ_o are wavelengths of the acoustic and optical waves, respectively, solutions of the Raman-Nath equations can be classified into three different regimes.

In the Raman-Nath regime, where $L \ll L_o$, the AO diffraction appears as a large number of different orders. The diffraction is similar to that of a thin phase grating. The direction of the various diffraction orders are given by the familiar grating equation, $\sin \theta_m = m\lambda_o/n\Lambda$, where m is the diffraction order. Solution of the Raman-Nath equations shows that the amplitude of the m th-order diffracted light is proportional to the m th-order Bessel functions. The maximum intensity of the first-order diffracted light (relative to the incident light) is about 34 percent. Due to this relatively low efficiency, AO diffraction in the Raman-Nath regime is of little interest to practical device applications.

In the opposite limit, $L \gg L_o$, the AO diffraction appears as a predominant first order and is said to be in the Bragg regime. The effect is called Bragg diffraction since it is similar to that of the x-ray diffraction in crystals. In the Bragg regime the acoustic column is essentially a plane wave of infinite extent. An important feature of the Bragg diffraction is that the maximum first-order diffraction efficiency obtainable is 100 percent. Therefore, practically all of today's AO devices are designed to operate in the Bragg regime.

In the immediate case, $L \leq L_o$, the AO diffraction appears as a few dominant orders. This region is referred as the near Bragg region since the solutions can be explained based on the near field effect of the finite length and height of the acoustic transducer.

Many modern AO devices are based on the light diffraction in anisotropic media. The Raman-Nath equations are no longer adequate and a new formulation is required. We have previously presented a plane wave analysis of AO interaction in anisotropic media.¹³ The analysis was patterned after that of Klienman¹⁹ used in the theory of nonlinear optics. Unlike the Raman-Nath equations wherein the optical plane waves are diffracted by an acoustic column, the analysis assumes that the acoustic wave is also a plane wave of infinite extent. Results of the plane wave AO interaction in anisotropic media are summarized as follows.

The AO interaction can be viewed as a parametric process where the incident optical plane wave mixes with the acoustic wave to generate a number of polarization waves, which in turn generate new optical plane waves at various diffraction orders. Let the angular frequency and optical wavevector of the incident optical wave be denoted by ω_m and $\vec{k}_o$, respectively, and those of the acoustic waves by ω_a and $\vec{k}_a$. The polarization waves and the diffracted optical waves consist of waves with angular frequencies $\omega_m = \omega_o + m\omega_a$ and wavevectors $\vec{K}_m = \vec{k}_o + m\vec{k}_a$ ($m = \pm 1, \pm 2, \dots$). The diffracted optical waves are new normal modes of the interaction medium with the angular frequencies $\omega_m = \omega_o + m\omega_a$ and wavevectors $\vec{k}_m$ making angles θ_m with the z axis.

The total electric field of the incident and diffracted light be expanded in plane waves as

$$\vec{E}(r, t) = \frac{1}{2} \sum \hat{e}_m E_m(z) \exp j(\omega_m t - \vec{k}_m \cdot \vec{r}) + \text{c.c.} \quad (10)$$

where $\hat{e}_m$ is a unit vector of the electric field of the m th wave, E_m is the slowly varying amplitude of the electric field and c.c. stands for the complex conjugate. The electric field of the optical wave satisfies the wave equation,

$$\nabla \times \nabla \times \vec{E} + \frac{1}{c^2} \left(\vec{\epsilon} \cdot \frac{\partial^2 \vec{E}}{\partial t^2} \right) = -\mu_o \frac{\partial^2 \vec{P}}{\partial t^2} \quad (11)$$

where $\vec{\epsilon}$ is the relative dielectric tensor and $\vec{P}$ is the acoustically induced polarization. Based on Pockels' theory of the elasto-optic effect,

$$\vec{P}(r, t) = \epsilon_o \vec{\chi} \vec{S}(r, t) \vec{E}(r, t) \quad (12)$$

where $\vec{\chi}$ is the elasto-optical susceptibility tensor defined in Eq. (7). $\vec{S}(r, t)$ is the strain of the acoustic wave

$$\vec{S}(r, t) = 1/2 (\hat{s} \exp j(\omega_a t - \vec{k}_a \cdot \vec{r}) + \text{c.c.}) \quad (13)$$

where $\hat{s}$ is a unit strain tensor of the acoustic wave and S is the acoustic wave amplitude. Substituting Eqs. (10), (12), and (13) into Eq. (11) and neglecting the second-order derivatives of electric-field amplitudes, we obtain the coupled wave equations for AO Bragg diffraction.

$$\frac{dE_m}{dz} = \frac{j(\omega_o/c)^2}{4k_m \cos \gamma_m} (\chi_m S E_{m-1} e^{-j\Delta \vec{k}_m \cdot \vec{r}} + \chi_{m+1} S^* E_{m+1} e^{j\Delta \vec{k}_{m+1} \cdot \vec{r}}) \quad (14)$$

where $\chi_m = n_m^2 n_{m-1}^2 p_m$, $p_m = \hat{e}_m \cdot \vec{p} \cdot \vec{s} \cdot \hat{e}_{m-1}$, γ_m is the angle between $\vec{k}_m$ and the z axis, and $\Delta \vec{k}_m = \vec{K}_m - \vec{k}_m = \vec{k}_o + m\vec{k}_a - \vec{k}_m$ is the momentum mismatch between the optical polarization waves and m th-order normal modes of the medium. Equation (14) is the coupled wave equation describing the AO interaction in an anisotropic medium. Solution of the equation gives the field of the optical waves in various diffraction orders.

Two-Wave AO Interaction In the Bragg limit, the coupled wave equation reduces to the two-wave interaction between the incident and the first-order diffracted light ($m = 0, 1$):

$$\frac{dE_d}{dz} = \frac{j\pi n_i^2 n_d p_e}{2\lambda_o \cos \gamma_o} S E_i e^{j\Delta \vec{k} \cdot \vec{z}} \quad (15)$$

$$\frac{dE_i}{dz} = \frac{j\pi n_d^2 n_i p_e}{2\lambda_o \cos \gamma_o} S^* E_d e^{-j\Delta \vec{k} \cdot \vec{z}} \quad (16)$$

where n_i and n_d are the refractive indices for the incident and diffracted light, $p_e = \hat{e}_d \cdot \hat{p} \cdot \hat{s} \cdot \hat{e}_i$ is the effective elasto-optic constant for the particular mode of AO interaction, γ_o is the angle between the z axis and the median of incident and diffracted light and, $\Delta \vec{k} \cdot \vec{z} = \Delta k_z$ is the component of the momentum mismatch $\Delta \vec{k}$ along the z axis, and $\Delta \vec{k}$ is the momentum mismatch between the polarization wave $\vec{K}_d$ and the free wave $\vec{k}_d$ of the diffracted light.

$$\Delta \vec{k} = \vec{K}_d - \vec{k}_d = \vec{k}_i + \vec{k}_a - \vec{k}_d \quad (17)$$

Equations (15) and (16) admit simple analytic solutions. At $z = L$, the intensity of the first-order diffracted light (normalized to the incident light) is

$$I_1 = \frac{I_d(L)}{I_i(0)} = \eta \operatorname{sinc}^2 \frac{1}{\pi} \left(\eta + \left(\frac{\Delta k_z L}{2} \right)^2 \right)^{1/2} \quad (18)$$

where $\operatorname{sinc}(x) = (\sin \pi x)/\pi x$, and

$$\eta = \frac{\pi^2}{2\lambda_o^2} \left(\frac{n^6 p^2}{2} \right) S^2 L^2 = \frac{\pi^2}{2\lambda_o^2} M_2 P_a \left(\frac{L}{H} \right) \quad (19)$$

In the above equation, we have used the relation $P_a = \rho V^3 S^2 LH/2$, where P_a is the acoustic power, H is the acoustic beam height, ρ is the mass density, V is the acoustic wave velocity, and $M_2 = n^6 p^2 / \rho V^3$ is a material figure of merit.

Far Bragg Regime Equation (18) shows that in the far-field limit ($L \rightarrow \infty$) the diffracted light will build up finite amplitude only when the exact phase matching is met. When $\Delta k = 0$, the diffracted light intensity becomes

$$I_1 = \sin^2 \sqrt{\eta} \quad (20)$$

where $\eta \ll 1$, $I_1 \approx \eta$ and the diffraction efficiency is linearly proportional to acoustic power. This is referred to as the weak interaction approximation (or low efficiency regime). As acoustic power increases, the diffraction efficiency approaches 100 percent. However, the acoustic power required is about 2.5 times that predicted by the low efficiency regime.

Near Bragg Regime In the near field the growth of the diffraction is determined by the accumulated phase mismatch over the interaction length. The fractional diffracted light I_1 can be approximated by

$$I_1 = \eta \sin^2 \psi \quad (21)$$

where $\psi = \Delta k_z L / 2\pi$ is the phase mismatch (normalized to 2π). In the following, we shall discuss first AO diffraction in the far Bragg limit, that is, when exact phase matching is satisfied.

Phase Matching

Particle Picture of AO Diffraction In the far Bragg limit ($L \rightarrow \infty$), the diffracted light will build up finite amplitude only when the exact phase matching is met.

$$\vec{k}_d = \vec{k}_i \pm \vec{k}_a \quad (22)$$

In this limit case, the AO diffraction can be viewed as the interaction of optical and acoustic plane waves of infinite extent. To analyze AO interaction, it is more conveniently to use the particle picture of plane waves. The optical and acoustic plane waves can be thought of as made of photons and phonons of well-defined momentum and energy. Equation (22) simply states the principle of conservation of total momentum in the collision process. The principle of conservation of energy may be written as

$$\omega_d = \omega_i \pm \omega_a \quad (23)$$

The optical frequency of the diffracted light is shifted by the frequency of the acoustic wave. In analog to the Doppler effect, the optical frequency is up or down shifted if the direction of the acoustic wave is the same as or opposite to that of light wave.

In the following, we shall apply the conservation principles to derive the basic characteristics of light diffraction by acoustic wave in the plane wave formulation. For the general case of AO interaction in an anisotropic medium, the magnitudes of the incident optical wavevector can be written as

$$k_i = \frac{2\pi n_i}{n_o \lambda} \quad k_d = \frac{2\pi n_d}{n_o \lambda} \quad k_a = \frac{2\pi}{\Lambda} \quad (24)$$

where $\lambda = \lambda_o / n_o$ is the wavelength of the o-wave in the medium, n_i and n_d are refractive indices for the incident and diffracted light, $\Lambda = V/f$ is the acoustic wavelength and V and f are the acoustic velocity and frequency, respectively. Since the acoustic frequency is typically below 10^{-4} of the optical frequency, the small optical frequency difference (due to acoustic frequency shift) of the incident and diffracted light beams are neglected in Eq. (24).

Isotropic Diffraction Consider first the case of isotropic diffraction. Figure 1a shows the wavevector interaction geometry in the interaction plane. The loci of the incident and diffracted optical wavevectors fall on a circle of radius n_o . The principle of momentum conservation requires that the acoustic and optical wavevectors form a closed triangle. For isotropic AO diffraction the triangle is isosceles, and the incident and diffracted optical wavevectors make the same angle with the acoustic wavefront at the Bragg angle θ_b .

$$\sin \theta_b = \frac{\lambda}{2\Lambda} = \frac{\lambda_o f}{2n_o V} \quad (25)$$

Consider, for example, the diffraction of a laser beam at 633 nm by a longitudinal mode acoustic wave in TeO₂, $n = 2.216$, $V = 4.2$ mm/ μ sec, Eq. (25) yields $\theta_b = 1.9^\circ$ at an acoustic frequency $f = 500$ MHz. For all practical purpose the Bragg angle θ_b is small and the incident optical beam is always nearly

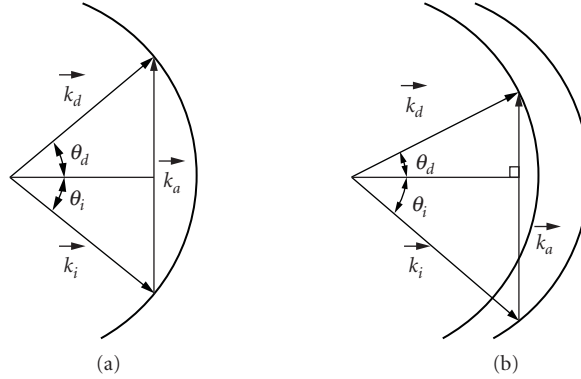


FIGURE 1 Vector diagram of acousto-optic interaction: (a) isotropic diffraction and (b) birefringent diffraction.

perpendicular to the acoustic wavefront. Thus, for an isotropic AO diffraction the primary effect of the acoustically driving polarization is to provide a transverse spatial light modulation (SLM). Similar to a grating, the transverse SLM acts as an acoustically driven optical beam deflector that scans the output light beam as the acoustic frequency is changed. The deflection angle θ_D , defined as the angle of separation between the incident and diffraction light, is given by (in the small-angle approximation)

$$\theta_D = \frac{\lambda}{\Lambda} = \frac{\lambda f}{V} \quad (26)$$

Anisotropic Diffraction Next, consider the AO diffraction in an optically anisotropic medium such as a birefringent crystal. In this case there are two distinct loci of the optical wavevectors (normal surfaces), an ordinary optical wave (polarized perpendicular to the c axis) and an extraordinary optical wave (polarized parallel to the c axis). The refraction indices are in general dependent on the propagation direction and polarization of the optical wave. The acoustic wave could couple the incident and diffracted optical wave of the same polarization ($o \leftrightarrow o$, $e \leftrightarrow e$) or orthogonal polarization ($o \leftrightarrow e$). In the latter case, the AO diffraction occurs between two loci of unequal refractive indices; this process is referred as the birefringent diffraction.

Phase Matching Equations Consider the AO interaction in a positive uniaxial crystal. Figure 1b shows the wavevector diagram for $o \leftrightarrow e$ type birefringent diffraction in the interaction plane. Applying the law of cosine to the triangle yields the following pair of equations for the incident and diffraction angle.

$$\sin \theta_i = \frac{\lambda_o}{2n_i \Lambda} \left[1 + \frac{\Lambda^2}{\lambda_o^2} (n_i^2 - n_d^2) \right] \quad (27)$$

$$\sin \theta_d = \frac{\lambda_o}{2n_d \Lambda} \left[1 - \frac{\Lambda^2}{\lambda_o^2} (n_i^2 - n_d^2) \right] \quad (28)$$

where n_i and n_d are the refractive indices for the incident and diffracted light, respectively. Equations (27) and (28) were first derived by Dixon⁸ for AO diffraction in a uniaxial crystal wherein the interaction plane is perpendicular to the c axis. The wavevector loci for the ordinary polarized light (o -wave) and the extraordinary polarized light (e -wave) are concentric circles; thus the two

equations can be separately solved. In general the refractive index for an extraordinary polarized light (e -wave) is a function of the polar angle of the incident and diffracted light. Thus, Dixon's equations are not applicable in practice since Eqs. (27) and (28) are coupled.

In order to decouple the input and diffracted angles from this pair of equations, we have chosen to express the wavevectors in the exact momentum matching condition using an elliptical parametric representation. Since the polarization is switched in the birefringent diffraction, the output characteristics of the diffracted light are same as the input characteristics of the incident light with the orthogonal polarization. Therefore, we need to consider only the input frequency-angle characteristics of the birefringent diffraction. The phase matching equations may be written in a form similar to the Dixon equations

$$e \rightarrow o: \quad \hat{s}_e \cdot \hat{s}_a = \frac{\lambda}{2\mu_e \Lambda} \left[1 + \frac{\Lambda^2}{\lambda^2} (\mu_e^2 - 1) \right] \quad (29)$$

$$o \rightarrow e: \quad \vec{\sigma}_o \cdot \vec{\sigma}_a = \frac{\lambda}{2\Lambda} \left[\vec{\sigma}_a^2 + \frac{\Lambda^2}{\lambda^2} (\vec{\sigma}_o^2 - 1) \right] \quad (30)$$

where

$$\begin{aligned} \hat{s}_a &= (\cos\theta_a, \sin\theta_a) & \hat{s}_e &= (\cos\theta_e, \sin\theta_e) & \mu_e &= (\cos^2\theta_e + e^{-2}\sin^2\theta_e)^{-1/2} \\ \vec{\sigma}_a &= (\cos\theta_a, e^{-1}\sin\theta_a) & \vec{\sigma}_o &= (\cos\theta_o, e^{-1}\sin\theta_o) \end{aligned} \quad (31)$$

where $e = n_e/n_o$ is the ratio of the principal refractive indices, θ_a , θ_o , and θ_e are the polar angles of the acoustic wave, ordinary, and extraordinary optical wave, respectively. The phase matching Eqs. (29) and (30) are the basic equations for AO diffraction in the Bragg regime.

To proceed with the analysis we introduce a dimensionless parameter a , defined as the ratio of optical wavelength (inside the medium) to the acoustic wavelength, $a = \lambda/\Lambda$. In terms of the wavelength ratio Eqs. (29) and (30) can be written in the form of quadratic equations.

$$q_{eo}(a) = a^2 - 2a\mu_e(\hat{s}_e \cdot \hat{s}_a) + \mu_e^2 - 1 = 0 \quad (32)$$

$$q_{oe}(a) = \vec{\sigma}_a^2 a^2 - 2a(\vec{\sigma}_a \cdot \vec{\sigma}_o) + \vec{\sigma}_o^2 - 1 = 0 \quad (33)$$

Equations (32) and (33) may be considered the dispersion relations for the acousto-optic "grating" for an extraordinary and ordinary light input, respectively. Consider, for example, the diffraction of an extraordinary wave into an ordinary wave in a uniaxial crystal. Solving the quadratic equation [Eq. (32)] yields the wavelength ratio a for the $e \rightarrow o$ type diffraction

$$a = \mu_e \left\{ \sin\theta_i \pm \sqrt{\sin^2\theta_i - (1 - \mu_e^{-2})} \right\} \quad (34)$$

where $\theta_i = \theta_e - \theta_a - \pi/2$ is the incidence angle. Equation (34) admits two real roots if $|\sin\theta_i| \geq \sqrt{1 - \mu_e^{-2}}$. Depending on the separation of the two roots, the AO diffraction can be classified into two regimes; referred as transverse and longitudinal spatial modulation.²⁰

Transverse Spatial Modulation By selecting the incidence angle near its critical value, the two roots are close to each other. This happens when direction of the acoustic wave is nearly perpendicular to that of the optical waves. The acoustically driven polarization provides a transverse spatial modulation (TSM) to the incident light beam. Similar to isotropic AO diffraction, the TSM acts as an optical beam deflector.

The characteristics birefringent diffraction can be utilized to provide significant performance advantages for the AO deflector. For instance, when the light beam is chosen to incident at the critical angle, Eq. (32) yields two equal roots and the acoustic roots. The AO diffraction is said to operate at the tangential phase matching (TPM) since acoustic wavevector is tangential to the locus of the diffracted wavevector. At TPM the AO deflector acquires a wideband frequency characteristics since a range of acoustic frequencies nearly satisfy the exact phase matching condition. The optic and acoustic wavelengths at TPM are related by

$$\lambda_o = \Lambda_t \sqrt{\mu_e^2 - 1} \approx \Lambda_t \sqrt{2n_o \Delta n} \sin \theta_e \quad (35)$$

where $\Delta n = n_e(e - 1)$ is the birefringence. At $\theta_e = 90^\circ$, the acoustic frequency reaches a maximum value.

$$\lambda_o \approx \Lambda_t \sqrt{2n_o \Delta n} \quad (36)$$

Longitudinal Spatial Modulation By selecting the optical incidence much larger than the critical angle, the two terms in Eq. (34) are about equal. Only the solution that corresponds to the difference of the two terms yields a low acoustic frequency near its minimum value. In this case the acoustic wavevector has a larger component along the direction of the optical wave, the AO diffraction thus provides a longitudinal spatial modulation (LSM) to the incident light. Unlike the previous case the LSM acts as an acoustically tuned optical bandpass filter. When the acoustic frequency is changed, the passband wavelength is tuned accordingly while the diffracted optical beam remains at a fixed deflection. This LSM-based AO device forms the basis of the acousto-optic tunable filter (AOTF).

Since the wavelength ratio is small, in this case the second order in Eq. (32) or (33) can be neglected. For an incident light with either polarization, an approximate solution of the center wavelength the AOTF passband is

$$\lambda_o \approx \frac{\Lambda \Delta n \sin^2 \theta_o}{\cos(\theta_a - \theta_o)} \quad (37)$$

where $\Delta n = n_e - n_o$ is the birefringence. At $\theta_a = 90^\circ$, the incident and diffracted optical waves are collinear with the acoustic wave, Eq. (37) reduced to

$$\lambda_o = \Lambda \Delta n \quad (38)$$

Equation (38) is the momentum-matching condition for the collinear AOTF.

Frequency Characteristics of AO Interaction

In the near-Bragg regime, the AO diffraction is determined by the z -component of phase mismatch function $\psi = \Delta k_z L / 2\pi = \Delta \sigma_z L$. In our analysis, this z -component mismatch is caused by the deviation of the light incidence from exact momentum matching. Based on this model it can be shown that the momentum mismatch (normalized to 2π) is $\Delta \sigma_z = (n_o/2)q$, where q is defined in Eqs. (32) and (33) for $e \rightarrow o$ and $o \rightarrow e$ type diffraction, respectively. Substituting $\psi = L n_o q / 2$ from Eq. (32) or (33) into Eq. (21), the bandpass function of AO diffraction can be expressed as a function of the variation of the acoustic angle θ_a and acoustic frequency f . It is convenient to normalize the acoustic frequency to a center frequency f_o (wavelength Λ_o). In terms of the normalized acoustic frequency ($F = f/f_o$), the phase mismatch ψ can be written as

$$\psi = \frac{l}{2}(F^2 - FF_m + F_t^2) \quad (39)$$

where

$$l = L/L_o \quad L_o = n_o \Lambda^2 / \lambda_o \quad F_m = 2(n_o \Lambda_o / \lambda_o) \mu_e \sin \theta_i \quad F_t = f_t / f_o \quad (40)$$

Isotropic Diffraction Bandshape By choosing $F_m = 1 + (\Delta F/2)^2$, the bandshape function $W(\psi)$ has equal loss at the two ends of the passband $F_t = 1 + (\Delta F/2)$, where ΔF is the fractional bandwidth of the AO interaction. The diffraction efficiency reduces to 0.5, where $\psi = 0.45$. This corresponds to a fractional bandwidth

$$\Delta F = 1.8/l \quad (41)$$

To realize octave bandwidth, ($\Delta F = 2/3$) for instance, the normalized interaction length l is equal to 2.7. Figure 2a shows the bandshape of isotropic AO diffraction.

Birefringent Diffraction Bandshape For $e \rightarrow o$ diffraction, it is possible to obtain a wide bandpass response by operating near tangential phase matching (TPM). At TPM the two

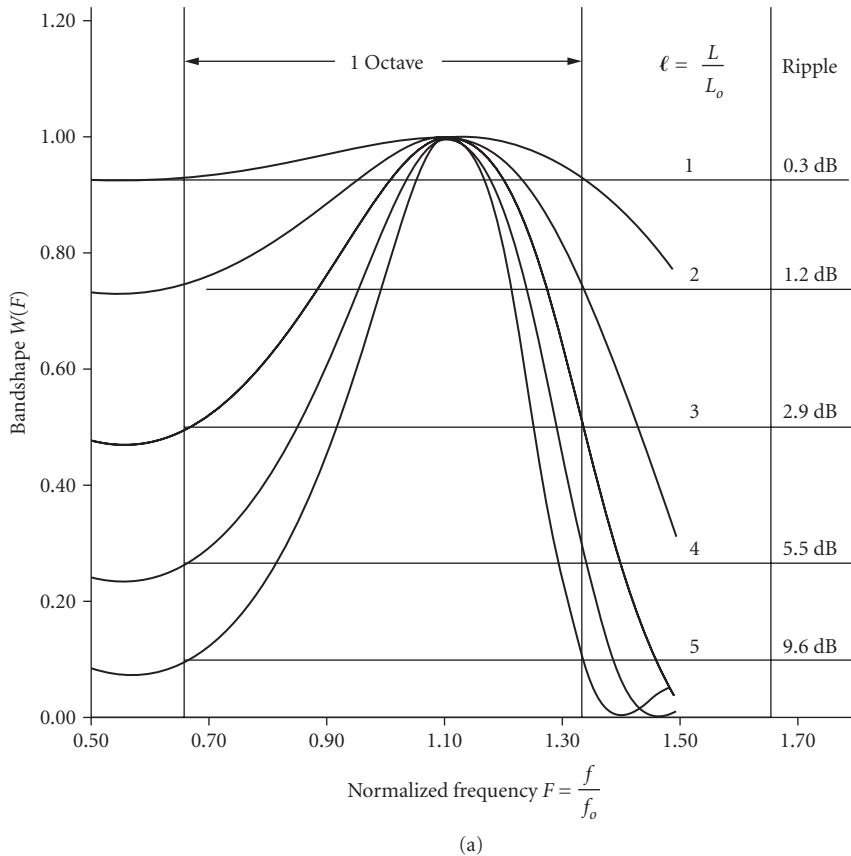


FIGURE 2 Acousto-optic bandshapes: (a) isotropic diffraction and (b) birefringent diffraction.

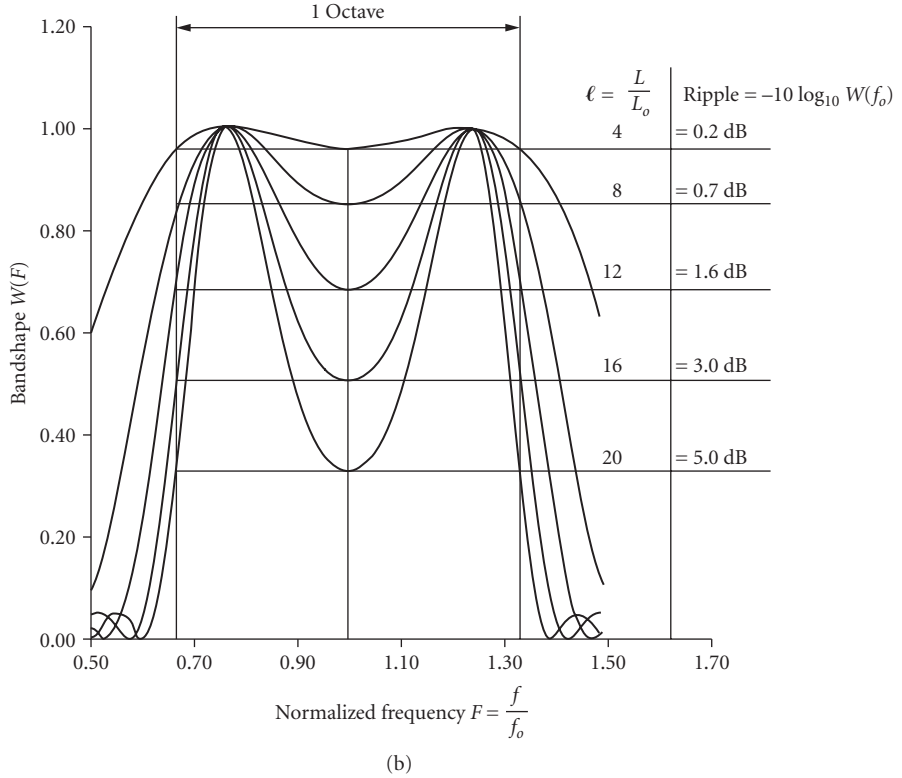


FIGURE 2 (Continued)

phase-matching frequencies coincide and the AO bandpass exhibits a flat-top shape with a bandwidth

$$\Delta F = \frac{\Delta f}{f} \approx \left(\frac{3.6}{l} \right)^{1/2} \quad (42)$$

For octave bandwidth ($\Delta F = 2/3$) the normalized interaction length l is equal to 8.1. This represents an efficiency advantage factor of 3 compared to isotropic AO diffraction. Figure 2b shows the bandshape of the birefringent AO diffraction.

Acousto-Optic Interaction of Finite Geometry

The plane wave analysis of AO interaction in the Bragg regime appears to be inadequate for real devices since the interaction geometry usually involves optical and acoustic beams of finite sizes with nonuniform amplitude distributions, for example, a Gaussian optical beam and a divergent acoustic beam. Gordon⁶ considered the AO diffraction in finite interaction geometry using the Green's function solution for the wave equation. To make the approach analytically tractable, the first-order Born's approximation is used, which is same as the weak interaction assumption.

Instead of the Green's function solution, an equivalent approach is to decompose the optical and acoustic waves into angular spectrum of plane waves (ASW) and add all the spatial frequency components of the

diffracted light that satisfy the exact phase-matching condition.^{21,22} Since the AO interaction is modeled as a filtering process in the spatial frequency domain, this approach is referred to as the frequency domain analysis.

As an example we use the frequency domain approach to determine the frequency and angular characteristics of an AO deflector. To simplify the analysis we will make a two-dimensional model for AO interaction in the interaction (x - z) plane. The reduction of diffraction efficiency in the height (y - z) plane will be calculated in the real domain as an overlapping integral of the acoustic and optical beams in the transverse plane along the height direction.

In the case of weak interaction, there is negligible depletion of the incident light. Assume that the incident light amplitude is a constant, Eq. (15) can be integrated to yield

$$E_d(\vec{\sigma}) = j\kappa S(\vec{\sigma}_a) E_i(\vec{\sigma}_i) \delta(\vec{\sigma} - \vec{\sigma}_i - \vec{k}_a) = j\kappa S(\vec{\sigma}_a) E_i(\vec{\sigma} - \vec{\sigma}_a) \quad (43)$$

where $\vec{\sigma}$, $\vec{\sigma}_i$, and $\vec{\sigma}_a$ are the wavevector components in the interaction plane for the diffracted light, incident light and the acoustic waves, respectively.

We assume that the acoustic beam consists of acoustic plane waves propagating along the x axis in the interaction (x - z) plane. It is generated by an acoustic transducer with a top electrode of effective length L along the z axis and height H along the y axis. We also assume that the incident light beam consists of optical plane waves propagating in the x - z plane at an angle of incidence γ_i with the z axis. The incident light beam has a Gaussian with a beam waist $2\omega_1$ and $2\omega_2$ at e^{-2} of the intensity profile in the interaction and height plane, respectively. By expanding the optical and acoustic field in terms of their spatial frequency components in the interaction plane and applying Eq. (43), the total power spectrum of the diffracted light intensity can be written as

$$I_d(F, \xi) = \eta W(F) U(\xi) \cdot J \quad (44)$$

where η is the small signal diffraction efficiency given by Eq. (19). The next three multiplying factors are briefly discussed as follows.

Overlapping Integral The overlapping integral J of optical and acoustic profile in the height plane is given by

$$J = \frac{2}{\pi\omega_1\omega_2} \int_{-H/2}^D \int_{-H/2}^{H/2} |V(x, y)|^2 e^{-y^2/\omega_2^2} e^{-(x-D/2)^2/\omega_1^2} dy dx \quad (45)$$

where $V(x, y)$ is the acoustic field distribution and can be determined by the Fresnel integral

$$V(x, y) = \int_{-H/2}^{H/2} \frac{1}{\sqrt{jB\Lambda x}} \left[\exp\left(\frac{j\pi(\gamma_o - y)^2}{B\Lambda x}\right) \right] dy_o \quad (46)$$

where B is the curvature of the acoustic slowness surface in the transverse plane. The overlapping integral represents a reduction of diffraction efficiency of the AO diffraction due to acoustic diffraction in the transverse plane along the height direction. For a transducer of height H , the acoustic radiation in a region near the transducer is approximately collimated. This collimated region, known as the near field, extends a distance D from transducer and is given by

$$H_o = \sqrt{B\Lambda D} = V \sqrt{\tau B/f} \quad (47)$$

Numerical calculation of the overlapping integral J in Eq. (43) shows its value is on the order of unity if the transducer height H is chosen to be equal to H_o given by Eq. (47).

Bandpass Response The bandpass function $W(F)$ as a function of the normalized frequency F is equal to the magnitude square of the Fourier transform, that is, the power spectrum of the acoustic

field along the z axis. For a standard rectangular window of the transducer field, the bandpass function is same as that given by

$$W(\psi) = \text{sinc}^2(\psi) \quad (48)$$

where $\psi = \Delta\sigma_z L = (n_o/2)q$ is the phase mismatch along optical direction (z axis) and is given by Eq. (39) as a function of the normalized frequency F .

Spatial Frequency Spectrum The spatial frequency spectrum $U(\xi)$ expressed as a function of spatial frequency component ξ signal direction (x axis) is determined by the power spectrum of the incident optical beam profile along the x axis, the direction perpendicular to the optical direction. For input Gaussian light with a beam waist $2\omega_1$ along the z axis in the interaction plane, the spatial frequency spectrum $U(\xi)$ is given by

$$U(\xi) = \exp -(\Delta\sigma_x^2 / \pi^2 \omega_1^2) \quad (49)$$

where $\Delta\sigma_x = \xi - \sigma_i \sin \theta_i - \sigma_o$, ξ is the spatial frequency of the diffracted light along x axis.

6.4 ACOUSTO-OPTIC MATERIALS

The significant progress of AO devices has been due largely to the development of superior materials such as TeO_2 and GaP. In this section we shall review the material issues related to AO device applications. A comprehensive list of tables summarizing the properties of AO materials is presented at the end of this section. The AO material issues and the selection guidelines have been discussed in a previous publication.¹²

The selection of AO materials depends on the specific device application.²³ An AO material suited for one type of device may not even be applicable for another. For example, GaP is perhaps the best choice for making wideband AO deflectors or modulators. However, since GaP is optically isotropic, it cannot be used for tunable filter purposes.

Some of the requirements for materials' properties apply to the more general cases of optical device applications, for example, high optical transparency over wavelength range of interest, availability in large single crystals, and the like. We shall restrict the discussion to material properties that are particularly required for AO device applications

Acousto-Optic Figures of Merit

A large AO figure of merit is desired for device applications. There are several AO figures of merit that have been used for judging the usefulness of an AO material. The relevant one to be used depends on the specific applications. Several AO figures of merit are defined in the literature. These include

$$M_1 = \frac{n^7 p^2}{\rho V} \quad M_2 = \frac{n^6 p^2}{\rho V^3} \quad M_3 = \frac{n^7 p^2}{\rho V^2} \quad M_4 = \frac{n^8 p^2 V}{\rho} \quad M_5 = \frac{n^8 p^2}{\rho V^3} \quad (50)$$

where n is the index of refraction, p is the relevant elasto-optic coefficient, ρ is the density, and V is the acoustic wave velocity. These figures of merit are generally listed as the normalized quantities M (normalized to values for fused silica).

The figure of merit M_1 relates the diffraction efficiency η , to the acoustic power P_a for a given device aspect ratio L/H in accordance with Eq. (18). M_2 is the AO figure of merit most often referred to in the literature and is widely used for the comparison of AO materials. This is a misconception, since from the viewpoint of device applications, M_2 is usually not appropriate. Comparison of AO materials (or modes) based on M_2 can lead to erroneous conclusions. M_2 is used only when efficiency is the only parameter of

concern. In most practical cases, other parameters such as bandwidth and resolution must also be considered. To optimize the efficiency bandwidth product, the relevant figure of merit is $M_1 = nV^2M_2$.

In the design of AO deflectors, or Bragg cells, besides efficiency and bandwidth, a third parameter of interest is the aperture time τ . A minimum acoustic beam height H must be chosen to ensure that the aperture is within the near field of the acoustic radiation. With choice, the relevant figure of merit for optimized efficiency for a specified bandwidth and aperture time is $M_3 = nVM_2$.

For wideband AO modulators, the acoustic power density P_d is often the limiting factor. The appropriate AO figure of merit is then M_4 , that is, $M_4 = nV^2M_1$.

In the design of AO tunable filters, the parameters to be optimized are the product of efficiency η , the resolving power $\lambda_o/\Delta\lambda$, and the solid angular aperture $\Delta\Omega$. In this case the appropriate AO figure of merit is $M_5 = n^2M_2$.

Acoustic Propagation and Attenuation

The performance of AO devices also depends on the acoustic properties of the interaction medium. As seen from the figures of merit listed above, a low acoustic velocity is one of the most important desired material parameters for AO devices. It not only provides the advantage of lower drive power, but it also allows the realization of large resolution for AO deflectors and tunable filters for a specified maximum crystal length.

The anisotropic acoustic propagation characteristic also plays an important role in the performance of AO devices. For instance, the small curvature of slowness surface of the transverse plane in GaP allows a smaller transducer height and greatly lowers the drive power. The use of large acoustic beam walk-off in TeO_2 allows the realization of a preferred interaction geometry to extend interaction length or optical aperture.

A low acoustic attenuation is also desired for increased resolution of deflectors or aperture of tunable filters. Generally, at room temperature $\omega\tau_{\text{th}} \ll 1$, where ω is the angular frequency of the acoustic wave and τ_{th} is the thermal phonon relaxation, the dominant contribution to acoustic attenuation is due to Akhieser loss caused by relaxation of the thermal phonon distribution toward equilibrium. A widely used result of this theory is the relation derived by Woodruff and Erhenrich. It states that acoustic attenuation measured in nepers per unit time is given by¹² $\alpha = \gamma^2 f^2 \kappa T / \rho V^4$, where γ is the Grüneisen constant, T is the temperature, and κ is the thermal conductivity. From this relation, it is seen that the AO figure of merit and the acoustic attenuation approximately follow an inverse relation. Notice also that the acoustic attenuation has a quadratic frequency-dependence for most crystals. In practice, in some crystals (such as GaP), it has been found that the frequency dependence of attenuation $\alpha \sim f^n$, where n varies from 1 to 2 as the frequency increases ranges. The deviation from a quadratic dependence may be attributed to the additional extrinsic attenuation caused by scattering from lattice imperfections.

Optical Birefringence

Optical birefringence is a requirement for materials used in AO tunable filters. The requirement is met by optically birefringent crystals so that the AO filter interaction can operate as a longitudinal spatial modulator (LSM). For AO deflectors and modulators, optical birefringence is not necessary. AO devices with high efficiency, wide bandwidth, and large resolution are realizable with superior isotropic materials such as GaP. However, in an optically birefringent crystal it is possible to operate at tangential phase matching (TPM), which provides an enhancement of interaction length l for a given fractional bandwidth. The optical birefringence in LiNbO_3 and TeO_2 has been largely responsible for the superior device performance.

Tabulation of Acousto-Optic Material Properties

To aid in the selection of AO materials, the relevant properties of some promising materials are listed below. Table 1 lists the values of elasto-optical tensor components. Table 2, taken from Ref. 24

TABLE 1 Elasto-Optic Coefficients of Materials

(a) Isotropic				
Material	λ (μm)	P_{11}	P_{12}	
Fused silica (SiO ₂)	0.63	+0.121	+0.270	
As ₂ S ₃ glass	1.15	+0.308	+0.299	
Water	0.63	0.31	0.31	
Ge ₃₃ Se ₅₅ As ₁₂ (glass)	1.06	0.21	0.21	
Lucite	0.63	0.30	0.28	
Polystyrene	0.63	0.30	0.31	
SF-59	0.63	0.27	0.24	
SF-8	0.63	0.198	0.262	
Tellurite glass	0.63	0.257	0.241	
(b) Cubic classes $\bar{4}3m$, 432, and m3m				
Material	λ (μm)	P_{11}	P_{12}	P_{44}
CdTe	10.60	-0.152	-0.017	-0.057
GaAs	1.15	-0.165	-0.140	-0.072
GaP	0.633	-0.151	-0.082	-0.074
Ge	2.0–2.2	-0.063	-0.0535	-0.074
	10.60	0.151	0.124	0.02
InP	1.5	0.18	0.06	0.06
NaCl	0.55–0.65	0.115	0.159	-0.011
NaF	0.633	0.08	0.20	-0.03
	0.589		-0.021	-0.10
Si	1.15	-0.101	0.0094	
	3.39	-0.094	0.017	-0.051
Y ₃ Fe ₅ O ₁₂ (YIG)	1.15	0.025	0.073	0.041
Y ₃ Al ₅ O ₁₂ (YAG)	0.633	-0.029	+0.0091	-0.0615
KRS5	0.633	0.18	0.27	± 0.15
KRS6	0.633	0.28	0.25	± 0.14
β -ZnS	0.633	0.091	-0.01	0.075
Y ₃ Ga ₅ O ₁₂	0.63	0.091	0.019	0.079
Diamond	0.59	-0.31	-0.03	
	0.59	-0.43	+0.19	-0.16
LiF	0.59	+0.02	+0.128	-0.064
MgO	0.59	-0.32	-0.08	
KBr	0.59	+0.22	+0.71	-0.026
KCl	0.59	+0.17	+0.124	
KI	0.59	+0.210	0.169	

(c) Cubic classes 23 and m3

Material	λ (μm)	P_{11}	P_{12}	P_{13}	P_{44}
Ba(NO ₃) ₂	0.63	0.15	0.35	0.29	0.02
NaBrO ₃	0.59	0.185	0.218	0.213	-0.0139
NaClO ₃	0.59	0.162	0.24	0.2	-0.198
BA(NO ₃) ₂	0.63	0.15	0.35	0.29	0.02
Bi ₁₂ GeO ₂₀	0.63		0.12		0.04
Bi ₁₂ SiO ₂₀	0.63		0.13		0.04

(d) Hexagonal system: classes 6m2, 6mm, 622, and 6/mm

Material	λ (μm)	P_{11}	P_{12}	P_{13}	P_{31}	P_{33}	P_{44}
CdS	0.63	-0.142	-0.066	-0.057	-0.041	-0.20	± 0.054
	10.60	0.104		0.011			
SnO	0.63	0.222	0.099	-0.111	0.088	-0.235	-0.0585
α -ZnS	0.63	-0.115	0.017	0.025	0.0271	-0.13	-0.0627
ZnO	0.63	0.222	0.199	-0.111	0.088	-0.235	-0.061

(e) Hexagonal system: classes 6, $\bar{6}$, and 6/m

Substance	λ (μm)	P_{11}	P_{12}	P_{13}	P_{31}	P_{33}	P_4	P_{45}	P_{16}
LiIO ₃	0.63	0.32			0.31				0.03

(f) Trigonal system: classes 3m, $\bar{3}2$, and $\bar{3}m$

Substance	λ (μm)	P_{11}	P_{12}	P_{13}	P_{14}	P_{31}	P_{33}	P_{41}	P_{44}
Al ₂ O ₃	0.644	-0.23	-0.03	0.02	0.00	-0.04	-0.20	0.01	-0.10
LiNbO ₃	0.633	-0.026	0.090	0.133	-0.075	0.179	0.071	-0.151	0.146
LiTaO ₃	0.633	-0.081	0.081	0.093	-0.026	0.089	-0.044	-0.085	0.028
SiO ₂ (quartz)	0.589	0.16	0.27	0.27	-0.030	0.29	0.10	-0.047	-0.079
Ag ₃ AsS ₃	0.633	± 0.10	± 0.19	± 0.22		± 0.24	± 0.20		
(proustite)	1.15	± 0.056	± 0.082	± 0.068		± 0.103	± 0.100	± 0.01	
Te	10.6	0.164	0.138	0.146		0.086	0.038		
CaCO ₃		± 0.095	± 0.189	± 0.215	-0.006	0.309	+0.178	+0.01	-0.090
HgS	0.63			0.445			0.115		
Tl ₃ AsSe ₃	3.39	0.4	0.22	0.24	0.04	0.2	0.22	0.018	0.15
Tl ₃ AsS ₃	3.39	0.36	0.13	0.2		0.15	0.36	0.02	

(Continued)

TABLE 1 Elasto-Optic Coefficients of Materials (*Continued*)

(g) Tetragonal system: classes 4mm, $\bar{4}2m$, 422, and 4/m													
Substance	λ (μm)	P_{11}	P_{12}	P_{13}	P_{31}	P_{33}	P_{44}	P_{66}					
(NH ₄)H ₂ PO ₄ (ADP)	0.589	0.319	0.277	0.169	0.197	0.167	−0.058	−0.091					
	0.63	0.296	0.243	0.208	0.188	0.228							
KH ₂ PO ₄ (KDP)	0.589	0.287	0.282	0.174	0.241	0.122	−0.019	−0.064					
	0.63	0.254	0.230	0.233	0.221	0.212							
Sr _{0.75} Ba _{0.25} Nb ₂ O ₆	0.63	0.16	0.10	0.08	0.11	0.47							
Sr _{0.5} Ba _{0.5} Nb ₂ O ₆	0.63	0.06	0.08	0.17	0.09	0.23							
TeO ₂	0.63	0.0074	0.187	0.340	0.0905	0.240	0.04	−0.0463					
TiO ₂ (rutile)	0.514	−0.001	0.113	−0.167	−0.106	−0.064	0.0095	−0.066					
	0.63	−0.011	0.172	−0.168	−0.0965	−0.058	±0.072						
ZrSiO ₄	0.63	0.06		0.13	0.07	0.09							
CdGeP ₂	0.63	0.21	−0.09	0.09		0.4	0.1	0.12					
Hg ₂ Cl ₂	0.63	0.551	0.44	0.256	0.137	0.01	0.047						
Hg ₂ Br ₂	0.63	0.262	0.175	0.148	0.177	0.116							
(h) Tetragonal system: classes 4, $\bar{4}$, 422, and 4/m													
Material	λ (μm)	P_{11}	P_{12}	P_{13}	P_{16}	P_{31}	P_{33}	P_{44}	P_{45}	P_{61}	P_{66}		
PbMoO ₄	0.63	0.24	0.24	0.255	0.017	0.175	0.3	0.067	−0.01	0.013	0.05		
CdMoO ₄	0.63	0.12	0.10	0.13	0.11	0.18							
NaBiMoO ₄	0.63	0.243	0.265	0.25		0.21	0.29						
LiBiMoO ₄	0.63	0.265	0.201	0.244		0.227	0.309						
CaMoO ₄	0.63	0.17	−0.15	−0.08	0.03	0.10	0.08	0.06	0.06	0.1	0.026		
(i) Orthorhombic all classes													
Material	λ (μm)	P_{11}	P_{12}	P_{13}	P_{21}	P_{22}	P_{23}	P_{31}	P_{32}	P_{33}	P_{44}	P_{55}	P_{66}
α -HIO ₃	0.63	0.406	0.277	0.304	0.279	0.343	0.305	0.503	0.310	0.334	0.092		
PbCO ₃	0.63	0.15	0.12	0.16	0.05	0.06	0.21	0.14	0.16	0.12			
BaSO ₄	0.59	0.21	0.25	0.16	0.34	0.24	0.19	0.27	0.22	0.31	0.22	−0.012	0.037
Gd ₂ (MoO ₄) ₃	0.63	0.19	0.31	0.175	0.215	0.235	0.175	0.185	0.23	0.115	−0.033	−0.028	0.035

TABLE 2 Selected AO Materials

Material	Optical Transmission (μm)	Density (g/cm^3)	Acoustic Mode	Acoustic Velocity ($\text{mm}/\mu\text{s}$)	Acoustic Attenuation ($\text{dB}/\mu\text{s-GHz}^2$)	Optical Polarization*	Refraction Index	Figure of Merit		
								M_1	M_2	M_3
Fused silica	0.2–4.5	2.2	L	5.96	7.2		1.46	1.0	1.0	1.0
LiNbO_3	0.4–4.5	4.64	L[100]	6.57	1.0	35° Y Rot.	2.2	8.5	4.6	7.7
	0.4–4.5	0.4–4.5	S(100)35°	~3.6	~1.0	[100]	2.2	2.3	4.2	3.8
TiO_2	0.45–6.0	4.23	L[110]	7.93	~1.0		2.58	18.6	6.0	14
PbMoO_4	0.42–5.5	6.95	L[001]	3.63	5.5	—	2.39	14.6	23.9	24
BGO	0.45–7.5	9.22	L[110]	3.42	1.6	Arb.	2.55	3.8	6.7	6.7
TeO_2	0.35–5.0	6.0	L[001]	4.2	6.3	—	2.26	17.6	22.9	25
	0.4–4.5	6.0	S[110]	0.62	17.9	Cir.	2.26	13.1	795	127
GaP	0.6–10.0	4.13	L[110]	6.32	8.0	—	3.31	75.3	29.5	71
	0.6–10.0	4.13	S[110]	4.13	2.0	—	3.31		16.6	26
Ti_3AsS_4	0.6–12.0	6.2	L[001]	2.15	5.0	—	2.83	152	523	416
Ti_3AsSe_3	1.26–13.0	7.83	L[100]	2.05	14.0	—	3.34	607	2259	1772
Hg_2Cl_2	0.38–28.0	7.18	L[100]	1.62	—	—	2.62	34	337	125
			S[110]	0.347	8.0		2.27	4.3	703	73
$\text{Ge}_{33}\text{As}_{12}\text{Se}_{33}$	1.0–14.0	4.4	L	2.52	1.7	—	2.7	54.4	164	129
GaAs	1.0–11.0	5.34	L[110]	5.15	15.5	—	3.37	118	69	137
As_2Se_3	0.9–11.0	4.64	L	2.25	27.5	—	2.89	204	722	539
Ge	2.0–20.0	5.33	L[111]	5.5	16.5	—	4.0	278	120	302

L: longitudinal; S: shear; arb.: arbitrary; cir.: circular birefringence.

*Optical polarization is normal or in the interaction plane.

lists the relevant properties of selected AO materials. The listed figures of merit M_1 , M_2 , and M_3 are normalized relative to that of fused silica, which has the following absolute values:

$$\begin{aligned} M_1 &= 7.83 \times 10^{-7} & [\text{cm}^2\text{s}g^{-1}] \\ M_2 &= 1.51 \times 10^{-18} & [\text{s}^3g^{-1}] \\ M_3 &= 1.3 \times 10^{-12} & [\text{cm}^2\text{s}^2g^{-1}] \end{aligned}$$

6.5 ACOUSTO-OPTIC DEFLECTOR

Acousto-optic interaction provides a simple means to deflect an optical beam in a sequential or random-access manner. As the driving frequency of the acoustic wave is changed, the direction of the diffracted beam can also be varied. The angle between the first-order diffracted beam and the incident beam for a frequency range Δf is approximately given by

$$\Delta\theta_d = \frac{\lambda\Delta f}{V} \quad (51)$$

In a deflector, the most important performance parameters are resolution and speed. The resolution, or the maximum number of resolvable spots, is defined as the ratio of the range of deflection angle divided by the angular spread of the diffracted beam, that is,

$$N = \frac{\Delta\theta_d}{\delta\theta_o} \quad (52)$$

The angular divergence of the incident optical beam is given by

$$\delta\theta_o = \xi \frac{\lambda}{D} \quad (53)$$

where D is the width of the incident beam and ξ is the a factor (near unity) that depends on the incident beam's amplitude distribution. For a nontruncated Gaussian beam $\xi = 4/\pi$. From Eqs. (51), (52), and (53) it follows

$$N \approx \tau \Delta f \quad (54)$$

where $\tau = D/V \cos \theta_o$ is the acoustic-transit time across the optical aperture. Notice that the acoustic-transit time also represents the (random) access time and is a measure of the speed of the deflector. Equation (54) shows that the resolution is equal to time (aperture) bandwidth product. This is the basic tradeoff relation between resolution and speed (or bandwidth) of AO deflectors.

It is instructive to specify the interaction geometry for AO deflectors using a dimensionless parameter A , the ratio of divergence angle between the optical and acoustic beams. For an acoustic wave generated from a flat transducer of width L , the acoustic beam divergence is given by

$$\delta\theta_a = \frac{\Lambda}{L} \quad (55)$$

TABLE 3 Interaction Geometry and Performance of AO Deflectors

Type I-Laser beam scanner
Regime of AO diffraction: Transverse SLM
Divergence ratio: $A \ll 1, N \gg 1$
Primary performance: High resolution, high peak efficiency
Type II-Signal processing Bragg cell
Regime of AO diffraction: Transverse SLM
Divergence ratio/Resolution: $A \ll 1, N \gg 1$
Primary performance: Wide bandwidth, large dynamic range

From the above equations, the divergence ratio A and resolution N are related by

$$A = \frac{\delta\theta_o}{\delta\theta_a} = \frac{l\Delta F}{N} \quad (56)$$

where $l = L/L_o$ is the normalized interaction length and $\Delta F = \Delta f/f_o$ is the fractional bandwidth. For isotropic AO diffraction $F, l\Delta F = 1.8$. Thus, an AO deflector is characterized by the requirement $A \ll 1$. As an example, to realize a high resolution AO deflector with 1000 resolvable spots, the divergence ratio must be less than 0.002. The use of expanded incident beam with diffracted limited optics becomes the most critical requirement in the design high-resolution AO deflectors.

In summary, the AO deflector can be viewed as a transverse spatial light modulator (SLM) with a large number of resolvable angular channels. Based on the primary performance goals and the interaction geometry the AO deflector can be further divided into type I-laser beam scanners and type II-signal processing Bragg cells. Table 3 lists the range of the divergence ratio and the key performance requirements for the two types of AO deflectors.

Deflector Interaction Geometry

Principle of Operation

Isotropic AO diffraction In the following we consider the design of AO deflectors.²⁵ Figure 3a and b shows the device geometry of an AO deflector in the interaction plane and transverse plane, respectively. A piezoelectric transducer is bonded to the appropriate face of a selected AO medium, oriented for efficient interaction. A top electrode deposited on the transducer defines the active area with equivalent acoustic beam length L and height H . An optical beam generally truncated to an aperture width, D is incident in the interaction plane at a proper Bragg angle with respect to the acoustic wavefront. The incident beam has a Gaussian profile with a beam waist $2\omega_1$ at $1/e^2$ of the intensity in the interaction plane and a beam waist $2\omega_2$ at $1/e^2$ intensity in the transverse plane.

In accordance to Eqs. (19) and (44) the peak diffraction efficiency of an AO diffraction under the weak interaction approximation is given by

$$\eta = \frac{\pi^2}{2\lambda_o^2} M_2 P_a \left(\frac{L}{H} \right) \cdot J \quad (57)$$

where L and H are the equivalent acoustic beam length and height and J is the overlapping integral given by Eq. (45). As an example, assume an optical Gaussian beam with a truncation ratio of 1.5 and a rectangular shaped electrode, numerical calculation shows when the electrode height is chosen to be equal to H_o so that the interaction region coincide with the acoustic near field. With this choice by Eq. (47), so that the AO interaction region extends from the transducer to the acoustic near field, the peak diffraction efficiency becomes

$$\eta_o = \frac{\pi^2 P_a}{2\lambda_o^3 f^{3/2} \tau^{1/2}} \left(\frac{M_3}{\sqrt{B}} \right) \cdot l \quad (58)$$

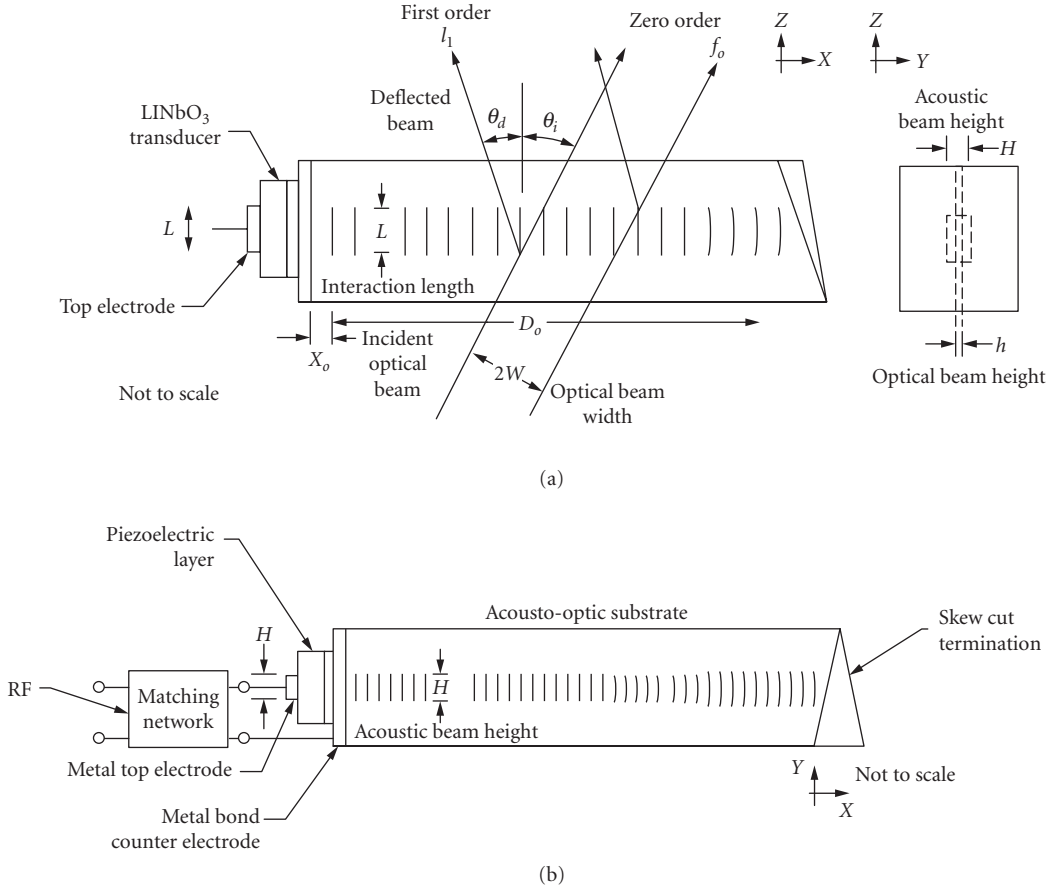


FIGURE 3 Acousto-optic deflector configuration: (a) interaction plane and (b) transverse plane.

Equation (58) is the basic equation for the design of the deflector. The multiplying factor J/h in Eq. (58) depends on the optical beam geometry and the transducer electrode pattern in the transverse plane. For example, for an optical Gaussian beam with a truncation ratio $D/2\omega_1 = 1.5$, and rectangular electrode, the value of J/h exhibits a broad maximum about unity when $h \approx 0.7$. The choice of normalized interaction length $l = L/L_o$ is determined by the specified bandwidth. For isotropic diffraction it is related to the fractional bandwidth by, from Eq. (41):

$$l = \frac{1.8}{\Delta F} \quad (59)$$

With the selection of l , all the key performance parameters can be determined. The frequency response and spatial frequency distribution are obtained from Eqs. (48) and (49).

Anisotropic acoustic beam collimation Equation (58) shows that increased diffraction efficiency can be obtained by selecting AO materials and modes with small slowness curvature B . This is referred to as anisotropic acoustic beam collimation. An example is the shear mode in GaP propagating

along [110] direction.²⁶ In this case $B = 0.026$. Compared to an acoustically isotropic direction, the transducer height can be reduced by a factor of 6.2.

Resolution The resolution of the AO deflector obtainable is limited by the maximum available crystal size, requirement of large optics, and acoustic attenuation across the aperture. For most crystalline solids, the acoustic attenuation is proportional to f^2 . If we allow a change of average attenuation of ξ (dB) across the band, the maximum deflector resolution is given by

$$N_{\max} = \frac{\xi}{\alpha_o f_o} \quad (60)$$

where α_o (dB/ μ sec-GHz²) is the acoustic attenuation coefficient.

It is instructive to estimate the maximum resolution achievable of AO deflectors. Consider, for example, a slow shear wave TeO₂ AO deflector with the following parameters: the octave bandwidth $f_o = 1.5 \Delta f$, acoustic velocity $V = 0.62$ mm/ μ sec, acoustic attenuation coefficient $\alpha_o = 17.9$ (dB/ μ sec-GHz²), maximum acoustic loss $\xi = 4$ dB, aperture size $D = 5$ cm. From Eq. (60), the maximum resolution $N_{\max}$ is estimated to be about 3000 resolvable spots. This is close to the limit of a practical single stage AO deflector.

Birefringent tangential phase matching The maximum interaction length is limited by the bandwidth requirement. One approach of increasing while maintaining the fractional bandwidth ΔF is the use of birefringent diffraction. It was shown in Sec. 6.3 that by choosing the birefringent AO interaction near the tangential phase matching (TPM), a large band of acoustic frequencies will simultaneously satisfy the momentum matching condition. Equivalently a larger l can be used without narrowing the bandwidth. Under the birefringent TPM condition l is related to the fractional bandwidth by, from Eq. (42):

$$l = \frac{3.6}{\Delta F^2} \quad (61)$$

Compared to isotropic diffraction, the birefringent phase matching achieves an efficiency advantage factor of $2/\Delta F$, which becomes particularly significant for smaller fractional bandwidths. For instance, at 50 percent fractional bandwidth, the birefringent phase matching has an advantage factor of 4.

Since the first report on the use of birefringent TPM,²⁷ this performance enhancement technique has been used in the design of high resolution AO deflectors. All of these TPM deflectors are based on the birefringent diffraction in TeO₂ by a slow shear wave propagating on or near the [110] axis. Because of the exceptional low velocity and large figure of merit, these TeO₂ deflectors can achieve a large number of resolvable spots with a relatively low drive power.

The birefringent TPM deflectors can be classified as on-axis, optically rotated (OR) and acoustically rotated (AR) types. The design of these high resolution (HR) AO deflectors shown are briefly discussed as follows.

ON-axis TPM The first TPM type of AO deflector uses a slow shear wave propagating along the [110] axis of the TeO₂ crystal. This is referred as 90° or on-axis design, since the acoustic angle θ_a is chosen to be equal to 90°. For lowering TPM frequency, the AO diffraction birefringent diffraction uses circular birefringence in the TeO₂ crystal along the c axis.²⁸ Due to the slow velocity $V = 0.62$ mm/ μ sec, large resolution is obtainable with a relatively small size crystal. For instance, a 70- μ sec cell realize more than 2000 resolvable spots with an optical aperture size of only 4.3 cm.

The on-axis design has several drawbacks. First, operation of this mode requires the use of circular polarizers. Second, the tangential frequency f_t is fixed and thus other performance parameters such as figure of merit cannot be independently chosen. Third, at the tangential phase-matching frequency, the second-order diffraction is also matched. The first-order diffracted light is rediffracted into the second order and results in a dip at the middle of the deflector passband.

To resolve this, two types of birefringent phase matching (TPM) are proposed. These include optically rotated (OR)²⁵ and acoustically rotated (AR)²⁹ TPM. For both types of phase-matching schemes the tangential matching frequency, $f_t = F_t f_o$, is given by, from Eq. (35):

$$f_t \approx \frac{V(\theta_a)}{\lambda_o} \sqrt{2n_o \Delta n \sin \theta_e} \quad (62)$$

Figure 4 shows the wavevector diagram for the two types of tangential phase matching.

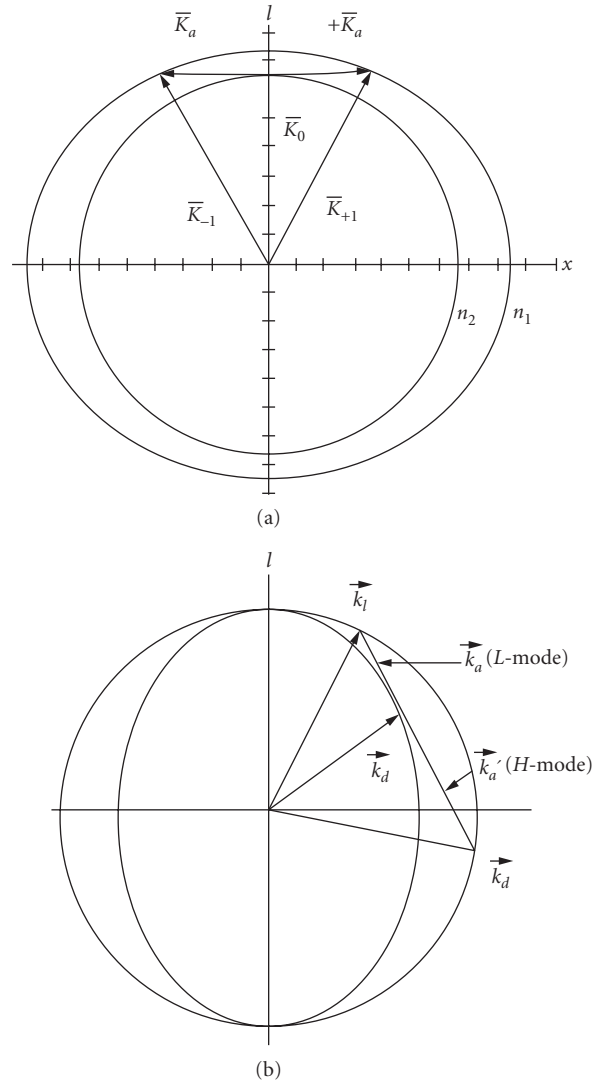


FIGURE 4 Wavevector diagram for tangential phase matching: (a) optically rotated geometry and (b) acoustically rotated geometry.

Optically rotated TPM The wavevector diagram shown in Fig. 4a describes the OR tangential phase matching. In this case the acoustic wavevector is chosen to be perpendicular to the optic axis. The incident light wavevector is allowed to rotate in different polar angles to achieve TPM at the selected frequency. Operation of this cell is more convenient than the on-axis device since the optical beam is linearly polarized. The OR type can realize long time aperture and higher efficiency since the acoustic wave is propagating along the isotropic direction. However, rediffraction into second order occurs since optical locus is symmetrical to the acoustic wave direction, the peak maximum efficiency remains limited by the dip in the center of the frequency response.

Acoustically rotated TPM Figure 4b shows the wavevector diagram for AR type of tangential phase matching. The constant azimuth plane is chosen as the interaction plane and the acoustic wavevector is rotated in the plane to be tangential to the locus of the diffracted light wavevector. Because of the rotation of the acoustic wavevector, the asymmetry of optical wavevector locus removes the degeneracy of the second-order diffraction. The dip in the bandpass is eliminated and thus can operate at the high-efficiency regime. This is the major advantage of the AR type cell. Since the acoustic wave is propagating along an anisotropic path, the large walk-off will reduce the optical aperture and thus limits the maximum resolution. A higher TPM frequency is chosen in order to achieve 2000 resolvable spots.

Phase array beam steering The birefringent type of tangential phase-matching technique is limited to AO diffraction in birefringent crystals. For isotropic materials such as wideband GaP Bragg cells an alternate technique for realizing TPM is to use acoustic beam steering.⁷ In this approach, a phase array of transducers is used so that the composite acoustic wavefront will effectively track the Bragg angle. The simplest phase array employs a fixed inter-relevant phase difference that corresponds to an acoustic delay of $PA/2$, where P is an integer. This is referred to as first-order beam steering and can be realized in either a stepped array^{7,30} or a planar configuration.³¹ The two types of phased array configuration are shown in Fig. 5. The stepped array configuration is more efficient, but fabrication difficulty makes it less practical. The following analysis addresses only the planar configuration of first-order beam steering.

Consider the simplest geometry of a planar first-order beam-steered transducer array where each element is driven with an inter-element phase difference of 180° . The acousto-optic bandpass response of this transmitter configuration is equal to the single-element bandshape multiplied by the interference (array) function,²⁵

$$W(F) = \left(\frac{\sin \pi X}{\pi X} \right)^2 \left(\frac{\sin N \pi Y}{N \sin \pi Y} \right)^2 \quad (63)$$

$$X = \frac{l_e F}{2} (F_b - F) \quad (64)$$

$$Y = \frac{d}{2} \left\{ F(F_b - F) + \frac{1}{d} \right\} \quad (65)$$

where $l_e = L_e/L_o$, $d = D/L_o$, L_e is the length of one element, N is the number of elements, and D is the center-to-center distance between adjacent elements. Since radiation pattern for a single element is broad, the bandpass function is primarily determined by the array functions. The bandpass characteristics of the phased array are similar to birefringent diffraction with an equivalent interaction length $\ell = Nd$ and tangential matching frequency $F_t = 1/\sqrt{d}$.

Referring to Eq. (59), notice that at the peak of the grating lobe for the phase-array radiation, the value of the single element radiation $\text{sinc}^2 x$ is approximately equal to 0.5. There is thus an additional 3 dB loss due to the planar phase array. This is the major disadvantage of the phase array approach.

Since the power density is proportional to ℓ^2 for both types of TPM techniques, the increase of ℓ will significantly reduce the power density. This is of practical importance, since high power density

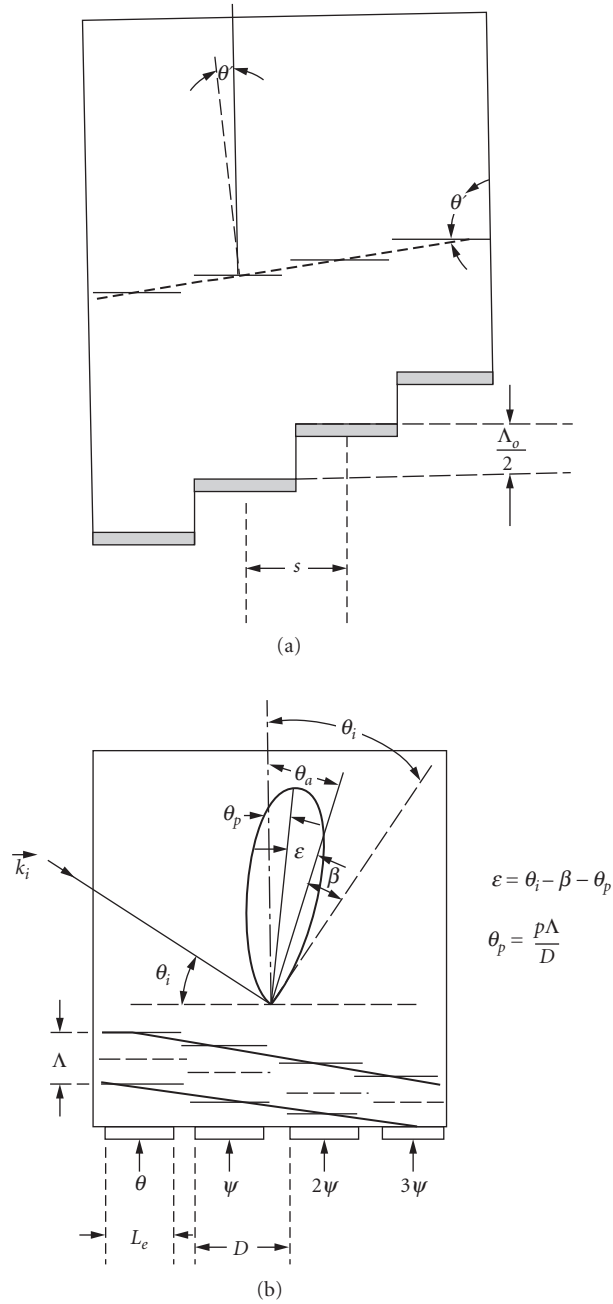


FIGURE 5 Acoustic beam steering using phased array: (a) stepped phased array and (b) planar phased array.

has been the dominant factor for performance degradation and even catastrophic device failure due to thermal gradient, nonlinear acoustics and overheating of transducers.

Birefringent phased array An interesting design combining the two performance enhancement techniques was reported in birefringent AO device using phased-array transducers.³² Theoretical analysis of the new interaction geometry was described and experimentally demonstrated in a wide-band Bragg cell. The new approach provides design flexibility in the choice of acoustic and optic modes for increased efficiency, bandwidth, suppression of nonlinear effects, and optimizing other performance characteristics. The approach was further explored in the design of AO deflectors for specific performance improvement. By applying the birefringent phased array design to on-axis,³³ OR and AR configuration,³⁴ some of the deficiencies in each configuration can be removed. The design flexibility of this approach was also utilized to improve other key performance features, including, for example, the development of wide angle AO Bragg cells.³⁵

AO Laser Beam Scanner The early development of AO deflectors was aimed at laser beam scanning applications. The primary goal is to provide a simple solid-state laser scanner that eliminates the inherent drawbacks of mechanical scanners due to moving parts, such as facet errors and the requirement of realignment because of bearing wear. For certain applications, such as beam-addressed optical memory, the rapid random access capability of AO deflectors offers a distinct advantage.

High Resolution AO Deflector The AO device for laser beam scanning is referred as the high resolution (HR) or type I AO deflector. As shown in Table 3, the primary design objective is to realize high resolution with the choice of a long-time aperture. Another key performance specification is high throughput or peak diffraction efficiency. Table 4 lists the performance of a few representative HR AO deflectors operated in the visible range. To satisfy the high peak efficiency specification, the degradation due to in-band dip must be minimized if OR type TPM rediffraction at high efficiency.

Resolution and Scan Rate In the design of the AO deflector for linear scanning, the most important system specification is the resolution, or maximum number of resolvable spots, N . Another key system parameter to specify is the scan time required to scan a single line and the flyback time. The acoustic transit time across the optical aperture sets the minimum flyback time for the scanner. The finite scan rate also degrades the resolution of the AO deflector. When the acoustic transit time τ (which is equal to flyback time) becomes an appreciable portion of the scan time T , there is a reduction of the effective aperture by the factor $1 - \tau/T$. This results in a resolvable number of spots for the scanning mode

$$N = \tau \Delta f [1 - \tau/(T + \tau)] \quad (66)$$

To minimize the loss of resolution, the transit time τ should be much less than the scan time T .

TABLE 4 Performance of High Resolution AO Deflectors

Material (Mode)	Center Frequency (MHz)	Bandwidth (MHz)	Aperture (μ s)	TB Product	Efficiency (%/W)
TeO ₂ (S)	50	30	70	2100	200
TeO ₂ (S)	90	50	40	2000	110
TeO ₂ (S)	160	100	20	2000	95
GaP(S)	300	200	2.5	500	50

Resolution Enhancement One early design for increased resolution involves the use of cascaded deflectors.³⁶ Another technique is to use higher-order AO diffraction. One interesting design was to utilize the second-order diffraction in an OR-type birefringent cell.³⁷ Since both the first- and second-order diffractions are degenerately phase-matched, efficient redirection of the first order into the second order was obtained. However, the use of the second-order diffraction allows the deflector resolution to be doubled for a given bandwidth and time aperture.

AO Signal Processing Bragg Cell The deflection of an incident optical wave according to a single RF frequency input is just one of the characteristics of the acoustically driven transverse spatial modulator (TSM). More generally, the TSM encodes the light beam with the complete spectrum information contained in the RF signal. An AO deflector simultaneously driven by multifrequencies provides a simple but powerful technique for processing wideband electronic signals. This multifrequency AO device used for signal processing is referred to as the AO Bragg cell.

Dynamic Range The design of Bragg cells is similar to that of the AO deflector. However, instead of optimization of the peak diffraction efficiency, the primary objective of the Bragg cell design is to achieve a large dynamic range rather than the peak diffraction. The dynamic range is defined as the ratio of the intensity of the diffracted light to the intensity of the spurious light in the Bragg cell. The spurious optical beams are caused by the nonlinear mixing of various diffraction orders when multiple frequencies are simultaneously present. Hecht³⁸ analyzed isotropic AO diffraction with multiple acoustic waves with different carrier frequencies based on the coupled wave theory. His analysis shows that the dominant in-band nonlinear signals for two simultaneous signals at f_1 and f_2 are the third-order intermodulation (IM) products occurring at $2f_1 - f_2$ and $2f_2 - f_1$. The magnitude of this spurious signal due to multiple AO diffraction is

$$I_{\text{AO}}(2, -1) = \frac{I_1^2 I_2}{36} \quad (67)$$

where I_1 and I_2 are the diffracted light efficiency at f_1 and f_2 , respectively. To obtain a spurious-free dynamic range of 40 dB, for instance, the Bragg cell must be operated with diffraction efficiency less than 5 percent.

In our experimental work on wideband Bragg cells, we have found another type of IM product in the nonlinear response of Bragg cells.³⁹ This second source of TM products is caused by the nonlinear acoustic (NA) interaction that occurs at high acoustic power densities. A key parameter for the NA type IM products process is the critical interaction length $L_c = (\pi \gamma_2 f \tau)$, where γ_2 is the ratio of second-order nonlinear elastic coefficient to the first-order elastic coefficient. The dominant NA modes are second harmonics and two-tone third-order IM product, which grow according to $(L_c/L)^2 P_1^2/4$ and $(L_c/L)^4 P_1^2 P_2/9$, respectively. Based on the new theory, the AO type of IM products dominates when the ratio L_c/L is small, as the acoustic power increases. The dynamic range of the Bragg cell is reduced by the NA type of IM product, initially by the successive AO diffraction from fundamental and second harmonics and finally by the single AO diffraction from the acoustic IM product.

Wideband Bragg Cell Two candidate materials for wideband Bragg cells are GaP and LiNbO₃. Both have very low acoustic attenuation and large M_3^* , making them ideal for this purpose. A number of high performance wideband cells using these two materials have been developed with bandwidth in excess of 1000 MHz. These include the AR-type birefringent LiNbO₃ cell⁴⁰ and the phased-array GaP cell.⁴¹ The best-known design is a LiNbO₃ device, which uses a y - z propagating off-axis x -polarized shear wave. The device demonstrated an overall bandwidth of 2 GHz, a peak diffraction efficiency of 6 percent per watt and about 600 resolvable spots. A summary of the representative wideband AO Bragg cells is shown in Table 5.

Multichannel Cell As an extension to the usual single-channel configuration there has been considerable activity in the development of multichannel Bragg cells (MCBC). The MCBC uses a pattern

TABLE 5 Performance of Acousto-Optic Bragg Cells ($\lambda_o = 830$ nm)

Material (Mode)	Center Frequency (MHz)	Bandwidth (MHz)	Aperture (μ s)	TB Product	Efficiency (%/W)
GaP(S)	1000	500	2.0	1000	30
LiNbO ₃ (S)	2500	1000	1.0	1000	10
GaP(L)	2500	1000	0.25	250	44
GaP(S)	3000	2000	0.25	500	8
LiNbO ₃ (S)	3000	2000	0.30	600	6

of multiple individually addressed transducer electrodes in the transverse plane. In addition to the design rules used for single-channel cells, other performance parameters such as crosstalk, amplitude, and phase tracking must be considered.

Early work on MCBC uses the y - z cut shear-wave wideband LiNbO₃ cell design.⁴² Good amplitude and phase tracking were obtained over the operating bandwidth of 1 GHz. The channel-to-channel isolation, typically about 25 dB, was limited by RF cross talk. In a later development of GaP MCBC,⁴³ reduction of electrical crosstalk (to -40 dB) was obtained by using stripline interconnection structures. The use of the anisotropic collimating modes in GaP also brings the advantage of lowering acoustic crosstalk.

6.6 ACOUSTO-OPTIC MODULATOR

Acousto-optic interaction has also been used to modulate light. The AO modulator uses the transverse spatial modulation (TSM) of AO interaction to convert an amplitude modulated RF signal into an intensity modulation light beam. Unlike AO deflectors, however, at the output end the diffracted light spectral components are focused into one channel to reproduce the video information via a collinear mixing process. Functionally the most important requirement for the AO modulators is its temporal response. Thus a basic design consideration for AO modulator is the limitation of temporal response caused by the finite acoustic transit time across the light beam. To reduce the transit time effect on the fidelity of the input video information, the AO modulator is operated in a focused beam geometry.

Following similar design procedure as the deflectors, we shall describe the characteristics of AO modulator by the basic parameters. As a modulator the key performance parameters are rise time or video modulation bandwidth, contrast ratio or rejection against the undeflected light beam, and the optical throughput, the diffracted light intensity integrated over the acceptance angular aperture. A thorough discussion on the tradeoff of key performance parameters of AO modulators is reviewed by Johnson.⁴⁴

Based on the optics configuration AO modulators can be divided into type-I focused beam modulator and type II-wide beam modulator. Table 6 lists the interaction geometry and key performance requirement for the two types of AO modulators.

TABLE 6 Interaction geometry and key performance parameters of AO modulators

Type I: Focused beam modulator
Regime of AO diffraction: Transverse SLM
Divergence ratio/Resolution: $A = 1, N = 2$
Primary performance: Wide modulation bandwidth, large optical throughput
Type II: Wide beam modulator
Regime of AO diffraction: Transverse SLM
Divergence ratio: $A \ll 1, N \gg 1$
Prime performance: High peak efficiency, high rejection ratio

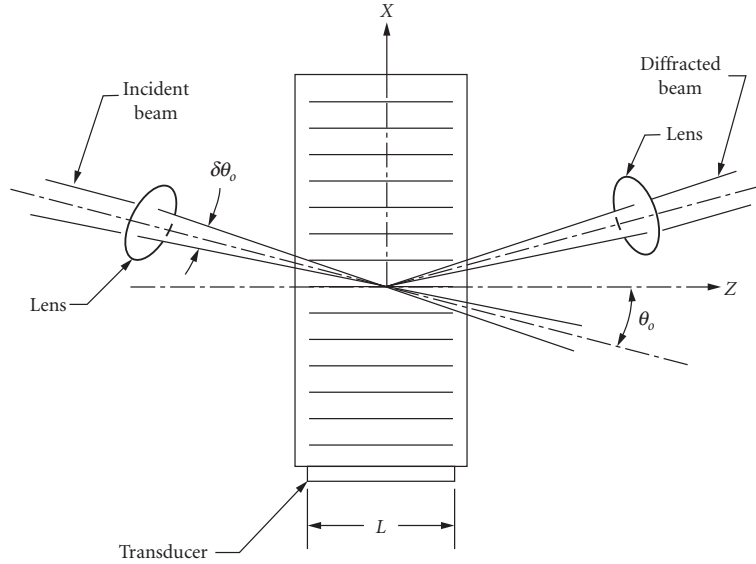


FIGURE 6 Diffraction geometry of acousto-optic modulator.

Principle of Operation Figure 6 shows the diffraction geometry of a focused-beam AO modulator. Unlike the AO deflector the optical beam is focused in both directions into the interaction region near the transducer. For an incident Gaussian beam with a beam waist d , ideally the rise time (10 to 90 percent) of the AO modulator response to a step function input pulse is given by

$$t_r = 0.64 \tau \quad (68)$$

where $\tau = d/V$ is the acoustic transit time across the optical beam. To reduce rise time, the optical beam is focused to a spot size as small as possible. However, focusing of the optical beam will increase the optical beam divergence $\delta\theta_o$, which may exceed the acoustic beam divergence $\delta\theta_a$, $A > 1$. This will result in a decrease of the diffracted light since the Bragg condition will no longer be satisfied. To make a compromise between the frequency response (spatial frequency bandwidth) and temporal response (rise time or modulation bandwidth) the optical and acoustic divergence should be approximately equal, $A = \delta\theta_o/\delta\theta_a \approx 1$. The actual value of the divergence ratio depends on the tradeoff between key performance parameters as dictated by the specific application.

Analog Modulation In the following, we shall consider the design of a focused beam AO modulator for analog modulation, the diffraction of an incident Gaussian beam by an amplitude-modulated (AM) acoustic wave. The carrier, upper, and lower sidebands of the AM acoustic wave will generate three correspondingly diffracted light waves traveling in separate directions. The modulated light intensity is determined by the overlapping collinear heterodyning of the diffracted optical carrier beam and the two sidebands. Using the frequency domain analysis, the diffracted light amplitudes of an AO modulator were calculated. The numerical results are summarized in the plot of modulation bandwidth and peak diffracted light intensity as a function of the optical to acoustic divergence ratio A .¹³ Design of the AO modulator based on the choice of the divergence ratio is discussed below.

Unity Divergence Ratio ($A \leq 1$) The characteristics of the AO modulator can be best described by the modulation transfer function (MTF), defined as the frequency domain response to a sinusoidal video signal. In the limit of $A \leq 1$, the MTF takes a simple form:

$$\text{MTF}(f) = \exp - (\pi f \tau)^2 / 8 \quad (69)$$

The modulation bandwidth f_m , the frequency at -3 dB is given by

$$f_m = \frac{0.75}{\tau} \quad (70)$$

From Eqs. (62) and (64), the modulator rise time and the modulation bandwidth are related by

$$f_m t_r = 0.48 \quad (71)$$

Minimum Profile Distortion ($0.5 < A < 0.67$) Equation (70) shows that the modulation bandwidth can be increased by further reducing the acoustic transit time τ . When the optical divergence $\delta\theta_o$ exceeds the acoustic divergence $\delta\theta_a$, that is, $A > 1$, the Bragg condition is no longer satisfied for all the optical plane waves. The light at the edges of the incident light beam will not be diffracted. This will result in an elliptically shaped diffracted beam. In many laser modulation systems this distortion of the optical beam is not acceptable. The effect of the parameter A on the eccentricity of the diffracted beam was analyzed based on numerical calculation.⁴⁵ The result shows that to limit the eccentricity to less than 10 percent, the divergence ratio value for A is about 0.67.

This distortion of the diffracted beam profile is caused by the finite acceptance angle of the isotropic AO diffraction. The limited angular aperture also results in the lowering the optical throughput. Based on curve fitting the numerical results the peak optical throughput can be expressed as a function of the divergence ratio⁴⁴

$$I_1(A) = 1 - 0.211A^2 + 0.026A^4 \quad (72)$$

As an example, at $A = 0.67$, the peak throughput I_1 is about 94.9 percent. However, with the choice of unity divergence ratio, $A = 1$, it reduces to 81.5 percent.

Digital Modulation ($1.5 < A < 2$) Another important case is the digital, or pulse modulation. Maydan⁴⁶ calculated the rise time and efficiency of pulsed AO modulators. His results show that an optimized choice of A is equal to 1.5, and that the corresponding rise time (10 to 90 percent) is

$$t_r = 0.85\tau \quad (73)$$

Contrast Ratio Another key performance parameter for AO modulator is the extinction ratio or the rejection against the undeflected light. To obtain an adequate extinction ratio, 30 dB, for instance, the angle of separation θ_D is chosen to be equal to twice that of the optical beam divergence. It follows that the acoustic frequency must be greater than

$$f_o = \frac{8}{\pi\tau} \quad (74)$$

Comparing to Eq. (70), the center frequency of the AO modulator is about 4 times that of the modulation frequency.

Birefringent Modulation In the above analysis of focused beam modulator it is assumed that the AO modulator is operated in the isotropic diffraction mode. Similar to the deflector case a large frequency bandwidth can be obtained by using birefringent diffraction. However, the input angular aperture according to Eq. (55) is proportional to $1/L$, an increase of L will in effect reduce the modulation bandwidth as well as the optical throughput. Based on this argument, it was well recognized that the birefringent diffraction offers no advantage for AO modulators. A careful reexamination of this basic limitation shows that a wide angle AO modulation is obtainable if the birefringent diffraction satisfies the noncritical phase matching (NPM) condition.⁴⁸ The concept of NPM will be discussed in Sec. 6.7, "Acousto-Optic Tunable Filter."

TABLE 7 Performance of Acousto-Optic Modulators

Material	Wavelength (μm)	Center Frequency (MHz)	Modulation Bandwidth (MHz)	Rise Time (nsec)	Efficiency (%)
TeO_2	0.44–0.64	200	40	10	80
GaP	0.63–0.83	200	40	10	80
GaP	1.06	160	32	15	70
GaAs	1.3–1.55	120	24	20	70
Ge	10.6	100	20	25	70

Focused Beam AO Modulator

Intensity modulator Simple intensity AO modulators using the focused beam interaction geometry have been standard approach for external modulation of lasers. Since the development of laser diode with wideband internal modulation capability, the use of AO modulators has been limited to gas and solid-state lasers for moderate modulation bandwidth f_m . Because of the high power requirement, in practice f_m of AO modulators is limited to about 50 MHz. Table 7 lists a few selected AO modulators and typical performances.

Intracavity modulator Since AO diffraction occurs in all crystal or amorphous solids, high optical quality, low optical absorption AO materials are readily available for intracavity applications. These intracavity modulator applications include Q-switching, mode locker, and cavity dumping. All of these intracavity AO modulator have been discussed in the early review paper.¹³ In the following we shall only briefly discuss the AO Q-switches.

Q-switching of YAG and other similar solid state lasers has been an importance requirement for industrial applications such as cutting, scribing, marking, and other material processing process. In a Q-switching operation, the optical loss introduced by the intracavity AO modulator keeps the laser below threshold. When the loss is suddenly removed by switching off the acoustic pulse, the laser bursts into a short pulse with extremely high intensity. During Q-switching, the AO modulator should add minimum loss to the laser cavity.

Besides low loss, other key performance requirement of the AO Q-switch include (a) high damage threshold, (b) fast pulse rise time, (c) good thermal and mechanical property, and (d) polarization-insensitive. The last requirement is due to a special characteristics of YAG lasers that it reaches maximum gain when operated in a unpolarized mode. To meet these requirement the standard design for high power laser is a shear wave AO modulator using UV grade fused silica as the interaction medium. Longitudinal mode Q-switch will have the advantage of lower drive power and faster rise time; however, since the diffracted efficiency for perpendicular polarization is five times that of the parallel polarization, the TAG laser will tend to operate in the mode with the lower diffraction efficiency. Thus the use of longitudinal mode Q-switch is limited to lower power polarized laser. More efficient materials such as PbMoO_4 or TeO_2 have been used but because of lower damage threshold these crystal Q-switch is also limited to low power lasers.

Widebeam AO Modulator

Image (Scophony) AO Modulator The focused-beam-type AO modulator has certain disadvantages. The diffraction spread associated with the narrow optical beam tends to lower the diffraction efficiency. More importantly, the focusing of the incident beam results in high peak intensity that can cause optical damage for even relatively low laser power levels. For these reasons, it is desirable to open up the optical aperture. Due to the basic issue of acoustic transit time, the temporal bandwidth of the wide-beam AO modulator will be severely degraded.

In certain applications, such as the laser display system, it is possible to use a much broader optical beam in the modulator than that which would be allowed by the transit time limitation. The operation of the wide-beam modulator is based on the ingenious technique of Scophony light modulation. A brief description of the Scophony light modulator is given below; refer to Johnson⁴⁷ for a thorough treatment.

The Scophony technique is applicable to any system that scans a line at a uniform scan velocity. The basic idea is to illuminate a number of picture elements, or pixels, in the modulator (window) onto the output line, such that the moving video signal in the modulator produces a corresponding image of the pixels that travels across the beam at sound velocity. The image can be made stationary by directing the image through a deflector that scans with equal and opposite velocity. Now, if the window contains N spots, then N picture elements can be simultaneously exposed in the image at any instant, and each picture element will be built up over the access time of the window that is equal to N times the spot time. Since the spots are immobilized, there is no loss of resolution in the image, provided that the modulator bandwidth is sufficient to produce the required resolution. The design of the wide-beam AO modulator is thus the same as that of the Bragg cell.

Acousto-Optic Frequency Shifters Because of the frequency shift associated with the acousto-optic diffraction, the AO frequency shifter (AOFS) provides a capability of shifting the optical frequency of an incident light beam, a unique ability shared by no other technique. Usually the frequency shift specified is either fixed or tunable over a small frequency range. Since the AOFS is often intended for use in interferometry, an important requirement is to minimize the distortion of diffracted light beam profile. For isotropic AO diffraction the use of an input optics design with low divergence ratio $A \ll 1$ is the preferred choice.

Another important requirement is the rejection of incident light, other diffraction order with opposite frequency shift, as well as various spurious signals caused by multiple acoustic reflections within the AO medium. To achieve high suppression of the residual AM components and obtain the ideal spectral purity, the acoustic wave reflection and diffraction in the medium must be considered in the device design and fabrications of the AOFS. Similar to focused beam AO modulator the wide angle design using birefringent phase array can provide large degree of freedom in the optimization of performance parameters such as the optical throughput. The use of crossed polarizers also provides the advantage of increased rejection ratio.

In the design of AOFS, the desired frequency shift f_s is generally chosen to be near the center frequency f_o . This constraint is obviously impractical for small frequency shift $f_s \ll f_o$. The AOFS in this case is typically constructed by putting two copropagating AO cells in cascade so that the two frequencies subtract to give the small frequency shift. The cascaded AO cells configuration can also be used to obtain variable frequency shifts without change of direction. For example, two counterpropagating AO cells can be configured so that the two frequencies are added while the directions cancel each other.

6.7 ACOUSTO-OPTIC TUNABLE FILTER

The acousto-optic tunable filter (AOTF) is an all-solid-state optical filter that operates on the principle of acousto-optic diffraction in an anisotropic medium. The center wavelength of the filter passband can be rapidly tuned across a wide spectral range by changing the frequency of the applied radio frequency (RF) signal. In addition to the electronic tunability, other outstanding features of the AOTF include large angular aperture while maintaining high spectral resolution, inherent intensity, and wavelength modulation capability.

The first AOTF, proposed by Harris and Wallace⁹ used a configuration in which the interacting optical and acoustic waves were collinear. Later, Chang¹⁰ extended the AOTF concept to a noncollinear configuration. The theory and practice of the AOTF have been discussed in a review paper.⁴⁹

Based on the viewpoint of device physics, the AOTF is a new AO device of a fundamentally different nature compared to AO deflectors and modulators. It operates on longitudinal spatial light modulation (SLM) that occurs only in an anisotropic medium. In accordance with the interaction geometry

TABLE 8 Interaction Geometry and Key Performance Goals of AO Tunable Filters

Regime of AO diffraction: Longitudinal SLM
Divergence ratio: $a \gg 1, R \gg 1$
Primary performance: Large angular aperture, high optical throughput
Type II Critical phase matching (CPM) AOTF
Regime of AO diffraction: Longitudinal SPM
Primary performance requirement: High resolution, low drive power

and primary performance requirement the AOTF can be divided into two types. Type I noncritical phase matching (NPM) AOTF exhibits large angular aperture and high optical throughput, is best suited for spectral imaging applications. Type II critical phase matching (CPF) AOTF emphasizes high spectral resolution and low drive power, has shown to be best suited for use as a dynamic-wavelength division multiplexing (WDM) component. Table 8 shows the range of divergence ratio and key performance requirements for the two types of AOTFs.

Principle of Operation

Collinear acousto-optic tunable filter Consider the collinear AO interaction in a birefringent crystal, a linearly polarized light beam will be diffracted into the orthogonal polarization if the momentum matching condition is satisfied. For a given acoustic frequency the diffracted light beam contains only for a small band of optical frequencies centered at the passband wavelength,

$$\lambda_o = \frac{V \Delta n}{f} \quad (75)$$

where Δn is the birefringence. Equation (75) shows that the passband wavelength of the filter can be tuned simply by changing the frequency of the RF signal. To separate the filtered light from the broadband incident light, a pair of orthogonal polarizers is used at the input and output of the filter. The spectral resolution of the collinear AOTF is given by

$$R = \frac{\Delta n L}{\lambda_o} \quad (76)$$

A significant feature of this type of electronically tunable filter is that the spectral resolution is maintained over a relatively large angular distribution of incident light. The total angular aperture (outside the medium) is

$$\Delta \psi = 2n \sqrt{\frac{\lambda_o}{\Delta n L}} \quad (77)$$

where n is the refractive index for the incident light wave.

This unique capability of collinear AOTF for obtaining high spectral resolution within a large angular aperture was experimentally demonstrated using a transmissive-type configuration shown in Fig. 7.⁵⁰ A longitudinal wave is mode converted at the prism interface into a shear wave that propagates down along the x axis of the CaMoO_4 crystal. An incident light with a wavelength satisfying Eq. (75) is diffracted into the orthogonal polarization and is coupled out by the output polarizer. The center wavelength of the filter passband was changed from 400 to 700 nm when the RF frequency was tuned from 114 to 38 MHz. The full width at half-maximum (FWHM) of the filter passband was measured to be 8 Å with an input light cone angle of $\pm 4.8^\circ$ ($F/6$). This angular aperture is more than one order of magnitude larger than that of a grating for the same spectral resolution.

Considering the potential of making electronically driven rapid-scan spectrometers, a dedicated effort was initiated to develop collinear AOTFs for the UV using quartz as the filter medium.⁵¹ Because of the collinear structure quartz AOTF demonstrated very high spectral resolution. With a 5-cm-long crystal, a filter

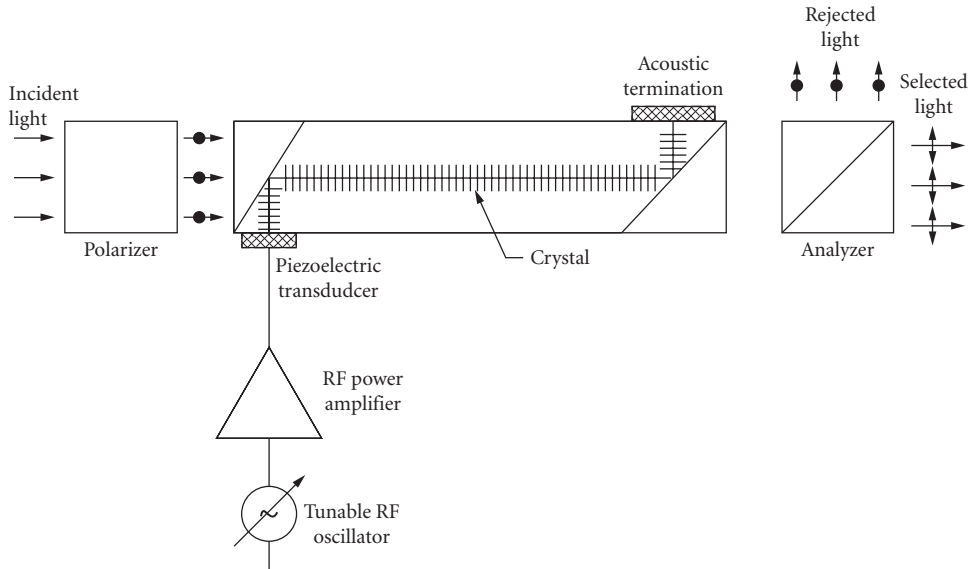


FIGURE 7 Collinear acousto-optic tunable filter with transmissive configuration.

bandwidth of 0.39 nm was obtained at 250 nm. One limitation of this mode conversion type configuration is the complicated fabrication procedures in the filter construction. To resolve this issue, a simpler configuration using acoustic beam walk-off was proposed and demonstrated.⁵² The walk-off AOTF allows the use of multiple transducers and thus could realize a wide tuning range. Experimentally, the passband wavelength was tunable from about 250 to 650 nm by changing the acoustic frequency from 174 to 54 MHz. The simple structure of the walk-off filter is particularly attractive for manufacturing.

Noncollinear AOTF The collinearity requirement limits the AOTF materials to rather restricted classes of crystals. Some of the most efficient AO materials (e.g., TeO_2) are excluded since the pertinent elasto-optic coefficient for collinear AO interaction is zero. Early work has demonstrated a noncollinear TeO_2 AOTF operation using a $S[110]$ on-axis design. However, since the phase-matching condition of the noncollinear AO interaction is critically dependent on the direction of the incident angle of the light beam, this type of filter has a very small angular aperture (on the order of milliradians), and its use must be restricted to well-collimated light sources. To overcome this deficiency, a new method was proposed to obtain a large-angle filter operation in a noncollinear configuration.

The basic concept of the noncollinear AOTF is shown in the wavevector diagram in Fig. 8. The acoustic wavevector is chosen so that the tangents to the incident and diffracted light wavevector loci are parallel. When the parallel tangents condition is met, the phase mismatch due to the change of angle incidence is compensated for by the angular change of birefringence. The AO diffraction thus becomes relatively insensitive to the angle of light incidence, a process referred to as the noncritical phase-matching (NPM) condition. The figure also shows the wavevector diagram for the collinear AOTF as a special case of noncritical phase-matching.

Figure 9 shows the schematic of a noncollinear acousto-optic tunable filter. The first experimental demonstration of the noncollinear AOTF was reported for the visible spectral region using TeO_2 as the filter medium. The filter had a FWHM of 4 nm at an $F/6$ aperture. The center wavelength is tunable from 700 to 450 nm as the RF frequency is changed from 100 to 180 MHz. Nearly 100 percent of the incident light is diffracted with a drive power of 120 mW. The filtered beam is separated from the incident beam with an angle of about 6° . The experimental result is in agreement

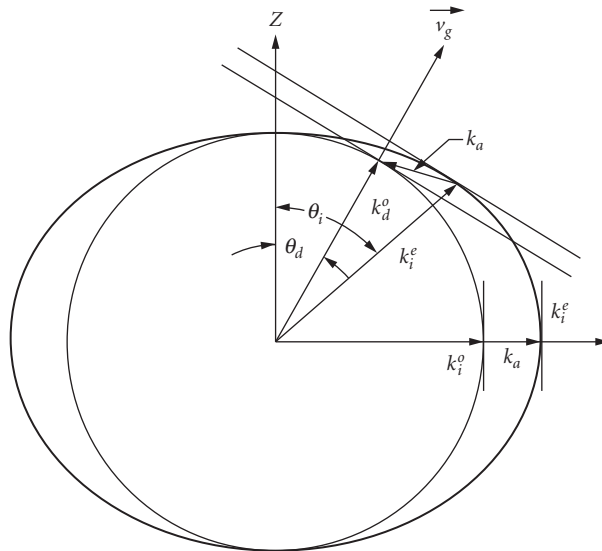


FIGURE 8 Wavevector diagram for noncollinear AOTF showing noncritical phase matching.

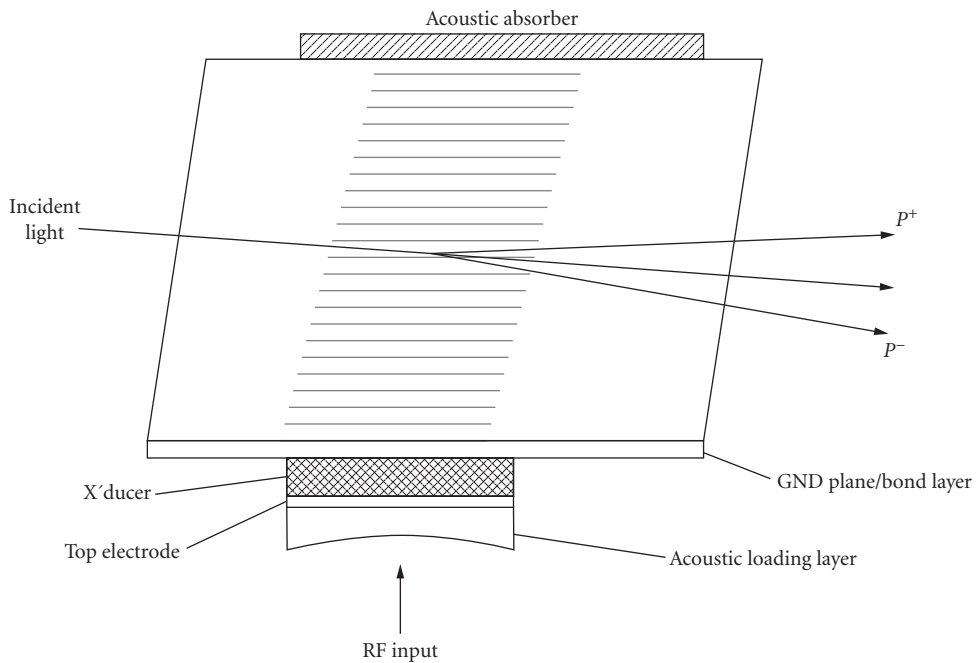


FIGURE 9 Schematic of noncollinear AOTF.

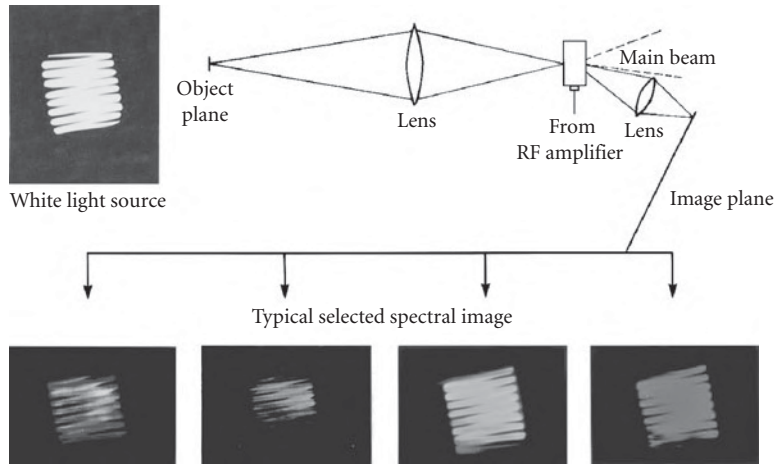


FIGURE 10 Color images of lamp filament through the first noncollinear AOTF.

with a theoretical analysis.⁵³ The early work on the theory and practice of the AOTF is given in a review paper.⁵⁴

After the first noncollinear TeO_2 AOTF, it was recognized that because of its larger aperture and simpler optical geometry, this new type of AOTF would be better suited to spectral imaging applications. A multispectral spectral experiment using a TeO_2 AOTF was performed in the visible region.⁵⁵ The white light beam was spatially separated from the filtered light and blocked by an aperture stop in the immediate frequency plane. A resolution target was imaged through the AOTF and relayed onto the camera. At a few wavelengths in the visible region selected by the driving RF frequencies, the spectral imaging of the resolution was measured. The finest bar target had a horizontal resolution of 144 lines/mm and a vertical resolution of 72 lines/mm. Figure 10 shows the measured spectral images of the lamp filament through the AOTF at a few selected wavelengths in the visible.

Other application concepts of the new AOTF have also been demonstrated. These include the detection of weak laser lines in the strong incoherent background radiation using the extreme difference in the temporal coherence⁵⁶ and the operation of multiwavelength AOTF driven by multifrequencies simultaneously.⁵⁷

It is instructive to summarize the advantages of the noncollinear AOTF: (a) it uses efficient AO materials; (b) it affords the design freedom of choosing the direction of optical and acoustic wave for optimizing efficiency, resolution, angular aperture, and so on; (c) it can be operated without the use of polarizers; and (d) it allows simple filter construction ease for manufacturing. As a result, practically all AOTFs are noncollinear types satisfying the NPM condition.

Noncritical Phase-Matching AOTF

AOTF characteristics The key performance characteristics of the AOTF operated in the noncritical phase-matching (NPM) mode will be reviewed. These characteristics include tuning relation, angle of deflection and imaging resolution, passband response, spectral resolution, angular aperture, transmission and drive power, out-of-band rejection, and sidelobe suppression.

Tuning relation For an AOTF operated at NPM, the tangents to the incident and diffracted optical wavevector surfaces must be parallel. The parallel tangent condition implies that the diffracted extraordinary ray is collinear with the incident ordinary ray,¹⁹

$$\tan \theta_e = e^2 \tan \theta_o \quad (78)$$

For a given incident light angle θ_o , the diffracted light angle θ_d is readily determined from the above equation. Thus, without loss of generality we can assume the incident optical beam to be either ordinary or extraordinary polarized. To minimize wavelength dispersion an ordinary polarized light (*o*-wave) is chosen as the incident light in the following analysis.

Substituting Eq. (78) into the phase matching Eq. (22), the acoustic wave direction and center wavelength of the passband can be expressed as a function of the incident light angle.

$$\tan \theta_a = (\cos \theta_o + \rho_o) / \sin \theta_o \quad (79)$$

$$\lambda_o = [n_o V(\theta_a) / f_a] [(\rho_o - 1)^2 \cos^2 \theta_o + (\delta \rho_o - 1)^2 \sin^2 \theta_o]^{1/2} \quad (80)$$

where $\rho_o = (1 + \delta \sin^2 \theta_o)^{1/2}$ and $\delta = (e^2 - 1)/2$. The above expressions are exact. For small birefringence $\Delta n = n_o |\delta| \ll n_o$, Eqs. (81) and (82) yield the approximate solution for the acoustic wave direction and acoustic frequency:

$$\tan \theta_a = 2 / \tan \theta_o - (1 + \delta) \tan \theta_o \quad (81)$$

$$\lambda_o = (V \Delta n / f_a) (\sin^4 \theta_o + \sin^2 2\theta_o)^{1/2} \quad (82)$$

Angle of deflection In a noncollinear AOTF the diffracted light is spatially separated from the incident light. The deflection angle θ_d is defined as the deviation of the diffracted light from that of the incident light. For small birefringence, θ_d is approximately given by

$$\theta_d \approx \Delta n \sin 2\theta_o \quad (83)$$

Equation (83) shows that θ_d depends only on the input optical angle. It reaches maximum value Δn when $\theta_o = 45^\circ$. For instance, in a TeO_2 AOTF operated at 633 nm, $\Delta n \approx 0.15$, the maximum deflection angle for TeO_2 AOTF is about 8.6° . As a result of the finite angle of deflection angle the noncollinear AOTF can operate without the use of polarizers.

Bandpass response For an acoustic column of uniform amplitude, the bandpass response is given by $T(\lambda_o) = T_o \text{sinc}^2 \Delta\sigma L$, where T_o is the peak transmission at exact momentum matching, L is the interaction length, and $\Delta\sigma$ is the momentum mismatch. It can be shown¹⁹ that

$$\Delta\sigma = -b \sin^2 \theta_i (\Delta\lambda / \lambda_o^2) + (\Delta n / 2\lambda_o) [F_1 (\Delta\theta)^2 + F_2 (\sin \theta_i \Delta\phi)^2 + F_3 (\Delta\theta)^3] \quad (84)$$

where

$$F_1 = 2 - 3 \sin^2 \theta_o \quad F_2 = 2 - \sin^2 \theta_o \quad F_3 = -\sin 2\theta_o \quad (85)$$

$\Delta\lambda$, $\Delta\theta$, and $\Delta\phi$ are deviations in wavelength, polar, and azimuth angles of the incident light beam, b is the dispersion constant defined by $b = \Delta n - \lambda_o \delta(\Delta n) / \delta \lambda_o$.

Resolution Equation (84) shows that half peak transmission occurs when $\Delta\sigma L \approx 0.44$. The full width at half-maximum (FWHM) of the AOTF is given by

$$\Delta\lambda = 0.9 \lambda_o^2 / b L \sin^2 \theta_i \quad (86)$$

The spectral resolution (i.e., longitudinal resolution) of the AOTF is given by

$$R_l = \lambda_o / \Delta\lambda = \Delta n L \sin^2 \theta_i / \lambda_o \quad (87)$$

Angular aperture The acceptance (half) angles in the polar plane ($\theta_o \neq 54.7$) and azimuth planes are given by

$$\Delta\theta = \pm n(\lambda_o/\Delta n L F_1)^{1/2} \quad (88)$$

$$\Delta\phi = \pm n(\lambda_o/\Delta n L F_2)^{1/2} \quad (89)$$

At $\theta_o \approx 54.7$, $F_1 = 0$, and the angular aperture in the polar plane reaches a maximum.

$$\Delta\theta = \pm n(\lambda_o/\Delta n L)^{1/3} \quad (90)$$

As a general rule, the noncollinear AOTF is operated without the use of polarizers. Thus the deflection angle θ_d sets the upper limit for the angular aperture for the AOTF. Assuming the incident is a rectangular beam of width D . This yields a transverse or imaging resolution

$$R_t = \frac{\theta_D}{\delta\theta_o} = \frac{D\Delta n}{\lambda_o} \sin 2\theta_o \quad (91)$$

Optical throughput The filter transmission for normalized input light distribution $I(\theta, \phi_i)$ is obtained by integrating the plane-wave transmission over the solid angle aperture

$$I_d(\Delta\lambda) = \int_{\alpha_i, \beta_i} T(\Delta\lambda, \theta_i, \phi_i) I(\theta_i, \phi_i) \sin\theta_i d\theta_i d\phi_i \quad (92)$$

Transmission and drive power The peak transmission of an AOTF is given by

$$T_o = \sin^2 \left(\frac{\pi^2}{2\lambda_o^2} M_2 P_d L^2 \right)^{1/2} \quad (93)$$

where P_d is the acoustic power density. Maximum transmission occurs when the drive power reaches the value

$$P_a = \frac{\lambda_o^2 A}{2M_2 L^2} \quad (94)$$

where A is the optical aperture. Because of the λ_o^2 dependence, the drive power of an AOTF increases rapidly as the wavelength increases. For instance, at $\lambda_o = 4 \mu\text{m}$, the required acoustic power for a 1 cm infrared TeO_2 AOTF exceeds about 10 W. The required high drive power is perhaps the most severe limitation of the infrared AOTF.

Sidelobe suppression As an optical filter, one of the most undesired characteristics of the AOTF is the nonuniform in-band frequency response and the high sidelobes/fall-off shape near the edge of the passband. The AOTF bandpass response is proportional to the amplitude square of the Fourier transform of the acoustic field. For uniform acoustic excitation, the AOTF exhibits a sinc²-type bandpass characteristic with many sidelobes. The nearest sidelobe is only about 13 dB below the main lobe. A more serious problem is that the envelope of the sidelobe decays slowly at a fall-off rate of -6 dB per octave. In principle, suppression of the high sidelobes can be obtained by the technique of amplitude apodization, a process that slowly reduces the acoustic amplitude profile at the transducer according to a prescribed weighting function. The real problem is the practicality of implementing the apodization. A number of transducer apodization techniques have been experimentally investigated.⁵⁸ These include (1) the use of segmented transducers weighted with designed resistive coupling or varying electrode area; and (2) tapered windows with an air gap or dielectric layer.

The effectiveness of each method varies depending on the size and thickness of the acoustic frequency, among other factors. Typically, the highest sidelobes of apodized devices are about 20 to 25 db below main peak. However, the result is very sensitive to the fabrication tolerance and design parameters. The price to pay is the reliability and high cost of the fabrication.

Spectral Tuning Range Prior to 1980, both the collinear and noncollinear types of AOTFs have been successfully demonstrated covering the spectral range from ultraviolet (UV) to long-wave infrared (LWIR). However, due to the various performance limitations such as polarization dependence, limited optical aperture and resolution and in particular high drive power, the TeO_2 AOTF is the only type that has been deployed in practical optical systems. Considerable effort were devoted to overcome the basic limitations of AOTF technology and extend the usable spectral range.⁵⁸ The progress of AOTFs for the various spectral regions are summarized as follows.

Ultraviolet (UV) AOTF Until recently, collinear quartz AOTF has been the only type for the UV region. The performance limitations of the quartz filter include (1) limited UV operating range: due to the optical absorption of calcite polarizers, the short wavelength cutoff of the UV AOTF is about 240 nm, (2) high drive power: because of the small figure of merit, the drive power is typically above 10 W; and (3) low optical throughput: the fixed on-axis of the collinear interaction geometry lacks the design flexibility with typically small solid angle $\times$ aperture characteristics. To overcome these drawbacks, two noncollinear UV AOTFs were developed.⁵⁹

These include (1) A KDP UV AOTF operated without polarizers was demonstrated at 253.7 nm using a Hg lamp. The device was tunable from 190 to 340 nm. The measured result include a full-width at half-maximum (FWHM) of 0.79 nm and a peak diffraction efficiency of 50 percent at 1.3 W input power. The deflection angle of the filtered light separated from the broadband incident light is 1.9° . (2) A Far UV AOTF made of MgF_2 was tested at 196.03 nm from a hollow cathode lamp. A Brewster-type polarizer made of a pile of thin MgF_2 plates was used in the test. The FWHM of the AOTF was about 3 nm. The diffracted light intensity was about 45 percent at a drive power of 6 W. Because of the UV transmission, the AOTF is extendable to 150 nm.

Mid-infrared AOTF The most severe technological issues that limit the extension of spectral range to thermal infrared region are (1) high drive power requirement due to the λ^2 dependence and (2) catastrophic device failure when operated at low temperatures. To resolve these critical issues a number of the TeO_2 AOTFs feasibility models for the mid-wave infrared (MWIR) have been built and tested.⁶⁰ Performance characteristics of the 1 cm^2 size acoustically resonant AOTFs measured at 3.39 μm include tuning range: 2.5 to 5 μm , bandwidth 15 nm, and drive power (at 80 percent efficiency) less than 0.5 W, with a resonance gain about 20. Repeated operation of the low-power MWIR AOTF was obtainable for temperature above 150 K. However, because of the power thermal conductivity of TeO_2 below the critical temperature the crystal or the transducer may crack. The reliability of the resonant AOTF operated at cryogenic temperature remains a major engineering issue.

Long wave infrared AOTF Operation of the AOTF were demonstrated in the LWIR region of 8 to 12 μm were demonstrated prior to 1980. Stimulated by the important potential application for hyperspectral imaging, there has been considerable effort to develop practical LWIR AOTF for optical remote sensing systems.

For AOTF operated in the long-wave infrared (LWIR) the high drive power is even a much serious problem. Because of its high figure of merit Ti_3AsSe_3 (TAS) is considered to be the best LWIR AOTF material. Even with its exceedingly large AO figure of merit (>2000), the TAS AOTF has to operate on pulsed basis with low duty cycle. However, because of extremely low thermal conductivity and the brittle nature of the TAS crystals, the potential of practical TAS AOTF operated at low temperature does not appear promising.

Performance of NPM AOTF The AOTF possesses many salient features that make it attractive for a variety of optical system applications. To realize the benefits of these merits the limitations of the AOTF must be considered when compared with the competing technologies. Based on the

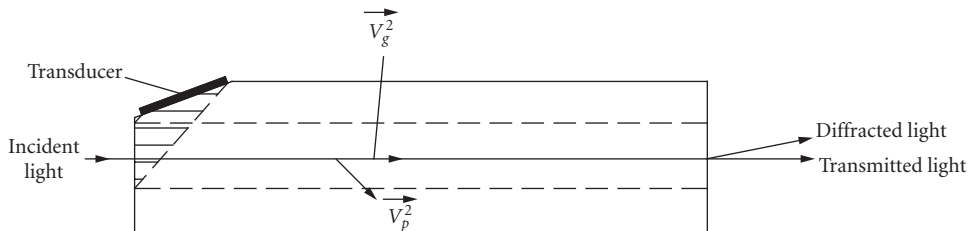
TABLE 9 Broadband and High Resolution Type NPM AOTF

Type	Broadband (8×8 mm)		High-Resolution AOTF	
Aperture	10×10 mm		5×5 mm	
Wavelength	400–1100	700–2500	400–650	650–1100
Bandwidth (cm^{-1})	25	20	5	7
$\Delta\lambda$ (nm)	1	5	0.12	0.5
At λ_o (μm)	633	1550	442	830
Efficiency (%)	80	70	95	95
RF power (W)	1	2	1	2

previous discussion of the usable spectral range, it appears that with the exception of possible new development in the UV, only the TeO_2 AOTFs operated in the visible to SWIR can be considered as a matured technology ready for system deployment. Considering the primary niche, it is pertinent to improve the basic performance of the AOTFs for meeting the system requirement. Table 9 shows two selected high performance NPM AOTFs. These include (1) broadband imaging AOTF with two octave tuning range in a single unit and (2) high resolution high efficiency AOTF suited as rapid random access laser tuner.

The AOTF has the unique capability of being able to simultaneously and independently add or drop multiwavelength signals. As such, it can serve as a WDM cross-connect for routing multiwavelength optical signals along a prescribed connection path determined by the signal's wavelength. Because of this unique attractive feature, the AOTF appears to be suited for use as dynamical reconfigurable components for the WDM network. However, due to the relatively high drive power and low resolution requirement, the AOTF has not been able to meet the requirement for dense wavelength division multiplexing (DWDM) applications. This basic drawback of the AOTF is the result of finite interaction length limited by the large acoustic beam walk-off in the TeO_2 crystal. To overcome this intrinsic limitation, a new type of noncollinear AOTF showing significant improvement of resolution and diffraction efficiency was proposed and demonstrated.^{60,61} The filter is referred to as the collinear beam (CB) AOTF, since the group velocity of the acoustic wave and light beams are chosen to be collinear. An extended interaction length was realized in a TeO_2 AOTF with narrow bandwidth and significantly lower drive power. Figure 11 shows the schematic of an in-line TeO_2 CBAOTF using internal mode conversion.

An initial test of the CBAOTF was performed at 1532 nm using a HeNe laser as the light source.⁶¹ Figure 12 shows the bandpass response of the CBAOTF obtained by monitoring the diffracted light intensity as the acoustic frequency was swept through the laser line. As shown in the figure, the slowly decaying bandpass response appears to be a Lorentzian shape with observable sidelobes. However, the falloff rate of the bandpass at -6 dB per octave wavelength change is the same as the envelope of the sinc^2 response of the conventional AOTF.

**FIGURE 11** Collinear beam AOTF using internal mode conversion.

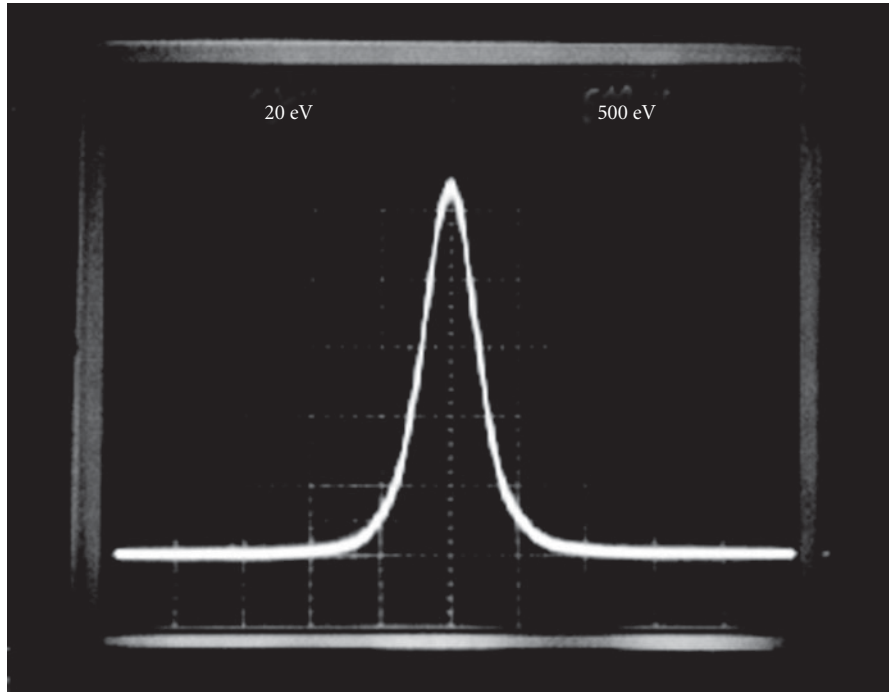


FIGURE 12 Bandpass of CBAOTF (RF swept through a 1.550- μm laser line).

The half-power bandwidth was measured to be 25 kHz, which corresponds to an optical FWHM of 1 nm or 4.3 cm^{-1} . The diffracted light reaches a peak value of 95 percent when the drive power is increased to about 55 mW. Compared to the state-of-the-art high resolution NPM, the measured result showed that the drive power was about 50 times smaller. The low drive power advantage of the CBAOTF is most important for WDM application, which requires simultaneously a large number of channels.

Although the CBAOTF has resolved the most basic limitation of high drive power requirement, to be used as a dynamic DWDM component, there still remains several critical technical bottleneck. For a 100-GHz (0.8 nm at 1550 nm) wavelength spacing system, the AOTF has to satisfy a set of performance goals. These include: polarization independent operation, full width at half-maximum (FWHM) of 0.4 nm, drive power of 150 mW per signal, and the sidelobe must be suppressed to be lower than at least 30 dB at 100 GHz away from the center wavelength. Significant progress has been made in the effort to overcome these critical issues. These are discussed below.

Polarization independence The operation of AOTF is critically dependent on the polarization state of the incoming light beam. To make it polarization-independent, polarization diversity configurations (PDL) are used.⁶² The scheme achieves polarization independence by dividing into two beams of orthogonal polarization, o- and e-rays with a polarization beam splitter (PBS) passing through two single polarization AOTFs in two paths, then combining the two diffracted o- and e-rays of selected wavelengths with a second PBS. Half-wave plates are used to convert the polarization of the light beam so that o-rays are incident onto the AOTF.

Sidelobe suppression Due to low crosstalk requirement, the sidelobe at the channel must be sufficiently low ($\sim 35\text{ dB}$). This is the most basic limitation and as such it will be discussed in some detail. There are three kinds of crosstalk. These include the interchannel crosstalk caused by the

TABLE 10 Measured Performance of CBAOTF

Tuning range	1300–1600 nm
Measured wavelength	1550 nm
FWHM (3 dB)	0.4 nm @ 1.55 nm
Sidelobe @ ch. spacing	–20 dB @ –0.8 nm
Sidelobe @ ch. spacing	–25 dB @ +0.8 nm
Peak efficiency	95(%)
Drive power	80 mW

sidelobe level of the AOTF bandpass; and the extinction ratio, an intrachannel crosstalk due to the finite extinction ratio. The most severe crosstalk is the coherent type that originates from the mixing of sidelobe of light beam λ_1 shifted by frequencies f_1 and f_2 . The optical interference of the two light beams will result in an amplitude-modulated type crosstalk at the difference frequency of $f_1 - f_2$. This type of interchannel crosstalk is much more severe since the modulation is proportional to the amplitude or square root of the sidelobe power level.

Several apodization techniques have been developed in order to suppress the high sidelobe. A simple approach of tilted configuration to simulate a various weighting function appears to be most practical.⁵⁸ A major advantage of this approach is the design flexibility to obtain a desired tradeoff between sidelobe suppression and bandwidth.

Another technique for reducing the sidelobe is to use two or more AOTFs in an incoherent optical cascade. The bandpass response of the incoherently cascaded AOTFs is equal to the product of the single stage and thus can realize a significantly reduced sidelobe. This doubling of sidelevel by cascaded cells has been experimentally demonstrated.⁵⁸

A number of prototype devices of polarization independent (PI) CBAOTF using the tilt configuration have been built. Typical measured results at 1550 nm include 1.0 nm FWHM, peak efficiency 95 percent at 80 mW drive power insertion loss; –3 dB polarization-independent loss; 0.1 dB polarization mode dispersion (PDL); 1 psec, and sidelobe below –27 dB at 3 nm from the center wavelength.⁵⁸

To further reduce the half power bandwidth, a higher angle cut design was chosen in the follow-on experimental work. A 65°, apodized CBAOTF was designed and fabricated. The primary design goal was to meet the specified narrow bandwidth and accept the sidelobe level based on the tradeoff relation. Test results of the 65° devices measured at 1550 are summarized in Table 10. Except the sidelobe suppression goal, the 65° device essentially satisfies all other specifications. The specified sidelobe level of –35 dB for the 100-GHz channel spacing can be met by using two CBAOTFs in cascade.

In conclusion, it is instructive to emphasize the unique advantage of the AOTF. Because of its random access wavelength tunability over large spectral range, the CBAOTF provides a low-cost implementation of a dynamic multiwavelength component for DWDM and other noncommunication type of application with nonuniform distribution of wavelengths.

6.8 REFERENCES

1. L. Brillouin, "Diffusion de la lumière et des ray x par un corps transparent homogène," *Ann. Phys.* **17**:80–122 (1992).
2. P. Debye and F. W. Sears, "On the Scattering of Light by Supersonic Waves," *Proc. Nat. Acad. Sci. (U.S.)* **18**:409–414 (1932).
3. R. Lucas and P. Biquard, "Propriétés optiques des milieux solides et liquides soumis aux vibration élastiques ultra sonores," *J. Phys. Rad.* **3**:464–477 (1932).
4. C. V. Raman and N. S. Nagendra Nath, "The Diffraction of Light by High Frequency Sound Waves," *Proc. Ind. Acad. Sci.* **2**:406–420 (1935); **3**:75–84 (1936); **3**:459–465 (1936).

5. M. Born and E. Wolf, *Principles of Optics*, 3rd ed., Pergamon Press, New York, 1965, Chap. 12.
6. E. I. Gordon, "A Review of Acoustooptical Deflection and Modulation Devices," *Proc. IEEE* **54**:1391–1401 (1966).
7. A. Korpel, R. Adler, Desmares, and W. Watson, "A Television Display Using Acoustic Deflection and Modulation of Coherent Light," *Proc. IEEE* **54**:1429–1437 (1966).
8. R. W. Dixon, "Acoustic Diffraction of Light in Anisotropic Media," *IEEE J. Quantum Electron.* **QE-3**:85–93 (Feb. 1967).
9. S. E. Harris and R. W. Wallace, "Acousto-Optic Tunable Filter," *J. Opt. Soc. Am.* **59**:744–747 (June 1969).
10. I. C. Chang, "Noncollinear Acousto-Optic Filter with Large Angular Aperture," *Appl. Phys. Lett.* **25**:370–372 (Oct. 1974).
11. E. K. Sittig, "Elasto-Optic Light Modulation and Deflection," Chap. VI, in *Progress in Optics*, vol. X, E. Wolf (ed.), North-Holland, Amsterdam, 1972.
12. N. Uchida and N. Niizeki, "Acoustooptic Deflection Materials and Techniques," *Proc. IEEE* **61**:1073–1092 (1973).
13. I. C. Chang, "Acoustooptic Devices and Applications," *IEEE Trans. Sonics Ultrason.* **SU-23**:2–22 (1976).
14. A. Korpel, "Acousto-Optics," Chap. IV, in *Applied Optics and Optical Engineering*, R. Kingslake and B. J. Thompson (eds.), Academic Press, New York, vol. VI, 1980.
15. J. F. Nye, *Physical Properties of Crystals*, Clarendon Press, Oxford, England, 1967.
16. F. Nelson and M. Lax, "New Symmetry for Acousto-Optic Scattering," *Phys. Rev., Lett.*, **24**:378–380 (Feb. 1970); "Theory of Photoelastic Interaction," *Phys. Rev.* **B3**:2778–2794 (Apr. 1971).
17. J. Chapelle and L. Tael, "Theorie de la diffusion de la lumiere par les cristaux fortement piezoelectriques," *C. R. Acad. Sci.* **240**:743 (1955).
18. W. R. Klein and B. D. Cook, "Unified Approach to Ultrasonic Light Diffraction," *IEEE Trans. Sonics Ultrason.* **SU-14**:123–134 (1967).
19. D. A. Klienman, "Optical Second Harmonic Generation," *Phys. Rev.* **128**:1761 (1962).
20. I. C. Chang, "Acousto-Optic Tunable Filters," *Opt. Eng.* **20**:824–828 (1981).
21. A. Korpel "Acoustic Imaging by Diffracted Light. I. Two-Dimensional Interaction," *IEEE Trans. Sonics Ultrason.* **SU-15**(3):153–157 (1968).
22. I. C. Chang and D. L. Hecht, "Device Characteristics of Acousto-Optic Signal Processors," *Opt. Eng.* **21**:76–81 (1982).
23. I. C. Chang, "Selection of Materials for Acoustooptic Devices," *Opt. Eng.* **24**:132–137 (1985).
24. I. C. Chang, "Acousto-Optic Devices and Applications," in *Optic Society of America Handbook of Optics*, 2nd ed, M. Bass (ed.), Vol. II, Chap. 12, pp. 12.1–54, 1995.
25. I. C. Chang, "Design of Wideband Acoustooptic Bragg Cells," *Proc. SPIE* **352**:34–41 (1983).
26. D. L. Hecht and G. W. Petrie, "Acousto-Optic Diffraction from Acoustic Anisotropic Shear Modes in GaP," *IEEE Ultrason. Symp. Proc.* p. 474, Nov. 1980.
27. E. G. H. Lean, C. F. Quate, and H. J. Shaw, "Continuous Deflection of Laser Beams," *Appl. Phys. Lett.* **10**:48–50 (1967).
28. W. Warner, D. L. White, and W. A. Bonner, "Acousto-Optic Light Deflectors Using Optical Activity in Praterullurite," *J. Appl. Phys.* **43**:4489–4495 (1972).
29. T. Yano, M. Kawabuchi, A. Fukumoto, and A. Watanabe, "TeO₂ Anisotropic Bragg Light Deflector Without Midband Degeneracy," *Appl. Phys. Lett.* **26**:689–691 (1975).
30. G. A. Couquin, J. P. Griffin, and L. K. Anderson, "Wide-Band Acousto-Optic Deflectors Using Acoustic Beam Steering," *IEEE Trans. Sonics Ultrason.* **SU-18**:34–40 (Jan. 1970).
31. D. A. Pinnow, "Acousto-Optic Light Deflection: Design Considerations for First Order Beamsteering Transducers," *IEEE Trans. Sonics Ultrason.* **SU-18**:209–214 (1971).
32. I. C. Chang, "Birefringent Phased Array Bragg Cells," *IEEE Ultrason. Symp. Proc.*, pp. 381–384, 1985).
33. E. H. Young, H. C. Ho, S. K. Yao, and J. Xu, "Generalized Phased Array Bragg Interaction in Birefringent Materials," *Proc. SPIE* **1476**: (1991).
34. A. J. Hoffman and E. Van Rooyen, "Generalized Formulation of Phased Array Bragg Cells in Uniaxial Crystals," *IEEE Ultrason. Symp. Proc.* p. 499, 1989.

35. R. T. Waverka and K. Wagner "Wide Angle Aperture Acousto-Optic Bragg Cell," *Proc. SPIE* **1562**:66–72 (1991).
36. W. H. Watson and R. Adler, "Cascading Wideband Acousto-Optic Deflectors," *IEEE Conf. Laser Eng. Appl.* Washington., D.C., June 1969.
37. I. C. Chang and D. L. Hecht, "Doubling Acousto-Optic Deflector Resolution Utilizing Second Order Birefringent Diffraction," *Appl. Phys. Lett.* **27**:517–518 (1975).
38. D. L. Hecht, "Multifrequency Acousto-Optic Diffraction," *IEEE Trans. Sonics and Ultrason.* **SU-24**:7 (1977).
39. I. C. Chang and R. T. Wererka, "Multifrequency Acousto-Optic Diffraction," *IEEE Ultrason. Symp. Proc.* p. 445, Oct. 1983.
40. I. C. Chang and S. Lee, "Efficient Wideband Acousto-Optic Bragg Cells," *IEEE Ultrason. Symp. Proc.* p. 427, Oct. 1983.
41. I. C. Chang et al., "Progress of Acousto-Optic Bragg Cells," *IEEE Ultrason. Symp. Proc.* p. 328, 1984.
42. I. C. Chang and R. Cadieux, "Multichannel Acousto-Optic Bragg Cells," *IEEE Ultrason. Symp. Proc.* p. 413, 1982.
43. W. R. Beaudot, M. Popek, and D. R. Pape, "Advances in Multichannel Bragg Cell Technology," *Proc. SPIE* **639**:28–33 (1986).
44. R. V. Johnson, "Acousto-Optic Modulator," in *Design and Fabrication of Acousto-Optic Devices*, A. Goutzoulis and D. Pape, (eds.), Marcel Dekker, New York, 1994.
45. E. H. Young and S. K. Yao, "Design Considerations for Acousto-Optic Devices," *Proc. IEEE* **69**:54–64 (1981).
46. D. Maydan, "Acousto-Optic Pulse Modulators," *J. Quantum Electron.* **QE-6**:15–24 (1967).
47. R. V. Johnson, "Scophony Light Valve," *Appl. Opt.* **18**:4030–4038 (1979).
48. I. C. Chang, "Large Angular Aperture Acousto-Optic Modulator," *IEEE Ultrason. Symp. Proc.* pp. 867–870, 1994.
49. I. C. Chang, "Acoustic-Optic Tunable Filters," in *Acousto-Optic Signal Processing*, N. Berg and J. M. Pellegrino (eds.), Marcel Dekker, New York, 1996.
50. S. E. Harris, S. T. K. Nieh, and R. S. Feigelson, "CaMoO₄ Electronically Tunable Acousto-Optical Filter," *Appl. Phys. Lett.* **17**:223–225 (Sep. 1970).
51. J. A. Kusters, D. A. Wilson, and D. L. Hammond, "Optimum Crystal Orientation for Acoustically Tuned Optic Filters," *J. Opt. Soc. Am.* **64**:434–440 (Apr. 1974).
52. I. C. Chang, "Tunable Acousto-Optic Filter Utilizing Acoustic Beam Walk-Off in Crystal Quartz," *Appl. Phys. Lett.* **25**:323–324 (Sep. 1974).
53. I. C. Chang, "Analysis of the Noncollinear Acousto-Optic Filter," *Electron. Lett.* **11**:617–618 (Dec. 1975).
54. D. L. Hecht, I. C. Chang, and A. Boyd, "Multispectral Imaging and Photomicroscopy Using Tunable Acousto-Optic Filters," *OSA Annual Meeting*, Boston, Mass., Oct. 1975.
55. I. C. Chang, "Laser Detection Using Tunable Acousto-Optic Filter," *J. Quantum Electron.* **14**:108 (1978).
56. I. C. Chang, et al., "Programmable Acousto-Optic Filter," *IEEE Ultrason. Symp. Proc.* p. 40, 1979.
57. R. T. Waverka, P. Katzka, and I. C. Chang, "Bandpass Apodization Techniques for Acousto-Optic Tunable Filters," *IEEE Ultrason. Symp.*, San Francisco, CA, Oct. 1985.
58. I. C. Chang, "Progress of Acoustooptic Tunable Filter," *IEEE Ultrason. Symp. Proc.* p. 819, 1996.
59. I. C. Chang and J. Xu, "High Performance AOTFs for the Ultraviolet," *IEEE Ultrason. Proc.* p. 1289, 1988.
60. I. C. Chang, "Collinear Beam Acousto-Optic Tunable Filter," *Electron Lett.* **28**:1255 (1992).
61. I. C. Chang et al, "Bandpass Response of Collinear Beam Acousto-Optic Filter," *IEEE Ultrason. Symp. Proc.* vol. 1, pp. 745–748.
62. I. C. Chang, "Acousto-Optic Tunable Filters in Wavelength Division Multiplexing (WDM) Networks," *1997 Conf. Laser Electro-Optics (CLEO)*, Baltimore, MD, May 1997.

This page intentionally left blank

ELECTRO-OPTIC MODULATORS

Georgeanne M. Purvinis

*The Battelle Memorial Institute
Columbus, Ohio*

Theresa A. Maldonado

*Department of Electrical and Computer Engineering
Texas A&M University
College Station, Texas*

7.1 GLOSSARY

$\mathbf{A}, [\mathbf{A}]$	general symmetric matrix
$[\mathbf{a}]$	orthogonal transformation matrix
b	electrode separation of the electro-optic modulator
$\mathbf{D}$	displacement vector
d	width of the electro-optic crystal
$\mathbf{E}$	electric field
$\mathbf{H}$	magnetic field
IL	insertion loss
$\mathbf{k}$	wavevector, direction of phase propagation
L	length of the electro-optic crystal
L/b	aspect ratio
N	number of resolvable spots
n_m	refractive index of modulation field
n_x, n_y, n_z	principal indices of refraction
r_{ijk}	third-rank linear electro-optic coefficient tensor
$\mathbf{S}$	Poynting (ray) vector, direction of energy flow
S_{ijkl}	fourth-rank quadratic electro-optic coefficient tensor
T	transmission or transmissivity
V	applied voltage
V_π	half-wave voltage
v_π	phase velocity
v_m	modulation phase velocity

v_s	ray velocity
w	beamwidth
w_o	resonant frequency of an electro-optic modulator circuit
$\mathbf{X}$	position vector in cartesian coordinates
$\mathbf{X}'$	electrically perturbed position vector in cartesian coordinates
(x, y, z)	unperturbed principal dielectric coordinate system
(x', y', z')	new electro-optically perturbed principal dielectric coordinate system
(x'', y'', z'')	wavevector coordinate system
(x''', y''', z''')	eigenpolarization coordinate system
β_1	polarization angle between x''' and x''
β_2	polarization angle between y''' and x''
Γ	phase retardation
Γ_m	amplitude modulation index
$\Delta\eta$	electro-optically induced change in the index of refraction or birefringence
$\Delta(1/\eta^2)$	electro-optically induced change in an impermeability tensor element
$\Delta\phi$	angular displacement of beam
$\Delta\nu$	bandwidth of a lumped electro-optic modulator
δ	phase modulation index
$[\mathcal{E}]$	permittivity tensor
ϵ_0	permittivity of free space
$[\mathcal{E}]^{-1}$	inverse permittivity tensor
$\epsilon_x, \epsilon_y, \epsilon_z$	principal permittivities
$[\epsilon]$	dielectric constant tensor
$[\epsilon]^{-1}$	inverse dielectric constant tensor
$\epsilon_x, \epsilon_y, \epsilon_z$	principal dielectric constants
η_m	extinction ratio
θ	half-angle divergence
θ_k, ϕ_k	orientation angles of the wavevector in the (x, y, z) coordinate system
ϑ	optic axis angle in biaxial crystals
λ	wavelength of the light
$[\lambda]$	diagonal matrix
ν_{tw}	bandwidth of a traveling wave modulator
ξ	modulation efficiency
ρ	modulation index reduction factor
τ	transit time of modulation signal
ϕ	phase of the optical field
Ω	plane rotation angle
ϖ	beam parameter for bulk scanners
ω_d	frequency deviation
ω_e	stored electric energy density
ω_m	modulation radian frequency
$[1/n^2]'$	electro-optically perturbed impermeability tensor
$[1/n^2]$	inverse dielectric constant (impermeability) tensor
Γ_{wg}	overlap correction factor
β	waveguide effective propagation constant

7.2 INTRODUCTION

Electro-optic modulators are used to control the amplitude, phase, polarization state, or position of an optical beam, or light wave carrier, by application of an electric field. The electro-optic effect is one of several means to impose information on, or modulate, the light wave. Other means include acousto optic, magneto optic, thermo optic, electroabsorption, mechanical shutters, and moving mirror modulation and are not addressed in this chapter, although the fundamentals presented in this chapter may be applied to other crystal optics driven modulation techniques.

There are basically two types of modulators: bulk and integrated optic. Bulk modulators are made of single pieces of optical crystals, whereas the integrated optic modulators are constructed using waveguides fabricated within or adjacent to the electro-optic material. Electro-optic devices have been developed for application in communications,^{1,4} analog and digital signal processing,⁵ information processing,⁶ optical computing,^{6,7} and sensing.^{5,7} Example devices include phase and amplitude modulators, multiplexers, switch arrays, couplers, polarization controllers, and deflectors.^{1,2} Given are rotation devices,⁸ correlators,⁹ A/D converters,¹⁰ multichannel processors,¹¹ matrix-matrix and matrix-vector multipliers,¹¹ sensors for detecting temperature, humidity, radio-frequency electrical signals,^{5,7} and electro-optic sampling in ultrashort laser applications.^{12,13} The electro-optic effect allows for much higher modulation frequencies than other methods, such as mechanical shutters, moving mirrors, or acousto-optic devices, due to a faster electronic response time of the material.

The basic idea behind electro-optic devices is to alter the optical properties of a material with an applied voltage in a controlled way. The direction dependent, electrically induced physical changes in the optical properties are mathematically described by changes in the second rank permittivity tensor. The tensor can be geometrically interpreted using the index ellipsoid, which is specifically used to determine the refractive indices and polarization states for a given direction of phase propagation. The changes in the tensor properties translate into a modification of some parameter of a light wave carrier, such as phase, amplitude, frequency, polarization, or position of the light as it propagates through the device. Therefore, understanding how light propagates in these materials is necessary for the design and analysis of electro-optic devices. The following section gives an overview of light propagation in anisotropic materials that are homogeneous, nonmagnetic, lossless, optically inactive, and nonconducting. Section 7.4 gives a geometrical and mathematical description of the linear and quadratic electro-optic effects. This section illustrates how the optical properties described by the index ellipsoid change with applied voltage. A mathematical approach is offered to determine the electrically perturbed principal dielectric axes and indices of refraction of any electro-optic material for any direction of the applied electric field as well as the phase velocity indices and eigenpolarization orientations for a given wavevector direction. Sections 7.5 and 7.6 describe basic bulk electro-optic modulators and integrated optic modulators, respectively. Finally, example applications, common materials, design considerations, and performance criteria are discussed.

The discussion presented in this chapter applies to any electro-optic material, any direction of the applied voltage, and any direction of the wavevector. Therefore, no specific materials are described explicitly, although materials such as lithium niobate (LiNbO_3), potassium dihydrogen phosphate (KDP), and gallium arsenide (GaAs) are just a few materials commonly used. Emphasis is placed on the general fundamentals of the electro-optic effect, bulk modulator devices, and practical applications.

7.3 CRYSTAL OPTICS AND THE INDEX ELLIPSOID

With an applied electric field, a material's anisotropic optical properties will be modified, or an isotropic material may become optically anisotropic. Therefore, it is necessary to understand light propagation in these materials. For any anisotropic (optically inactive) crystal class there are two allowed orthogonal linearly polarized waves propagating with differing phase velocities for a given wavevector $\mathbf{k}$. Biaxial crystals represent the general case of anisotropy. Generally, the allowed waves exhibit *extraordinary-like* behavior; the wavevector and ray (Poynting) vector directions differ.

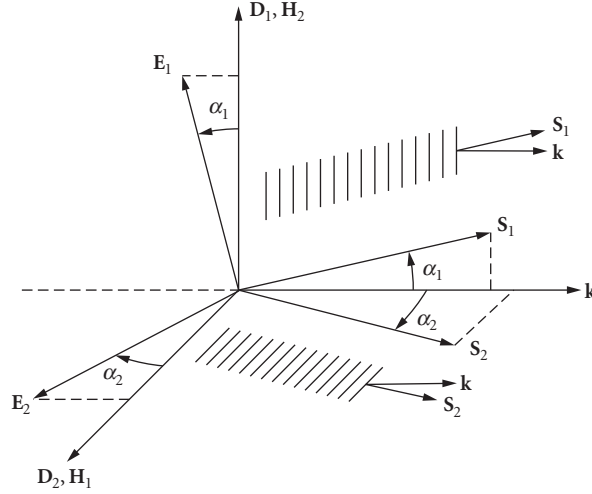


FIGURE 1 The geometric relationships of the electric quantities $\mathbf{D}$ and $\mathbf{E}$ and the magnetic quantities $\mathbf{B}$ and $\mathbf{H}$ to the wavevector $\mathbf{k}$ and the ray vector $\mathbf{S}$ are shown for the two allowed extraordinary-like waves propagating in an anisotropic medium.¹⁴

In addition, the phase velocity, polarization orientation, and ray vector of each wave change distinctly with wavevector direction. For each allowed wave, the electric field $\mathbf{E}$ is not parallel to the displacement vector $\mathbf{D}$ (which defines polarization orientation) and, therefore, the ray vector $\mathbf{S}$ is not parallel to the wavevector $\mathbf{k}$ as shown in Fig. 1. The angle α_1 between $\mathbf{D}$ and $\mathbf{E}$ is the same as the angle between $\mathbf{k}$ and $\mathbf{S}$, but for a given $\mathbf{k}$, $\alpha_1 \neq \alpha_2$. Furthermore, for each wave $\mathbf{D} \perp \mathbf{k} \perp \mathbf{H}$ and $\mathbf{E} \perp \mathbf{S} \perp \mathbf{H}$, forming orthogonal sets of vectors. The vectors $\mathbf{D}$, $\mathbf{E}$, $\mathbf{k}$, and $\mathbf{S}$ are coplanar for each wave.¹⁴

The propagation characteristics of the two allowed orthogonal waves are directly related to the fact that the optical properties of an anisotropic material depend on direction. These properties are represented by the constitutive relation $\mathbf{D} = [\epsilon] \mathbf{E}$, where $[\epsilon]$ is the permittivity tensor of the medium and $\mathbf{E}$ is the corresponding optical electric field vector. For a homogeneous, nonmagnetic, lossless, optically inactive, and nonconducting medium, the permittivity tensor has only real components. Moreover, the permittivity tensor and its inverse, $[\epsilon]^{-1} = 1/\epsilon_0 [1/n^2]$, where n is the refractive index, are symmetric for all crystal classes and for any orientation of the dielectric axes.^{15–17} Therefore the matrix representation of the permittivity tensor can be diagonalized, and in principal coordinates the constitutive equation has the form

$$\begin{pmatrix} D_x \\ D_y \\ D_z \end{pmatrix} = \begin{pmatrix} \epsilon_x & 0 & 0 \\ 0 & \epsilon_y & 0 \\ 0 & 0 & \epsilon_z \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \quad (1)$$

where reduced subscript notation is used. The principal permittivities lie on the diagonal of $[\epsilon]$.

The index ellipsoid is a construct with geometric characteristics representing the phase velocities and the vibration directions of $\mathbf{D}$ of the two allowed plane waves corresponding to a given optical wave-normal direction $\mathbf{k}$ in a crystal. The index ellipsoid is a quadric surface of the stored electric energy density ω_e of a dielectric,^{15,18}

$$\omega_e = \frac{1}{2} \mathbf{E} \times \mathbf{D} = \frac{1}{2} \sum_i \sum_j E_i \epsilon_{ij} E_j = \frac{1}{2} \epsilon_0 \mathbf{E}^T [\epsilon] \mathbf{E} \quad i, j = x, y, z \quad (2a)$$

where T indicates the transpose.

In principal coordinates, that is, when the dielectric principal axes are parallel to the reference coordinate system, the ellipsoid is simply

$$\omega_e = \frac{1}{2} \epsilon_0 (E_x^2 \epsilon_x + E_y^2 \epsilon_y + E_z^2 \epsilon_z) \quad (2b)$$

where the convention for the dielectric constant subscript is $\epsilon_{xx} = \epsilon_x$, and so on. The stored energy density is positive for any value of electric field; therefore, the quadric surface is always given by an ellipsoid.^{15,18–20}

Substituting the constitutive equation, Eq. (2b) assumes the form $(D_x^2/\epsilon_x) + (D_y^2/\epsilon_y) + (D_z^2/\epsilon_z) = 2\omega_e/\epsilon_0$. By substituting $x = D_x/(2\omega_e/\epsilon_0)^{1/2}$ and $n_x^2 = \epsilon_x$ and similarly for y and z , the ellipsoid is expressed in cartesian principal coordinates as

$$\frac{x^2}{n_x^2} + \frac{y^2}{n_y^2} + \frac{z^2}{n_z^2} = 1 \quad (3a)$$

In a general orthogonal coordinate system, that is, when the reference coordinate system is not aligned with the principal dielectric coordinate system, the index ellipsoid of Eq. (3a) can be written in summation or matrix notation as

$$\sum_i \sum_j X_i (1/n_{ij}^2) X_j = \mathbf{X}^T [1/n^2] \mathbf{X} = (x \ y \ z) \begin{pmatrix} 1/n_{xx}^2 & 1/n_{xy}^2 & 1/n_{xz}^2 \\ 1/n_{xy}^2 & 1/n_{yy}^2 & 1/n_{yz}^2 \\ 1/n_{xz}^2 & 1/n_{yz}^2 & 1/n_{zz}^2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (3b)$$

where $\mathbf{X} = [x, y, z]^T$, and all nine elements of the inverse dielectric constant tensor, or impermeability tensor, may be present. For sections that follow, the index ellipsoid in matrix notation will be particularly useful.

Equation (3b) is the general index ellipsoid for an optically biaxial crystal. If $n_{xx} = n_{yy}$, the surface becomes an ellipsoid of revolution, representing a uniaxial crystal. In this crystal, one of the two allowed eigenpolarizations will always be an *ordinary* wave with its Poynting vector parallel to the wavevector and $\mathbf{E}$ parallel to $\mathbf{D}$ for any direction of propagation. An isotropic crystal ($n_{xx} = n_{yy} = n_{zz}$) is represented by a sphere with the principal axes having equal length. Any wave propagating in this crystal will exhibit ordinary characteristics. The index ellipsoid for each of these three optical symmetries is shown in Fig. 2.

For a general direction of phase propagation $\mathbf{k}$, a cross section of the ellipsoid through the origin perpendicular to $\mathbf{k}$ is an ellipse, as shown in Fig. 2. The major and minor axes of the ellipse represent the

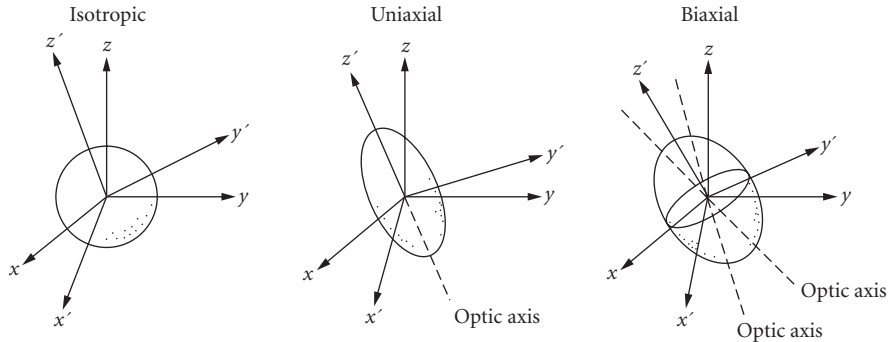


FIGURE 2 The index ellipsoids for the three crystal symmetries are shown in nonprincipal coordinates (x', y', z') relative to the principal coordinates (x, y, z). For isotropic crystals, the surface is a sphere. For uniaxial crystals, it is an ellipsoid of revolution. For biaxial crystals, it is a general ellipsoid.²¹

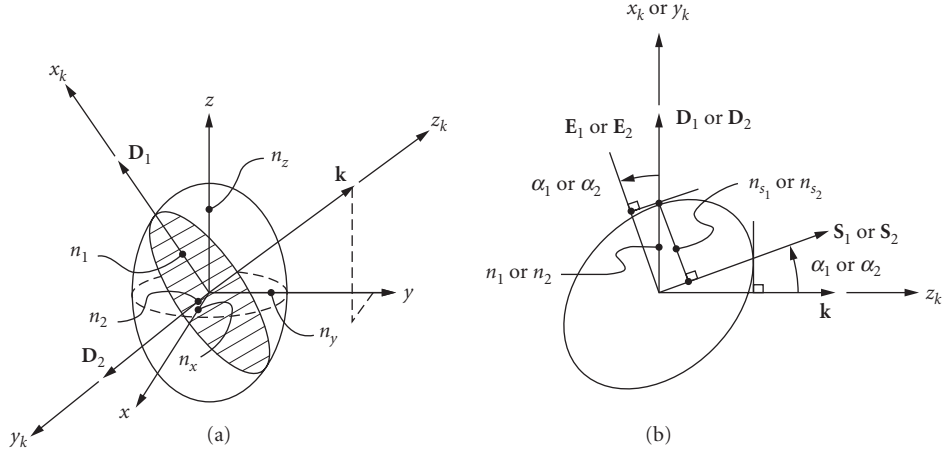


FIGURE 3 (a) The index ellipsoid cross section (crosshatched) that is normal to the wavevector $\mathbf{k}$ has the shape of an ellipse. The major and minor axes of this ellipse represent the directions of the allowed polarizations $\mathbf{D}_1$ and $\mathbf{D}_2$ and (b) for each eigenpolarization (1 or 2) the vectors $\mathbf{D}$, $\mathbf{E}$, $\mathbf{S}$, and $\mathbf{k}$ are coplanar.²¹

orthogonal vibration directions of $\mathbf{D}$ for that particular direction of propagation. The lengths of these axes correspond to the phase velocity refractive indices. They are, therefore, referred to as the “fast” and “slow” axes. Figure 3b illustrates the field relationships with respect to the index ellipsoid. The line in the $(\mathbf{k}, \mathbf{D}_i)$ plane ($i=1$ or 2) that is tangent to the ellipsoid at $\mathbf{D}_i$ is parallel to the ray vector $\mathbf{S}_i$; the electric field $\mathbf{E}_i$ also lies in the $(\mathbf{k}, \mathbf{D}_i)$ plane and is normal to $\mathbf{S}_i$. The line length denoted by n_{s_i} gives the ray velocity as $v_{s_i} = c/n_{s_i}$ for $\mathbf{S}_i$. The same relationships hold for either vibration, $\mathbf{D}_1$ or $\mathbf{D}_2$.

In the general ellipsoid for a biaxial crystal there are two cross sections passing through the center that are circular. The normals to these cross sections are called the *optic axes* (denoted in Fig. 2 in a nonprincipal coordinate system), and they are coplanar and symmetric about the z principal axis in the x - z plane. The angle ϑ of an optic axis with respect to the z axis in the x - z plane is

$$\tan \vartheta = \frac{n_z}{n_x} \sqrt{\frac{n_y^2 - n_x^2}{n_z^2 - n_y^2}} \quad (4)$$

The phase velocities for $\mathbf{D}_1$ and $\mathbf{D}_2$ are equal for these two directions: $v_1 = v_2 = c/n_y$. In an ellipsoid of revolution for a uniaxial crystal, there is one circular cross section perpendicular to the z principal axis. Therefore, the z axis is the optic axis, and $\vartheta = 0^\circ$ in this case.

7.4 THE ELECTRO-OPTIC EFFECT

At an atomic level, an electric field applied to certain crystals causes a redistribution of bond charges and possibly a slight deformation of the crystal lattice.¹⁶ In general, these alterations are not isotropic; that is, the changes vary with direction in the crystal. Therefore, the dielectric tensor and its inverse, the impermeability tensor, change accordingly. The linear electro-optic effect, or Pockels, is a change in the impermeability tensor elements that is proportional to the magnitude of the externally applied electric field. Only crystals lacking a center of symmetry or macroscopically ordered dipolar molecules exhibit the Pockels effect. On the other hand, all materials, including amorphous materials and liquids, exhibit a quadratic (Kerr) electro-optic effect. The changes in the impermeability tensor elements are proportional to the square of the applied field. When the linear effect is present, it generally dominates over the quadratic effect.

Application of the electric field induces changes in the index ellipsoid and the impermeability tensor of Eq. (3b) according to

$$\mathbf{X}^T \left[\frac{1}{n^2} + \Delta \frac{1}{n^2} \right] \mathbf{X} = 1 \quad (5)$$

where the perturbation is

$$\Delta \left[\frac{1}{n^2} \right] = \left[\frac{1}{n^2} \right]_{E \neq 0} - \left[\frac{1}{n^2} \right]_{E=0} = \sum_{k=1}^3 r_{ijk} E_k + \sum_{k=1}^3 \sum_{l=1}^3 s_{ijkl} E_k E_l \quad (6)$$

Since n is dimensionless, and the applied electric field components are in units of V/m, the units of the linear r_{ijk} coefficients are in m/V and the quadratic coefficients s_{ijkl} are in m²/V².

The linear electro-optic effect is represented by a third rank tensor r_{ijk} with $3^3 = 27$ independent elements, that if written out in full form, will form the shape of a cube. The permutation symmetry of this tensor is $r_{ijk} = r_{ikj}$, $i, j, k = 1, 2, 3$ and this symmetry reduces the number of independent elements to 18.²² Therefore, the tensor can be represented in contracted notation by a 6×3 matrix; that is, $r_{ijk} \Rightarrow r_{ij}$, $i = 1, \dots, 6$ and $j = 1, 2, 3$. The first suffix is the same in both the tensor and the contracted matrix notation, but the second two tensor suffixes are replaced by a single suffix according to the following relation.

Tensor notation	11	22	33	23,32	31,13	12,21
Matrix notation	1	2	3	4	5	6

Generally, the r_{ij} coefficients have very little dispersion in the optical transparent region of a crystal.²³ The electro-optic coefficient matrices for all crystal classes are given in Table 1. References 16, 23, 24, and 25, among others, contain extensive tables of numerical values for indices and electro-optic coefficients for different materials.

The quadratic electro-optic effect is represented by a fourth rank tensor s_{ijkl} . The permutation symmetry of this tensor is $s_{ijkl} = s_{jikl} = s_{ijlk}$, $i, j, k, l = 1, 2, 3$. The tensor can be represented by a 6×6 matrix; that is, $s_{ijkl} \Rightarrow s_{kl}$, $k, l = 1, \dots, 6$. The quadratic electro-optic coefficient matrices for all crystal classes are given in Table 2. Reference 16 contains a table of quadratic electro-optic coefficients for several materials.

The Linear Electro-Optic Effect

An electric field applied in a general direction to a noncentrosymmetric crystal produces a linear change in the constants $(1/n^2)_i$ due to the linear electro-optic effect according to

$$\Delta(1/n^2)_i = \sum_j r_{ij} E_j \quad \begin{matrix} i = 1, \dots, 6 \\ j = x, y, z = 1, 2, 3 \end{matrix} \quad (7)$$

where r_{ij} is the ij th element of the linear electro-optic tensor in contracted notation. In matrix form Eq. (7) is

$$\begin{pmatrix} \Delta(1/n^2)_1 \\ \Delta(1/n^2)_2 \\ \Delta(1/n^2)_3 \\ \Delta(1/n^2)_4 \\ \Delta(1/n^2)_5 \\ \Delta(1/n^2)_6 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \\ r_{61} & r_{62} & r_{63} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \quad (8)$$

TABLE 1 The Linear Electro-Optic Coefficient Matrices in Contracted Form for All Crystal Symmetry Classes¹⁶Centrosymmetric ($\bar{1}$, 2/m, mmm, 4/m, 4/mmm, $\bar{3}$, $\bar{3}2/m$, 6/mmm, m3, m3m):

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Triclinic:

$$1^* \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \\ r_{61} & r_{62} & r_{63} \end{pmatrix}$$

Cubic:

$$\begin{matrix} \bar{4}3m, 23 & 432 \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & r_{41} & 0 \\ 0 & 0 & r_{41} \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

Monoclinic:

$$\begin{matrix} 2(2 \parallel x_2) & 2(2 \parallel x_3) \\ \begin{pmatrix} 0 & r_{12} & 0 \\ 0 & r_{22} & 0 \\ 0 & r_{32} & 0 \\ r_{41} & 0 & r_{43} \\ 0 & r_{52} & 0 \\ r_{61} & 0 & r_{63} \end{pmatrix} & \begin{pmatrix} 0 & 0 & r_{13} \\ 0 & 0 & r_{23} \\ 0 & 0 & r_{33} \\ r_{41} & r_{42} & 0 \\ r_{51} & r_{52} & 0 \\ 0 & 0 & r_{63} \end{pmatrix} \\ m(m \perp x_2) & m(m \perp x_3) \\ \begin{pmatrix} r_{11} & 0 & r_{13} \\ r_{21} & 0 & r_{23} \\ r_{31} & 0 & r_{33} \\ 0 & r_{42} & 0 \\ r_{51} & 0 & r_{52} \\ 0 & r_{62} & 0 \end{pmatrix} & \begin{pmatrix} r_{11} & r_{12} & 0 \\ r_{21} & r_{22} & 0 \\ r_{31} & r_{32} & 0 \\ 0 & 0 & r_{43} \\ 0 & 0 & r_{53} \\ r_{61} & r_{62} & 0 \end{pmatrix} \end{matrix}$$

Tetragonal:

$$\begin{matrix} 4 & \bar{4} & 422 \\ \begin{pmatrix} 0 & 0 & r_{13} \\ 0 & 0 & r_{13} \\ 0 & 0 & r_{33} \\ r_{41} & r_{51} & 0 \\ r_{51} & -r_{41} & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & r_{13} \\ 0 & 0 & -r_{13} \\ 0 & 0 & 0 \\ r_{41} & -r_{51} & 0 \\ r_{51} & r_{41} & 0 \\ 0 & 0 & r_{63} \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & -r_{41} & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ 4mm & \bar{4}2m(2 \parallel x_1) \\ \begin{pmatrix} 0 & 0 & r_{13} \\ 0 & 0 & r_{13} \\ 0 & 0 & r_{33} \\ 0 & r_{51} & 0 \\ r_{51} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & r_{41} & 0 \\ 0 & 0 & r_{63} \end{pmatrix} \end{matrix}$$

Orthorhombic:

$$\begin{matrix} 222 & 2mm \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & r_{52} & 0 \\ 0 & 0 & r_{63} \end{pmatrix} & \begin{pmatrix} 0 & 0 & r_{13} \\ 0 & 0 & r_{23} \\ 0 & 0 & r_{33} \\ 0 & r_{42} & 0 \\ r_{51} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

Trigonal:

$$\begin{matrix} 3 & 32 \\ \begin{pmatrix} r_{11} & -r_{22} & r_{13} \\ -r_{11} & -r_{22} & r_{13} \\ 0 & 0 & r_{33} \\ r_{41} & r_{51} & 0 \\ r_{51} & -r_{41} & 0 \\ -r_{22} & -r_{11} & 0 \end{pmatrix} & \begin{pmatrix} r_{11} & 0 & 0 \\ -r_{11} & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & -r_{41} & 0 \\ 0 & -r_{11} & 0 \end{pmatrix} \end{matrix}$$

Hexagonal:

$$\begin{matrix} 6 & 6mm & 622 \\ \begin{pmatrix} 0 & 0 & r_{13} \\ 0 & 0 & r_{13} \\ 0 & 0 & r_{33} \\ r_{41} & r_{51} & 0 \\ r_{51} & -r_{41} & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & r_{13} \\ 0 & 0 & r_{13} \\ 0 & 0 & r_{33} \\ 0 & r_{51} & 0 \\ r_{51} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & -r_{41} & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \bar{6} & \bar{6}m2(m \perp x_1) & \bar{6}m2(m \perp x_2) \\ \begin{pmatrix} r_{11} & -r_{22} & 0 \\ -r_{11} & r_{22} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -r_{22} & r_{11} & 0 \end{pmatrix} & \begin{pmatrix} 0 & -r_{22} & 0 \\ 0 & r_{22} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -r_{22} & 0 & 0 \end{pmatrix} & \begin{pmatrix} r_{11} & 0 & 0 \\ -r_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -r_{11} & 0 \end{pmatrix} \\ 3m(m \perp x_1) & 3m(m \perp x_2) \\ \begin{pmatrix} 0 & -r_{22} & r_{13} \\ 0 & r_{22} & r_{13} \\ 0 & 0 & r_{33} \\ 0 & r_{51} & 0 \\ r_{51} & 0 & 0 \\ -r_{22} & 0 & 0 \end{pmatrix} & \begin{pmatrix} r_{11} & 0 & r_{13} \\ -r_{11} & 0 & r_{13} \\ 0 & 0 & r_{33} \\ 0 & r_{51} & 0 \\ r_{51} & 0 & 0 \\ 0 & -r_{11} & 0 \end{pmatrix} \end{matrix}$$

^aThe symbol over each matrix is the conventional symmetry-group designation.

TABLE 2 The Quadratic Electro-Optic Coefficient Matrices in Contracted Form for All Crystal Symmetry Classes¹⁶

Triclinic:

$$1, \bar{1} \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} & s_{36} \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} & s_{46} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & s_{56} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_{66} \end{pmatrix}$$

Monoclinic:

$$2, m, 2/m \begin{pmatrix} s_{11} & s_{12} & s_{13} & 0 & s_{15} & 0 \\ s_{21} & s_{22} & s_{23} & 0 & s_{25} & 0 \\ s_{31} & s_{32} & s_{33} & 0 & s_{35} & 0 \\ 0 & 0 & 0 & s_{44} & 0 & s_{46} \\ s_{51} & s_{52} & s_{53} & 0 & s_{55} & 0 \\ 0 & 0 & 0 & s_{64} & 0 & s_{66} \end{pmatrix}$$

Orthorhombic:

$$2mm, 222, mmm \begin{pmatrix} s_{11} & s_{12} & s_{13} & 0 & 0 & 0 \\ s_{21} & s_{22} & s_{23} & 0 & 0 & 0 \\ s_{31} & s_{32} & s_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{66} \end{pmatrix}$$

Tetragonal:

$$4, \bar{4}, 4/m \begin{pmatrix} s_{11} & s_{12} & s_{13} & 0 & 0 & s_{16} \\ s_{12} & s_{11} & s_{13} & 0 & 0 & -s_{16} \\ s_{31} & s_{31} & s_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & s_{45} & 0 \\ 0 & 0 & 0 & -s_{45} & s_{44} & 0 \\ s_{61} & -s_{61} & 0 & 0 & 0 & s_{66} \end{pmatrix}$$

$$422, 4mm, \bar{4}2m, 4/mm \begin{pmatrix} s_{11} & s_{12} & s_{13} & 0 & 0 & 0 \\ s_{12} & s_{11} & s_{13} & 0 & 0 & 0 \\ s_{31} & s_{31} & s_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{66} \end{pmatrix}$$

Hexagonal:

$$6, \bar{6}, 6/m \begin{pmatrix} s_{11} & s_{12} & s_{13} & 0 & 0 & -s_{61} \\ s_{12} & s_{11} & s_{13} & 0 & 0 & s_{61} \\ s_{31} & s_{31} & s_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & s_{45} & 0 \\ 0 & 0 & 0 & -s_{45} & s_{44} & 0 \\ s_{61} & -s_{61} & 0 & 0 & 0 & \frac{1}{2}(s_{11} - s_{12}) \end{pmatrix}$$

$$622, 6mm, \bar{6}m2, 6/mmm \begin{pmatrix} s_{11} & s_{12} & s_{13} & 0 & 0 & 0 \\ s_{12} & s_{11} & s_{13} & 0 & 0 & 0 \\ s_{31} & s_{31} & s_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}(s_{11} - s_{12}) \end{pmatrix}$$

Cubic:

$$23, m\bar{3} \begin{pmatrix} s_{11} & s_{12} & s_{13} & 0 & 0 & 0 \\ s_{13} & s_{11} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{13} & s_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{44} \end{pmatrix}$$

(Continued)

TABLE 2 The Quadratic Electro-optic Coefficient Matrices in Contracted Form for All Crystal Symmetry Classes¹⁶ (Continued)

Trigonal:	Cubic:
$ \begin{array}{c} 3, \bar{3} \\ \left(\begin{array}{cccccc} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & -s_{61} \\ s_{12} & s_{11} & s_{13} & -s_{14} & -s_{15} & s_{61} \\ s_{31} & s_{31} & s_{33} & 0 & 0 & 0 \\ s_{41} & -s_{41} & 0 & s_{44} & s_{45} & -s_{51} \\ s_{51} & -s_{51} & 0 & -s_{45} & s_{44} & s_{41} \\ s_{61} & -s_{61} & 0 & -s_{15} & s_{14} & \frac{1}{2}(s_{11} - s_{12}) \end{array} \right) \end{array} $	$ \begin{array}{c} 432, m\bar{3}m, \bar{4}3m \\ \left(\begin{array}{cccccc} s_{11} & s_{12} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{11} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{12} & s_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{44} \end{array} \right) \end{array} $
$ \begin{array}{c} 32, 3m, \bar{3}m \\ \left(\begin{array}{cccccc} s_{11} & s_{12} & s_{13} & s_{14} & 0 & 0 \\ s_{12} & s_{11} & s_{13} & -s_{14} & 0 & 0 \\ s_{13} & s_{13} & s_{33} & 0 & 0 & 0 \\ s_{41} & -s_{41} & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{44} & s_{41} \\ 0 & 0 & 0 & 0 & s_{14} & \frac{1}{2}(s_{11} - s_{12}) \end{array} \right) \end{array} $	$ \begin{array}{c} \text{Isotropic:} \\ \left(\begin{array}{cccccc} s_{11} & s_{12} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{11} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{12} & s_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2}(s_{11} - s_{12}) & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2}(s_{11} - s_{12}) & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}(s_{11} - s_{12}) \end{array} \right) \end{array} $

E_x , E_y , and E_z are the components of the applied electric field in principal coordinates. The magnitude of $\Delta(1/n^2)$ is typically on the order of less than 10^{-5} . Therefore, these changes are mathematically referred to as perturbations. The new impermeability tensor $[1/n^2]'$ in the presence of an applied electric field is no longer diagonal in the reference principal dielectric axes system. It is given by

$$[1/n^2]' = \begin{pmatrix} 1/n_x^2 + \Delta(1/n^2)_1 & \Delta(1/n^2)_6 & \Delta(1/n^2)_5 \\ \Delta(1/n^2)_6 & 1/n_y^2 + \Delta(1/n^2)_2 & \Delta(1/n^2)_4 \\ \Delta(1/n^2)_5 & \Delta(1/n^2)_4 & 1/n_z^2 + \Delta(1/n^2)_3 \end{pmatrix} \quad (9)$$

and is determined by the unperturbed principal refractive indices, the electro-optic coefficients, and the direction of the applied field relative to the principal coordinate system. However, the field-induced perturbations are symmetric, so the symmetry of the tensor is not disturbed. The new index ellipsoid is now represented by

$$(1/n^2)'_1 x^2 + (1/n^2)'_2 y^2 + (1/n^2)'_3 z^2 + 2(1/n^2)'_4 yz + 2(1/n^2)'_5 xz + 2(1/n^2)'_6 xy = 1 \quad (10)$$

or equivalently, $\mathbf{X}^T [1/n^2]' \mathbf{X} = 1$, where $\mathbf{X} = [x \ y \ z]^T$.^{19,26} The presence of cross terms indicates that the ellipsoid is rotated and the lengths of the principal dielectric axes are changed. Determining the new orientation and shape of the ellipsoid requires that $[1/n^2]'$ be diagonalized, thus determining its eigenvalues and eigenvectors. After diagonalization, in a suitably rotated new coordinate system $\mathbf{X}' = [x' \ y' \ z']$ the perturbed ellipsoid will then be represented by a square sum:

$$\frac{x'^2}{n_{x'}^2} + \frac{y'^2}{n_{y'}^2} + \frac{z'^2}{n_{z'}^2} = 1 \quad (11)$$

The eigenvalues of $[1/n^2]'$ are $1/n_{x'}^2$, $1/n_{y'}^2$, $1/n_{z'}^2$. The corresponding eigenvectors are $\mathbf{x}' = [x_{x'} \ y_{x'} \ z_{x'}]^T$, and $\mathbf{y}' = [x_{y'} \ y_{y'} \ z_{y'}]^T$, respectively.

The Quadratic or Kerr Electro-Optic Effect

An electric field applied in a general direction to any crystal, centrosymmetric or noncentrosymmetric, produces a quadratic change in the constants $(1/n^2)_i$ due to the quadratic electro-optic effect according to

$$\begin{pmatrix} \Delta(1/n^2)_1 \\ \Delta(1/n^2)_2 \\ \Delta(1/n^2)_3 \\ \Delta(1/n^2)_4 \\ \Delta(1/n^2)_5 \\ \Delta(1/n^2)_6 \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} & s_{36} \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} & s_{46} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & s_{56} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_{66} \end{pmatrix} \begin{pmatrix} E_x^2 \\ E_y^2 \\ E_z^2 \\ E_y E_z \\ E_x E_z \\ E_x E_y \end{pmatrix} \quad (12)$$

E_x , E_y , and E_z are the components of the applied electric field in principal coordinates. The perturbed impermeability tensor and the new index ellipsoid have the same form as Eqs. (9) and (10).

Normally, there are two distinctions made when considering the Kerr effect: the ac Kerr effect and the dc Kerr effect. The induced changes in the optical properties of the material can occur as the result of a slowly varying applied electric field, or it can result from the electric field of the light itself. The former is the dc Kerr effect and the latter is the ac or optical Kerr effect. The dc Kerr effect is given by

$$\Delta n = \lambda K E^2 \quad (13)$$

where K is the Kerr constant in units of m/V^2 and λ is the freespace wavelength. Some polar liquids such as nitrobenzene ($\text{C}_6\text{H}_5\text{NO}_2$), which is poisonous, exhibit very large Kerr constants which are much greater than those of transparent crystals.²⁷ In contrast to the linear Pockels electro-optic effect, larger voltages are required for any significant Kerr modulation.

The ac or optical Kerr effect occurs when a very intense beam of light modulates the optical material. The optical Kerr effect is given by

$$n = n_o + n_2 I \quad (14)$$

which describes the intensity dependent refractive index n , where n_o is the unmodulated refractive index, n_2 is the second order nonlinear refractive index (m^2/W), and I is the intensity of the wave (W). Equation 14 is derived from the expression for the electric field induced polarization in a material as a function of the linear and nonlinear susceptibilities. This intensity dependent refractive index is used as the basis for Kerr-lens mode-locking of ultrashort lasers. It is also responsible for nonlinear effects of self-focusing and self-phase modulation.²⁸

A Mathematical Approach: The Jacobi Method

The analytical design and study of electro-optic modulators require robust mathematical techniques due to the small, anisotropic perturbations to the refractive index profile of a material. Especially with the newer organic crystals, polymers, and tailored nanostructured materials, the properties are often biaxial before and after applied voltages. The optimum modulator configuration may not be along principal axes. In addition, sensitivities in modulation characteristics of biaxial materials (natural and/or induced) can negatively impact something as simple as focusing a beam onto the material. Studying the electro-optic effect is basically an eigenvalue problem.

Although the eigenvalue problem is a familiar one, obtaining accurate solutions has been the subject of extensive study.^{29–32} A number of formalisms are suggested in the literature to address the specific problem of finding the new set of principal dielectric axes relative to the zero-field principal

dielectric axes. Most approaches, however, do not provide a consistent means of labeling the new axes. Also, some methods are highly susceptible to numerical instabilities when dealing with very small off-diagonal elements as in the case of the electro-optic effect. In contrast to other methods,^{15,22,26,30,33,34} a similarity transformation is an attractive approach for diagonalizing a symmetric matrix for the purpose of determining its eigenvalues and eigenvectors.^{21,29,30,32,35}

The Jacobi method utilizes the concepts of rigid-body rotation and the properties of ellipsoids to determine the principal axes and indices of a crystal by constructing a series of similarity transformations that consist of elementary plane rotations. The method produces accurate eigenvalues and orthogonal eigenvectors for matrices with very small off-diagonal elements, and it is a systematic procedure for ordering the solutions to provide consistent labeling of the principal axes.^{21,31} The sequence of transformations are applied to the perturbed index ellipsoid and convert from one set of orthogonal axes $\mathbf{X} = [x, y, z]$ to another set $[x', y', z']$, until a set of axes coincides with the new principal dielectric directions and the impermeability matrix is diagonalized. Since similarity is a transitive property, several transformation matrices can be multiplied to generate the desired cumulative matrix.^{21,29} Thus, the problem of determining the new principal axes and indices of refraction of the index ellipsoid in the presence of an external electric field is analogous to the problem of finding the cumulative transformation matrix $[a] = [a_m] \cdots [a_2][a_1]$ that will diagonalize the perturbed impermeability tensor. The transformation required matrix, $[a]$, is simply the product of the elementary plane rotation matrices multiplied in the order in which they are applied.

When plane rotations are applied to the matrix representation of tensors, the magnitude of a physical property can be evaluated in any arbitrary direction. When the matrix is transformed to diagonal form, the eigenvalues lie on the diagonal and the eigenvectors are found in the rows or columns of the corresponding transformation matrices. Specifically, a symmetric matrix $[\mathbf{A}]$ can be reduced to diagonal form by the transformation $[a][\mathbf{A}][a]^T = [\lambda]$, where $[\lambda]$ is a 3×3 diagonal matrix and $[a]$ is the orthogonal transformation matrix. Since the eigenvalues of $[\mathbf{A}]$ are preserved under similarity transformation, they lie on the diagonal of $[\lambda]$, as in Eq. (1).

In terms of the index ellipsoid, first recall that the perturbed index ellipsoid in the original (zero field) coordinate system is $\mathbf{X}^T [1/n^2]' \mathbf{X} = 1$, where $[1/n^2]'$ is given by Eq. (9). A suitable matrix, $[a]$, will relate the “new” principal axes $\mathbf{X}'$ of the perturbed ellipsoid to the “old” coordinate system; that is, $\mathbf{X}' = [a]\mathbf{X}$, or $\mathbf{X} = [a]^T \mathbf{X}'$. Substituting these relationships into the index ellipsoid results in

$$\begin{aligned} ([a]^T \mathbf{X}')^T [1/n^2]' [a]^T \mathbf{X}' &= 1 \\ \mathbf{X}'^T [a] [1/n^2]' [a]^T \mathbf{X}' &= 1 \\ \mathbf{X}'^T [1/n^2]'' \mathbf{X}' &= 1 \end{aligned} \quad (15)$$

where $[a][1/n^2]'[a]^T = [1/n^2]''$ is the diagonalized impermeability matrix in the new coordinate system and $[a]$ is the cumulative transformation matrix.

Using the Jacobi method, each simple elementary plane rotation that is applied at each step will zero an off-diagonal element of the impermeability tensor. The goal is to produce a diagonal matrix by minimizing the norm of the off-diagonal elements to within a desired level of accuracy. If m transformations are required, each step is represented by

$$[1/n^2]_m = [a_m][1/n^2]_{m-1}[a_m]^T \quad (16)$$

To determine the form of each elementary plane rotation, the Jacobi method begins by first selecting the largest off-diagonal element $(1/n^2)_{ij}$ and executing a rotation in the (i, j) plane, $i < j$, so as to zero that element. The required rotation angle Ω is given by

$$\tan(2\Omega) = \left(\frac{2(1/n^2)_{ij}}{(1/n^2)_{ii} - (1/n^2)_{jj}} \right) \quad i, j = 1, 2, 3 \quad (17)$$

For example, if the largest off-diagonal element is $(1/n_{12}^2) = (1/n_{21}^2)$, then the plane rotation is represented by

$$[a] = \begin{pmatrix} \cos \Omega & \sin \Omega & 0 \\ -\sin \Omega & \cos \Omega & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (18)$$

which is a counter clockwise rotation about the three-axis. If $(1/n^2)_{ii} = (1/n^2)_{jj}$, which can occur in isotropic and uniaxial crystals, then $|\Omega|$ is taken to be 45° , and its sign is taken to be the same as the sign of $(1/n^2)_{ij}$. The impermeability matrix elements are updated with the following equations, which are calculated from the transformation of Eq. (15):

$$\begin{pmatrix} (1/n_{11}^2)' & 0 & (1/n_{13}^2)' \\ 0 & (1/n_{22}^2)' & (1/n_{23}^2)' \\ (1/n_{13}^2)' & (1/n_{23}^2)' & (1/n_{33}^2)' \end{pmatrix} = \begin{pmatrix} \cos \Omega & \sin \Omega & 0 \\ -\sin \Omega & \cos \Omega & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/n_{11}^2 & 1/n_{12}^2 & 1/n_{13}^2 \\ 1/n_{12}^2 & 1/n_{22}^2 & 1/n_{23}^2 \\ 1/n_{13}^2 & 1/n_{23}^2 & 1/n_{33}^2 \end{pmatrix} \begin{pmatrix} \cos \Omega & -\sin \Omega & 0 \\ \sin \Omega & \cos \Omega & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (19)$$

Once the new elements are determined, the next iteration step is performed, selecting the new largest off-diagonal element and repeating the procedure with another suitable rotation matrix. The process is terminated when all of the off-diagonal elements are reduced below the desired level (typically 10^{-10}). The next step is to determine the cumulative transformation matrix $[a]$. One way is to multiply the plane rotation matrices in order, either as $[a] = [a_m] \cdots [a_2][a_1]$ or equivalently for the transpose of $[a]$ as

$$[a]^T = [a_1]^T [a_2]^T \cdots [a_m]^T \quad (20)$$

The set of Euler angles, which also defines the orientation of a rigid body, can be obtained from the cumulative transformation matrix $[a]$.^{36,37} These angles are given in the Appendix. Several examples for using the Jacobi method are given in Ref. 21.

Determining the Eigenpolarizations and Phase Velocity Indices of Refraction

After the perturbed impermeability matrix is diagonalized, the polarization directions of the two allowed linear orthogonal waves $\mathbf{D}_1$ and $\mathbf{D}_2$ that propagate independently for a given wavevector direction $\mathbf{k}$ can be determined along with their respective phase velocity refractive indices $n_{x''}$ and $n_{y''}$. These waves are the only two that can propagate with unchanging orientation for the given wavevector direction. Figure 4a depicts these axes for a crystal in the absence of an applied field. Figure 4b depicts the x''' and y''' axes, which define the fast and slow axes, when an electric field is applied in a direction so as to reorient the index ellipsoid. The applied field, in general, rotates the allowed polarization directions in the plane perpendicular to the direction of phase propagation as shown in Fig. 4b. Determining these “eigenpolarizations,” that is, $\mathbf{D}_1$, $\mathbf{D}_2$, n_1 , and n_2 , is also an eigenvalue problem.

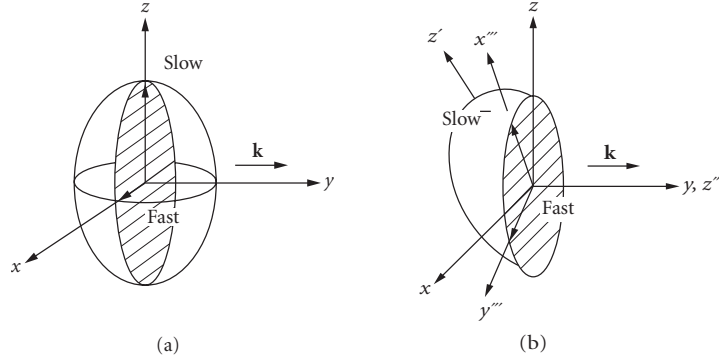


FIGURE 4 (a) The cross-section ellipse for a wave propagating along the y principal axis is shown with no field applied to the crystal; (b) with an applied electric field the index ellipsoid is reoriented, and the eigenpolarizations in the plane transverse to $\mathbf{k}$ are rotated, indicated by x'' and y'' .

The perturbed index ellipsoid resulting from an external field was given by Eq. (10) in the original principal-axis coordinate system. For simplicity, the coefficients may be relabeled as

$$Ax^2 + By^2 + Cz^2 + 2Fyz + 2Gxz + 2Hxy = 1$$

or

$$\mathbf{X}^T \begin{pmatrix} A & H & G \\ H & B & F \\ G & F & C \end{pmatrix} \mathbf{X} = 1 \quad (21)$$

where x , y , and z represent the original dielectric axes with no applied field and $\mathbf{X}^T = [x \ y \ z]$. However, before the eigenpolarizations can be determined, the direction of light propagation, $\mathbf{k}$, through the material whose index ellipsoid has been perturbed, must be defined. The problem is then to determine the allowed eigenpolarizations and phase velocity refractive indices associated with this direction of propagation. The optical wavevector direction $\mathbf{k}$ is conveniently specified by the spherical coordinates angles θ_k and ϕ_k in the (x, y, z) coordinate system as shown in Fig. 5. Given $\mathbf{k}$, the cross

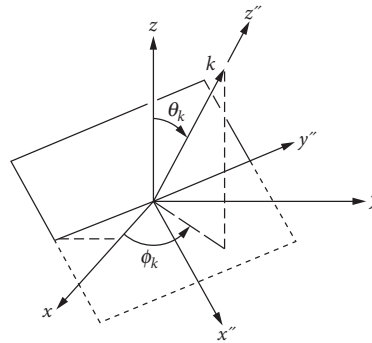


FIGURE 5 The coordinate system (x'', y'', z'') of the wavevector $\mathbf{k}$ is defined with its angular relationship (ϕ_k, θ_k) with respect to the unperturbed principal dielectric axes coordinate system (x, y, z) .²¹

section ellipse through the center of the perturbed ellipsoid of Eq. (21) may be drawn. The directions of the semiaxes of this ellipse represent the fast and slow polarization directions of the two waves $\mathbf{D}_1$ and $\mathbf{D}_2$ that propagate independently. The lengths of the semiaxes are the phase velocity indices of refraction. The problem is to determine the new polarization directions x''' of $\mathbf{D}_1$ and y''' of $\mathbf{D}_2$ relative to the (x, y, z) axes and the corresponding new indices of refraction $n_{x''}$ and $n_{y''}$.

The first step is to do a transformation from the (x, y, z) (lab or principal axis) coordinate system to a coordinate system (x'', y'', z'') aligned with the direction of phase propagation. In this example, (x'', y'', z'') is chosen such that $z'' \parallel \mathbf{k}$, and x'' is lying in the (z, z') plane. The (x'', y'', z'') system is, of course, different from the (x', y', z') perturbed principal axes system. Using the spherical coordinate angles of $\mathbf{k}$, the (x'', y'', z'') system may be produced first by a counterclockwise rotation ϕ_k about the z axis followed by a counterclockwise rotation θ_k about y'' as shown in Fig. 5. This transformation is described by $\mathbf{X}'' = [\mathbf{a}]\mathbf{X}$, or $[\mathbf{a}]^T \mathbf{X}'' = \mathbf{X}$ and is explicitly,

$$\mathbf{X} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos \phi_k & -\sin \phi_k & 0 \\ \sin \phi_k & \cos \phi_k & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \theta_k & 0 & -\sin \theta_k \\ 0 & 1 & 0 \\ \sin \theta_k & 0 & \cos \theta_k \end{pmatrix} \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} \quad (22)$$

The equation for the cross section ellipse normal to $\mathbf{k}$ is determined by substituting Eq. (22) into Eq. (21) and setting $z'' = 0$ or by matrix substitution as follows:

$$([\mathbf{a}]^T \mathbf{X}'')^T [1/n^2]' ([\mathbf{a}]^T \mathbf{X}'') = 1$$

$$\mathbf{X}''^T \underbrace{[\mathbf{a}] [1/n^2]' [\mathbf{a}]^T}_{[1/n^2]''} \mathbf{X}'' = 1 \quad (23)$$

which results in

$$\mathbf{X}''^T [1/n^2]'' \mathbf{X}'' = 1$$

or

$$\begin{pmatrix} x'' & y'' & z'' \end{pmatrix} \begin{pmatrix} A'' & H'' & G'' \\ H'' & B'' & F'' \\ G'' & F'' & C'' \end{pmatrix} \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} = 1 \quad (24)$$

The coefficients of the cross section ellipse equation described by Eq. (24), with z'' set to zero, are used to determine the eigenpolarization directions and the associated phase velocity refractive indices for the chosen direction of propagation. The cross section ellipse normal to the wavevector direction $\mathbf{k} \parallel z''$ is represented by the 2×2 submatrix of $[1/n^2]''$:

$$\begin{pmatrix} x'' & y'' \end{pmatrix} \begin{pmatrix} A'' & H'' \\ H'' & B'' \end{pmatrix} \begin{pmatrix} x'' \\ y'' \end{pmatrix} = A'' x''^2 + B'' y''^2 + 2H'' x'' y'' = 1 \quad (25)$$

The polarization angle β_1 of x''' ($\mathbf{D}_1$) with respect to x'' , as shown in Fig. 6, is given by

$$\beta_1 = \frac{1}{2} \tan^{-1} \left[\frac{2H''}{(A'' - B'')} \right] \quad (26)$$

The polarization angle β_2 of y''' ($\mathbf{D}_2$) with respect to x'' is $\beta_1 + \pi/2$. The axes are related by a plane rotation $\mathbf{X}''' = [\mathbf{a}_{\beta_1}]\mathbf{X}''$ or

$$\begin{pmatrix} x''' \\ y''' \end{pmatrix} = \begin{pmatrix} \cos \beta_1 & -\sin \beta_1 \\ \sin \beta_1 & \cos \beta_1 \end{pmatrix} \begin{pmatrix} x'' \\ y'' \end{pmatrix} \quad (27)$$

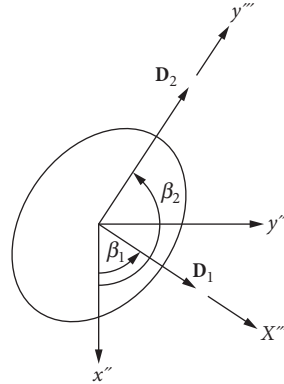


FIGURE 6 The polarization axes (x''' , y''') are the *fast* and *slow* axes and are shown relative to the (x'' , y'') axes of the wavevector coordinate system. The wavevector $\mathbf{k}$ and the axes z'' and z''' are normal to the plane of the figure.²¹

The refractive indices, $n_{x''}$ and $n_{y''}$, may be found by performing one more rotation in the plane of the ellipse normal to $\mathbf{k}$, using the angle β_1 or β_2 and the rotation of Eq. (27). The result is a new matrix that is diagonalized,

$$[1/n_2]''' = \begin{pmatrix} 1/n_{x'''} & 0 \\ 0 & 1/n_{y'''} \end{pmatrix} = [a_{\beta_1}]^T [1/n_2]'' [a_{\beta_1}] \quad (28)$$

The larger index corresponds to the slow axis and the smaller index to the fast axis.

7.5 MODULATOR DEVICES

An electro-optic modulator is a device with operation based on an electrically induced change in index of refraction or change in natural birefringence. Depending on the device configuration, the following properties of the light wave can be varied in a controlled way: phase, polarization, amplitude, frequency, or direction of propagation. The device is typically designed for optimum performance at a single wavelength, with some degradation in performance with wideband or multimode lasers.^{16,38,39}

Electro-optic devices can be used in analog or digital modulation formats. The choice is dictated by the system requirements and the characteristics of available components (optical fibers, sources/detectors, etc.). Analog modulation requires large signal-to-noise ratios (SNR), thereby limiting its use to narrow-bandwidth, short-distance applications. Digital modulation, on the other hand, is more applicable to large-bandwidth, medium to long distance systems.^{38,39}

Device Geometries

A bulk electro-optic modulator can be classified as one of two types, *longitudinal* or *transverse*, depending on how the voltage is applied relative to the direction of light propagation in the device. Basically a bulk modulator consists of an electro-optic crystal sandwiched between a pair of electrodes and, therefore, can be modeled as a capacitor. In general, the input and output faces are parallel for the beam to undergo a uniform phase shift over the beam cross section.¹⁶ Waveguide modulators are discussed later in the section “Waveguide or Integrated-Optic Modulators” and have a variety of electrode configurations that are analogous to longitudinal and transverse orientations, although the distinction is not as well defined.

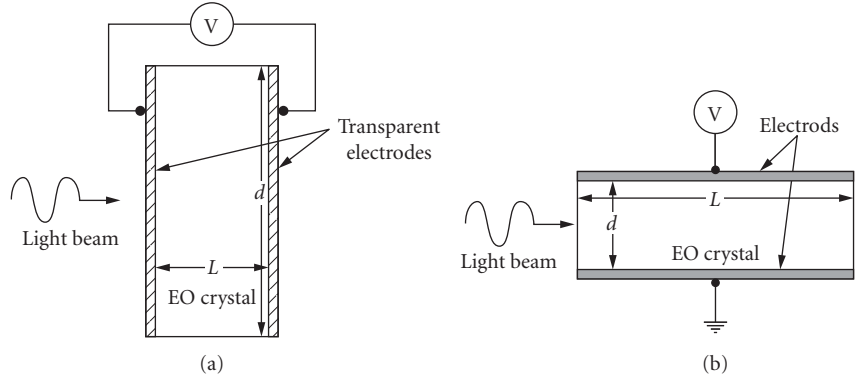


FIGURE 7 (a) A longitudinal electro-optic modulator has the voltage applied parallel to the direction of light propagation and (b) a transverse modulator has the voltage applied perpendicular to the direction of light propagation.¹⁶

In the bulk longitudinal configuration, the voltage is applied parallel to the wavevector direction as shown in Fig. 7a.^{16,25,40–43} The electrodes must be transparent to the light either by the choice of material used for them (metal-oxide coatings of SnO, InO, or CdO) or by leaving a small aperture at their center at each end of the electro-optic crystal.^{25,41–43} The ratio of the crystal length L to the electrode separation b is defined as the *aspect ratio*. For this configuration $b = L$, and, therefore, the aspect ratio is always unity. The magnitude of the applied electric field inside the crystal is $E = V/L$. The induced phase shift is proportional to V and the wavelength λ of the light but not the physical dimensions of the device. Therefore, for longitudinal modulators, the required magnitude of the applied electric field for a desired degree of modulation cannot be reduced by changing the aspect ratio, and it increases with wavelength. However, these modulators can have a large acceptance area and are useful if the light beam has a large cross-sectional area.

In the transverse configuration, the voltage is applied perpendicular to the direction of light propagation as shown in Fig. 7b.^{16,40–43} The electrodes do not obstruct the light as it passes through the crystal. For this case, the aspect ratio can be very large. The magnitude of the applied electric field is $E = V/d$, ($b = d$), and d can be reduced to increase E for a given applied voltage, thereby increasing the aspect ratio L/b . The induced phase shift is inversely proportional to the aspect ratio; therefore, the voltage necessary to achieve a desired degree of modulation can be greatly reduced. Furthermore,

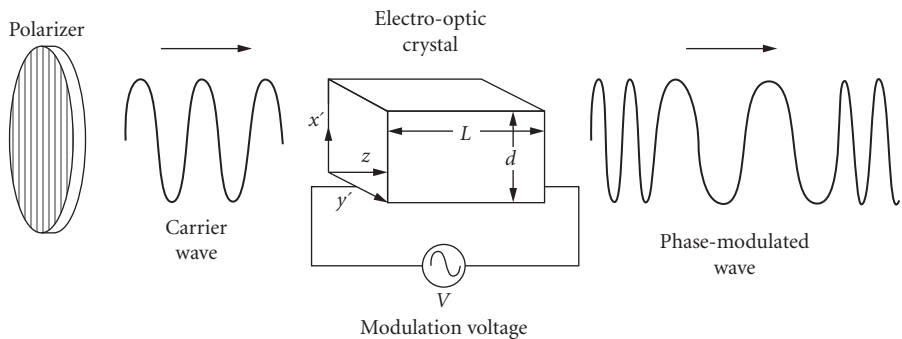


FIGURE 8 A longitudinal phase modulator is shown with the light polarized along the new x' principal axis when the modulation voltage V is applied.¹⁶

the interaction length can be long for a given field strength. However, the transverse dimension d is limited by the increase in capacitance, which affects the modulation bandwidth or speed of the device, and by diffraction for a given length L , since a beam with finite cross section diverges as it propagates.^{16,41,44}

Bulk Modulators

The modulation of phase, polarization, amplitude, frequency, and position of light can be implemented using an electro-optic bulk modulator with polarizers and passive birefringent elements. Three assumptions are made in this section. First, the modulating field is *uniform* throughout the length of the crystal; the change in index or birefringence is uniform unless otherwise stated. Second, the modulation voltage is dc or very low radian frequency ω_m ($\omega_m \ll 2\pi/\tau$); the light experiences the same induced Δn during its transit time τ through the crystal of length L , and the capacitance is negligible. Finally, light propagation is taken to be along a principal axis, before and after the voltage is applied; therefore, the equations for the eigenpolarizations are presented in terms of the *optical electric field* $\mathbf{E}$, rather than the *displacement vector* $\mathbf{D}$, which is a common practice in various optical references. For other general configurations, the equations should be expressed in terms of the eigenpolarizations $\mathbf{D}_1$ and $\mathbf{D}_2$. However, the electric field will determine the direction of energy flow, which is generally not in the same direction as the wavevector. References 16 and 41, among others, provide examples of modulator devices using potassium dihydrogen phosphate (KDP), lithium niobate (LiNbO_3), lithium tantalate (LiTaO_3), gallium arsenide (GaAs), and barium titanate (BaTiO_3).

Phase Modulator A light wave can be phase modulated, without change in polarization or intensity, using an electro-optic crystal and an input polarizer in the proper configuration. This is the simplest electro-optic modulator. As an example, consider a longitudinal device that is made of a LiNbO_3 crystal as shown in Fig. 8, with the voltage applied in the z direction. In general, an applied voltage V will rotate the principal axes in the crystal cross section. However, for phase modulation, the input polarizer must be aligned parallel to one of the principal axes that will have a preserved orientation when the voltage is on or off. The LiNbO_3 crystal is uniaxial with symmetry 3m, and it is a common material for electro-optic modulators. Substituting into Eq. (7) results in

$$\begin{pmatrix} 0 & -r_{22} & r_{13} \\ 0 & r_{22} & r_{13} \\ 0 & 0 & r_{33} \\ 0 & r_{51} & 0 \\ r_{51} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ E_z \end{pmatrix} = \begin{pmatrix} r_{13}E_z \\ r_{13}E_z \\ r_{33}E_z \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (29)$$

Figure 8 indicates the polarizer along x' provides an input optical electric field $E_{\text{in}}(t) = E_x \cos \omega t$. The voltage is applied in the z direction and thus, has only one component, E_z . In this case, the perturbed index ellipsoid will be

$$x^2(1/n_o^2 + r_{13}E_z) + y^2(1/n_o^2 + r_{13}E_z) + z^2(1/n_e^2 + r_{33}E_z) = 1 \quad (30)$$

where $n_x = n_y = n_o$ is the ordinary refractive index and $n_z = n_e$ is the extraordinary refractive index of this uniaxial crystal. The orientation of the principal axes remains unchanged, but the lengths of the axes, and hence the phase velocity refractive indices, have been modified by the applied electric field. The optical wave at the output of the crystal at $z = L$ is

$$E_{\text{out}}(t) = E_x \cos(\omega t - \phi) \quad (31)$$

where

$$\phi = \frac{2\pi}{\lambda}(n_{x'} + \Delta n_{x'})L = \phi_o + \Delta\phi_{x'} \quad (32)$$

is the total phase shift consisting of a natural phase term $\phi_o = (2\pi/\lambda)Ln_{x'}$, with $n_{x'}$ being the unperturbed index in the x' ($= x$, in this example) direction, and an electrically induced phase term $\Delta\phi_{x'} = (2\pi/\lambda)L\Delta n_{x'}$ for a polarization along x' . The new x' axis has length $2n_{x'}$, where

$$(1/n_{x'})^2 = (1/n_o^2) + r_{13}E_z \quad (33)$$

Using the derivative $\Delta(1/n^2)' = -2n^{-3}\Delta n$ results in

$$\Delta n_{x'}' \approx (-1/2)n_o^3 r_{13} E_z \quad (34)$$

an approximation often seen in literature.

For a longitudinal modulator the applied electric field is $E_z = V/L$, and the induced phase shift is $\Delta\phi_{x'} = \pi/\lambda n_{x'}^3 r V$, which is independent of L and is linearly related to V . For a transverse modulator $E = V/d$ and the induced phase shift is $\Delta\phi_{x'} = \pi/\lambda n_{x'}^3 r V(L/d)$, which is a function of the aspect ratio L/d and V . The voltage that would produce an induced phase shift of $\Delta\phi_{x'} = \pi$ is the *half-wave voltage*. The half-wave voltage is $V_\pi = \lambda/n_{x'}^3 r$ for a longitudinal modulator and $V_\pi = (\lambda/n_{x'}^3 r)(d/L)$ for a transverse modulator.

Whenever possible, it is desired to take advantage of the largest electro-optic coefficient. In LiNbO_3 , $r_{13} = 10$ and $r_{33} = 32$. To use r_{33} , the optical input signal should be polarized in the z direction.

If a dc applied voltage is used, one of two possibilities is required for a crystal and its orientation. The first possibility is a crystal having principal axes which will not rotate with applied voltage V ; an example is LiNbO_3 with V applied in the z direction and an input polarization along the $x' = x$ axis propagating along $z' = z$. The second possibility is a crystal having a characteristic plane, that is, a plane exhibiting a constant refractive index perpendicular to the direction of propagation. If a field is applied such that the axes rotate in this plane, the input wave must be polarized along one of the new principal axes. Therefore, it will always be polarized along a principal axis, whether the voltage is on or off. An example is KDP with V along the z axis and an input wave polarized along the new principal axis x' and propagating along $z' = z$. Phase modulation is then achieved by turning the voltage on and off.

If the applied modulation voltage is sinusoidal in time ($V = V_m \sin \omega_m t$), the corresponding electric field can be represented by $E = E_m \sin \omega_m t$.

The magnitude of the field varies only with time, not space; it is a stationary wave applied in the same direction for all the time. In other words, this time-varying voltage signal is to be distinguished from a traveling wave voltage which will be discussed in the next section. In this case,

$$\begin{aligned} \phi &= \left(\frac{2\pi}{\lambda} \right) \left(n_{x'} - \frac{1}{2} n_{x'}^3 r E_m \sin \omega_m t \right) L \\ &= \left(\frac{2\pi}{\lambda} \right) n_{x'} L - \delta \sin \omega_m t \end{aligned} \quad (35)$$

The parameter $\delta = (\pi/\lambda)n_{x'}^3 r E_m L = \pi V_m/V_\pi$, where V_π is the half-wave voltage for a given configuration, is the *phase modulation index* or *depth-of-phase modulation*. By neglecting the constant phase term ϕ_o , applying the identity $\cos(\delta \sin \omega_m t) + j \sin(\delta \sin \omega_m t) = \exp[j\delta \sin \omega_m t] = \sum_{l=-\infty}^{\infty} J_l(\delta) \exp[jl\omega_m t]$, and equating the real and imaginary parts, the output light wave becomes

$$\begin{aligned} E_o(t) &= E_i [j_0(\delta) \cos \omega t + J_1(\delta) \cos(\omega + \omega_m)t - J_1(\delta) \cos(\omega - \omega_m)t \\ &\quad + J_2(\delta) \cos(\omega + 2\omega_m)t + J_2(\delta) \cos(\omega - 2\omega_m)t + \dots] \end{aligned} \quad (36)$$

The output consists of components at frequencies ω and $(\omega + n\omega_m)$, $n = \pm 1, \pm 2, \dots$. For no modulation, $\delta = 0$ and $J_0(0) = 1$, $J_n(0) = 0$ for $n \neq 0$ and $E_o(t) = E_i \cos \omega t = E_{x'}(t)$.¹⁶ For $\delta \approx 2.4048$, $J_0(\delta) = 0$ all the power is transferred to harmonic frequencies.⁴¹ For the case of small modulation index, $\delta \ll 1$, most of the power resides in the carrier frequency at ω and a small amount resides in the first two sidebands at frequencies $\omega \pm \omega_m$. This condition makes phase modulators useful in laser mode-locking. Increasing δ results in the presence of more sidebands.

Polarization Modulator (Dynamic Retardation) Polarization modulation involves the coherent addition of two orthogonal waves, resulting in a change of the input polarization state at the output. As with a phase modulator, the basic components for an electro-optic polarization modulator (or dynamic retardation plate or polarization state converter) is an electro-optic crystal and an input polarizer. The crystal and applied voltage V (dc assumed) are configured to produce dynamically the fast and slow axes in the crystal cross section. In this case, however, the polarizer is positioned such that the input light wave is decomposed equally into the two orthogonal linear eigenpolarizations along these axes as shown in Fig. 9. If the light is polarized along the x axis and propagates along the z principal axis, for example, the propagating fields are

$$\begin{aligned} E_{x'} &= E_o \cos[\omega t - (2\pi/\lambda)n_x z] \\ E_{y'} &= E_o \cos[\omega t - (2\pi/\lambda)n_y z] \end{aligned} \quad (37)$$

where the fast and slow axes are x' and y' . The corresponding refractive indices are

$$\begin{aligned} n_{x'} &\approx n_x - \frac{1}{2}r_x n_x^3 E = n_x - \Delta n_x \\ n_{y'} &\approx n_y - \frac{1}{2}r_y n_y^3 E = n_y - \Delta n_y \end{aligned} \quad (38)$$

where n_x and n_y are the indices in the absence of an applied field and r_x, r_y are the appropriate electro-optic coefficients for the material being used and the orientation of the applied voltage. As the two polarizations propagate at different speeds through the crystal, a phase difference (relative phase) or *retardation* Γ evolves between them as a function of length:

$$\begin{aligned} \Gamma &= \frac{2\pi}{\lambda}(n_{x'} - n_{y'})L \\ &= \frac{2\pi}{\lambda}(n_x - n_y)L - \frac{\pi}{\lambda}(r_x n_x^3 - r_y n_y^3)EL = \Gamma_o + \Gamma_i \end{aligned} \quad (39)$$

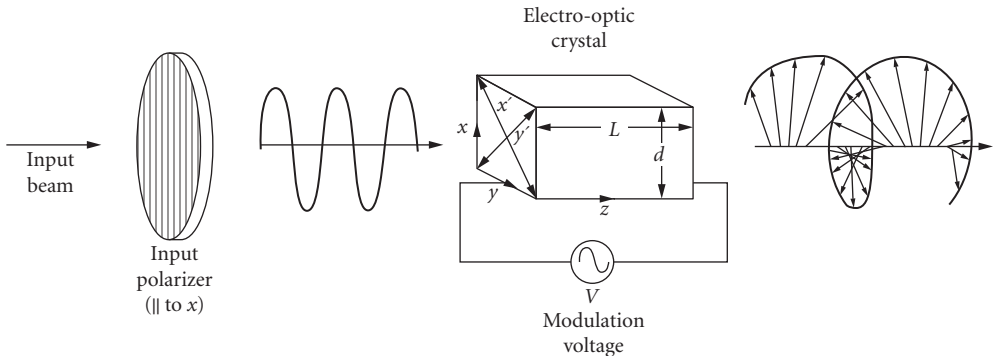


FIGURE 9 A longitudinal polarization modulator is shown with the input polarizer oriented along the x principal axis at 45° with respect to the perturbed x' and y' axes.

where Γ_o is the natural phase retardation in the absence of an applied voltage and Γ_i is the induced retardation linearly related to V .

For a longitudinal modulator the applied electric field is $E = V/L$, and the induced retardation is $\Gamma_i = (\pi/\lambda)(r_x n_y^3 - r_x n_x^3)V$, which is independent of L and linearly related to V .

For a transverse modulator $E = V/d$, and the induced retardation is $\Gamma_i = (\pi/\lambda)(r_y n_y^3 - r_x n_x^3)V(L/d)$, which is dependent on the aspect ratio L/d and V .

The optical fields at the output can be expressed in terms of Γ :

$$\begin{aligned} E_{x'} &= \cos \omega t \\ E_{y'} &= \cos(\omega t - \Gamma) \end{aligned} \quad (40)$$

Therefore, the desired output polarization is obtained by applying the appropriate voltage magnitude. Figure 10 illustrates the evolution of the polarization state as a function of propagation distance z . In terms of an active device, Fig. 10 also can be interpreted as a change in polarization state as a function of applied voltage for fixed length. The eigenpolarizations $E_{x'}$ and $E_{y'}$ are in phase at $z = 0$. They have the same frequency but different wavelengths. Light from one polarization gradually couples into the other. In the absence of natural birefringence, $n_x - n_y = 0$, the voltage that would produce a retardation of $\Gamma = \Gamma_i = \pi$, such that a vertical polarization input becomes a horizontal polarization

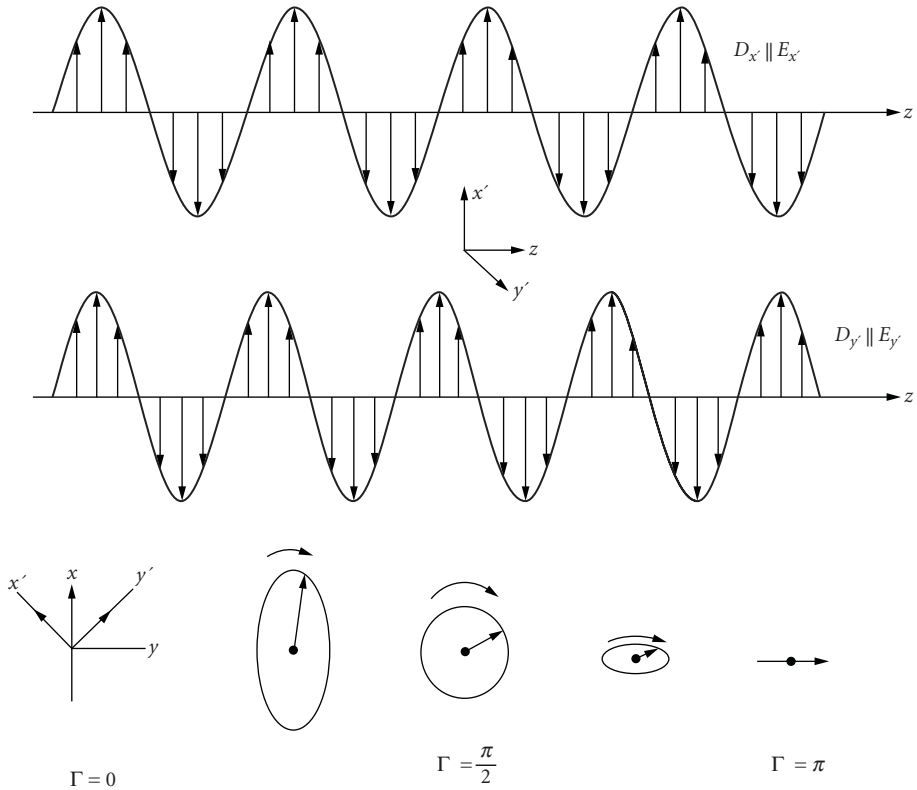


FIGURE 10 The polarization state of an input vertical linear polarization is shown as a function of crystal length L or applied voltage V . The retardation $\Gamma = \pi$ for a given length L_π in a passive $\lambda/2$ wave plate or applied voltage V_π in an electro-optic polarization modulator.¹⁶

output, is the *half-wave voltage* V_π . For a longitudinal modulator $V_\pi = \lambda/(r_x n_x^3 - r_y n_y^3)$, which is independent of L . For a transverse modulator $V_\pi = \lambda/(r_x n_x^3 - r_y n_y^3)(d/L)$, which is dependent on the aspect ratio L/d . The total retardation in terms of V_π (calculated assuming no birefringence) is

$$\Gamma = \Gamma_o + \pi \left(\frac{v}{V_\pi} \right) \quad (41)$$

To cancel the effect of natural birefringence, the phase retardation Γ_o can be made a multiple of 2π by slightly polishing the crystal to adjust the length or by applying a bias voltage. If birefringence is present, an effective V_π can be calculated that would give a total retardation of $\Gamma = \pi$.

To achieve polarization modulation, a birefringence must exist in the crystal cross section. If the cross section is a characteristic plane, then the input polarization propagates through the crystal unchanged when $V = 0$. If an applied voltage causes the axes to rotate 45° in this cross section with respect to the input polarization, as in Fig. 9, then the input will decompose into two equal components and change polarization state at the output. If the cross section has natural birefringence, then the input polarization state will change with $V = 0$ as well as with an applied voltage.

Amplitude Modulator The intensity (optical energy) of a light wave can be modulated in several ways. Some possibilities include using (1) a dynamic retarder configuration with a crossed polarizer at the output, (2) a dynamic retarder configuration with a parallel polarizer at the output, (3) a phase modulator configuration in a branch of a Mach-Zehnder interferometer, or (4) a dynamic retarder with push-pull electrodes. The intensity modulator parameter of interest is the *transmission* $T = I_o/I_i$, the ratio of output to input intensity.

An intensity modulator constructed using a dynamic retarder with crossed polarizers is shown in Fig. 11. The transmission for this modulator is

$$T(V) = \sin^2\left(\frac{\Gamma}{2}\right) = \sin^2\left(\frac{\Gamma_o}{2} + \frac{\pi V}{2V_\pi}\right) \quad (42)$$

For linear modulation, where the output is a replica of the modulating voltage signal, a fixed bias of $\Gamma_o = \pi/2$ must be introduced either by placing an additional phase retarder, a $\lambda/4$ wave plate (Fig. 11), at the output of the electro-optic crystal or by applying an additional dc voltage of $V_\pi/2$. This bias produces a transmission of $T = 0.5$ in the absence of a modulating voltage. If the crystal cross section has natural birefringence, then a variable compensator (Babinet-Soleil) or a voltage less than $V_\pi/2$ must be used to tune the birefringence to give a fixed retardation of $\pi/2$.

For a sinusoidal modulation voltage $V = V_m \sin \omega_m t$, the retardation at the output of the crystal, including the bias, is

$$\Gamma = \Gamma_o + \Gamma_i = \frac{\pi}{2} + \Gamma_m \sin \omega_m t \quad (43)$$

where $\Gamma_m = \pi V_m / V_\pi$ is the *amplitude modulation index* or *depth-of-amplitude modulation* and V_π is the half-wave voltage. The transmission becomes

$$\begin{aligned} T(V) &= \sin^2\left(\frac{\pi}{4} + \frac{\Gamma_m}{2} \sin \omega_m t\right) \\ &= \frac{1}{2} \left[1 - \cos\left(\frac{\pi}{2} + \Gamma_m \sin \omega_m t\right) \right] \end{aligned} \quad (44)$$

If the modulation voltage is small ($V_m \ll 1$), then the modulation depth is small ($\Gamma_m \ll 1$) and

$$T(V) = \frac{1}{2} [1 + \Gamma_m \sin \omega_m t] \quad (45)$$

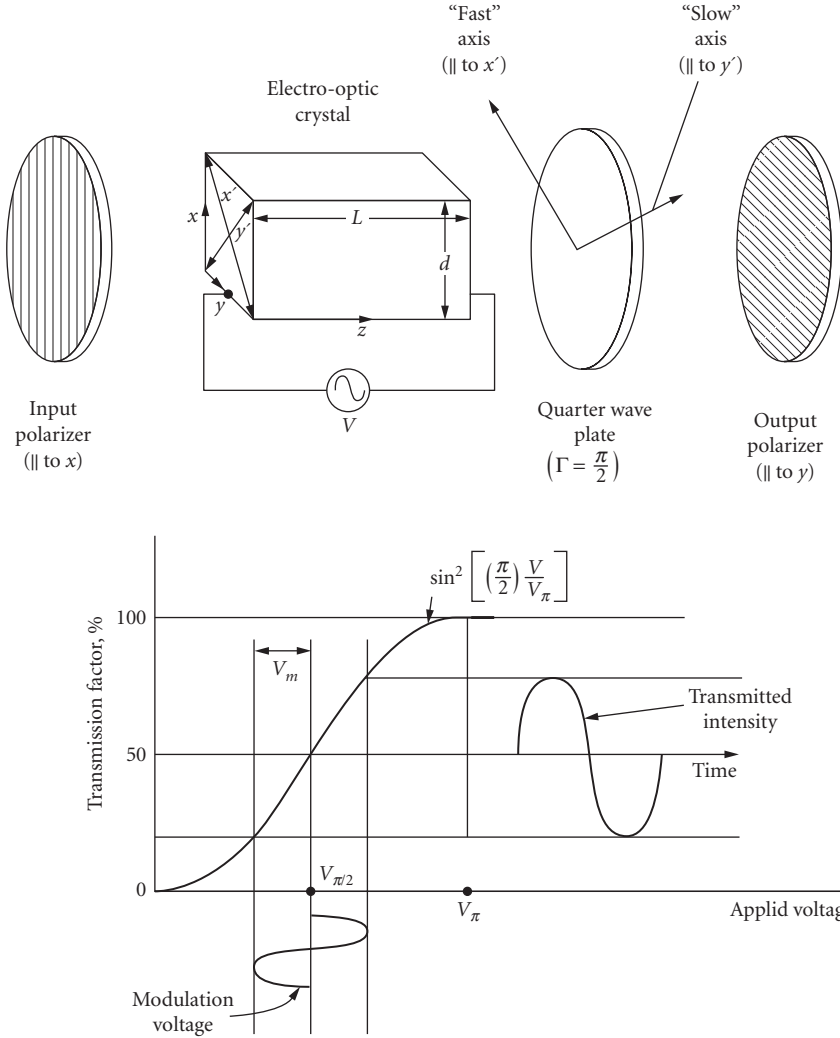


FIGURE 11 A longitudinal intensity modulator is shown using crossed polarizers with the input polarization along the x principal axis. A $\lambda/4$ wave plate is used as a bias to produce linear modulation.¹⁶

Therefore, the transmission or output intensity is linearly related to the modulating voltage. If the signal is large, then the output intensity becomes distorted and higher-order odd harmonics appear.¹⁶ The dynamic retarder with parallel polarizers has a transmission of¹⁶

$$\begin{aligned}
 t &= \cos^2\left(\frac{\Gamma}{2}\right) = \cos^2\left(\frac{\pi}{4} + \frac{\Gamma_m}{2} \sin \omega_m t\right) \\
 &= \frac{1}{2} \left[1 + \cos\left(\frac{\pi}{2} + \Gamma_m \sin \omega_m t\right) \right]
 \end{aligned} \tag{46}$$

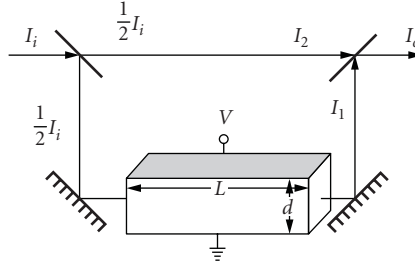


FIGURE 12 An intensity modulator is shown implementing a Mach-Zehnder interferometer configuration with a phase modulator in one branch.⁴²

For small modulation, $T(V) = 1/2[-\Gamma_m \sin \omega_m t]$ and again, the output is a replica of the modulating voltage.

Similarly, the output of a Mach-Zehnder interferometer is given by

$$I_o = I_1 + I_2 + \frac{1}{2}[I_i \cos \Gamma_o + I_i] = I_i \cos^2 \left(\frac{\Gamma_o}{2} \right) \quad (47)$$

where Γ_o is the relative phase shift between the two branches. An intensity modulator is produced by placing a phase modulator in one branch as shown in Fig. 12. The total retardation is $\Gamma = \Gamma_o + \Gamma_p$, as before. The transmission is

$$T = \frac{I_o}{I_i} = \cos^2 \left(\frac{\Gamma}{2} \right) \quad (48)$$

The push-pull modulator is based on the Mach-Zehnder interferometer. In this case, a phase modulator is placed in each branch with opposite polarity voltages applied to the arms; the phase modulators are driven 180° out of phase. This configuration requires lower drive voltages and provides a shorter transit time for the light for a defined degree of modulation.⁴⁵

Frequency Modulator In frequency modulation a shift or deviation in the frequency by ω_d from the optical carrier instantaneous frequency ω is desired. One approach to achieve a shift in frequency is to use an intensity modulator configuration of an electro-optic crystal between left- and right-hand circular polarizers. The electrodes on the modulator must be designed to produce an applied circular electric field.⁴⁶

A crystal and its orientation are selected such that there is no birefringence in the crystal cross section at zero voltage. When a circular electric field with frequency ω_m , is applied; however, a birefringence is induced in the crystal cross section, and the induced principal axes rotate with angular velocity $\omega_m/2$ in the opposite sense with respect to the modulating field. The relative rotation between the axes and the modulating field creates a frequency shift in the optical electric field as it propagates through the crystal.

An example of such a device⁴⁶ is shown in Fig. 13. There are two sets of electrodes in the transverse configuration. The applied voltages are 90° out of phase to produce a left circular modulating field:

$$\begin{aligned} E_x &= E_m \cos \omega_m t \\ E_y &= E_m \sin \omega_m t \end{aligned} \quad (49)$$

The principal axes in the crystal cross section rotate through an angle

$$\beta_1(t) = -\frac{1}{2}(\omega_m t + \Phi) \quad (50)$$

where Φ is a fixed angle that depends on the electro-optic coefficients r_{ij} of the crystal and not on the electric field magnitude.

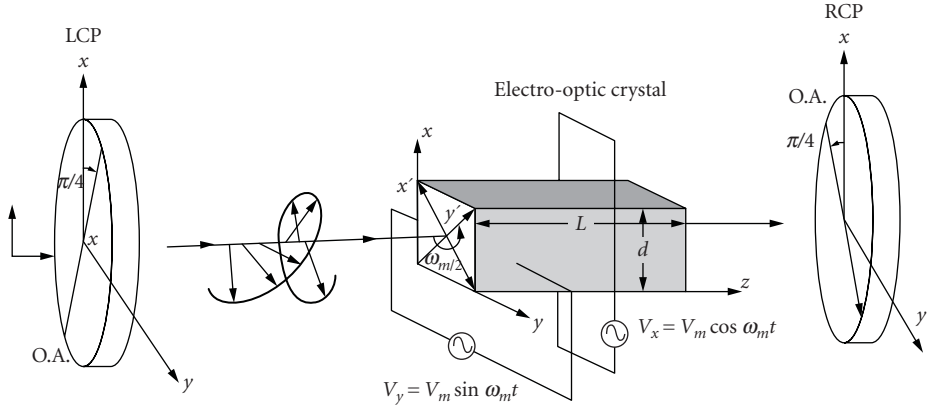


FIGURE 13 A frequency modulator using a phase modulator with two pairs of transverse electrodes set 90° out of phase to produce a circular applied electric field. The phase modulator is placed between left and right circular polarizers to create a frequency deviation in the output optical field.

The input optical wave (after the left circular polarizer) has field components

$$\begin{aligned} E_{i_x} &= E_i \cos \omega t \\ E_{i_y} &= E_i \sin \omega t \end{aligned} \quad (51)$$

The induced retardation $\Gamma_i = \Gamma = (2\pi/\lambda)\Delta nL$ is independent of time, since Δn is constant although the principal axes are rotating at constant angular velocity. The optical field components along the induced principal axes at the output of the crystal are

$$\begin{aligned} E_{o_{x'}} &= E_i \cos \left(\omega t - \beta_1 + \frac{\Gamma}{2} \right) \\ E_{o_{y'}} &= E_i \sin \left(\omega t - \beta_1 + \frac{\Gamma}{2} \right) \end{aligned} \quad (52)$$

In terms of the original stationary x, y axes, the optical field components at the output are

$$\begin{aligned} E_{o_x} &= E_i \cos(\Gamma/2) \cos \omega t - E_i \sin(\Gamma/2) \sin[(\omega + \omega_m)t + \Phi] \\ E_{o_y} &= E_i \cos(\Gamma/2) \sin \omega t - E_i \sin(\Gamma/2) \cos[(\omega + \omega_m)t + \Phi] \end{aligned} \quad (53)$$

The first terms in E_{o_x} and E_{o_y} represent left circular polarization at the original optical frequency ω with constant amplitude $E_i \cos(\Gamma/2)$ and phase independent of the rotating principal axes (that is, of β_1). The second terms represent right circular polarization at frequency $(\omega + \omega_m)$ with constant amplitude $E_i \sin(\Gamma/2)$ and phase proportional to $2\beta_1 = \omega_m t$. Therefore, the frequency shift ω_d is the modulation frequency ω_m . If the retardation $\Gamma = \pi$, the frequency of the light has complete deviation to $(\omega + \omega_m)$. If Γ is very small, the component optical fields at frequency $(\omega + \omega_m)$ are linearly related to Γ and therefore, to the applied voltage.

A shift in frequency to $(\omega - \omega_m)$ is obtained if the optical and the applied modulating electric fields rotate in the opposite sense.

Scanners The position of an optical beam can be changed dynamically by using an electro-optic deflecting device or scanner. Analog scanners are based on refraction phenomena: (1) refraction at a dielectric interface (prism) and (2) refraction by an index gradient that exists perpendicular to the direction of light propagation. Digital scanners or switches are based on birefringence.

One of the most important parameters characterizing the performance of a scanner is its resolution, the number of independent resolvable spots it can scan, which is limited by the diffraction occurring at the aperture of the device. The Rayleigh criterion states that two spots are just resolved when the angular displacement of the beam $\Delta\varphi$ is equal to the half-angle divergence θ due to diffraction.¹⁵ Therefore, the total number of resolvable spots is given by the ratio of the total deflection angle φ to the half-angle divergence θ

$$N = \frac{\varphi}{\theta} \quad (54)$$

The half-angle divergence is $\theta = \varpi/w$, where w is the beamwidth, $\varpi = 1$ for a rectangular beam of uniform intensity, $\varpi = 1.22$ for a circular beam of uniform intensity, and $\varpi = 1.27$ for a beam of Gaussian intensity distribution.⁴⁷

An analog scanner can be constructed by a prism of electro-optic material with electrodes on the crystal faces as shown in Fig. 14. The resolution is maximum in this isosceles-shaped prism when the beam is transmitted through at the minimum deviation angle and is^{47,48}

$$N = \Delta n \left(\frac{l}{\varpi\lambda} \right) \left(\frac{w}{W} \right) \quad (55)$$

where l is the base length of the prism and w/W is the ratio of the beamwidth to input aperture of the prism. This type of device typically requires a voltage much higher than the half-wave voltage V_π to resolve N spots.⁴²

An analog scanner based on a gradient index of refraction⁴⁹ is shown in Fig. 15. The voltage is applied to a crystal such that the change in index is linear with distance perpendicular to the direction of light propagation; that is, $n(x) = n_o + (2\Delta n/W)x$, where W is the width of the crystal. For linear gradient and small refraction angles, the wavefront remains planar. The (small) deflection angle of the ray after propagating a distance L in the crystal is approximated to be⁴⁷

$$\varphi = L \frac{dn}{dx} \quad (56)$$

and the resolution is

$$N = \frac{\varphi}{\theta} = 2\Delta n \left(\frac{L}{\varpi\lambda} \right) \left(\frac{w}{W} \right) \quad (57)$$

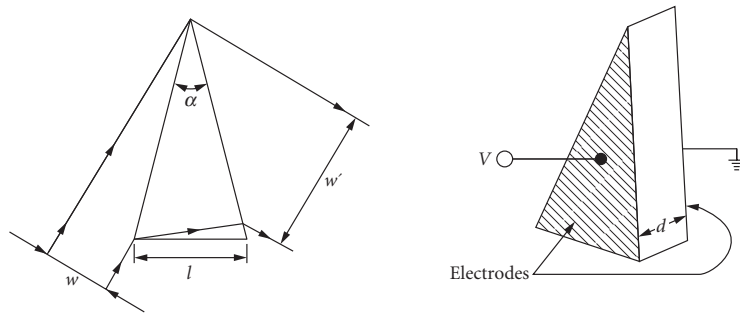


FIGURE 14 An analog scanner can be constructed using an isosceles-shaped prism with the beam transmitted at the minimum deviation angle.⁴⁸

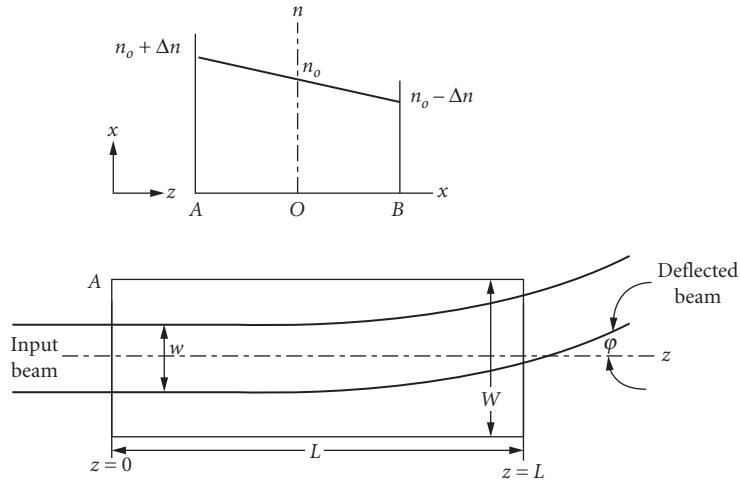


FIGURE 15 An analog scanner based on an index gradient is shown.⁴⁹

A large crystal length L is needed to obtain appreciable deflection, since Δn is very small ($\sim 10^{-4}$). Laser beams, however, are very narrow to make such a device practical.

Digital light deflection can be implemented with a number of *binary units*,^{42,50} each of which consists of a polarizer followed by a birefringent crystal (or discriminator) as shown in Fig. 16. The polarizer is oriented to give a horizontal polarization input to the birefringent crystal, such as calcite, and passes through undeflected. With an applied voltage, the principal axes rotate 45° and the input is then decomposed into orthogonal components of equal amplitude. The voltage

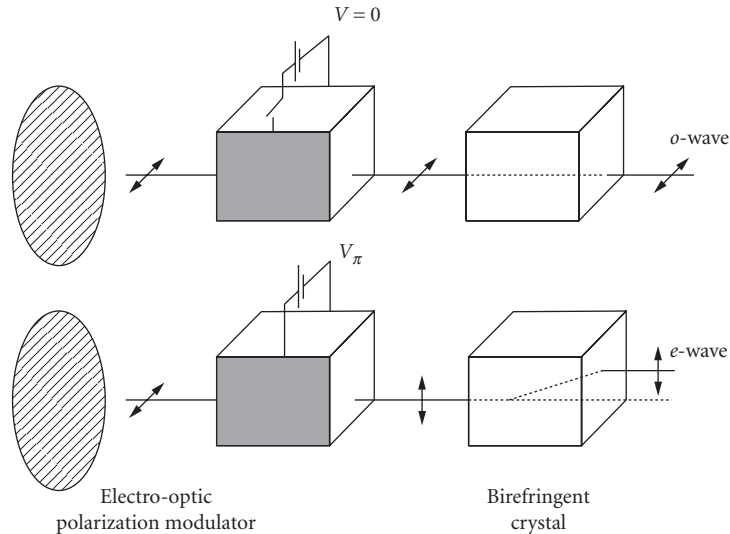


FIGURE 16 The component of a digital scanner is a binary unit. It consists of a polarizer modulator followed by a birefringent crystal which serves as a discriminator.⁴²

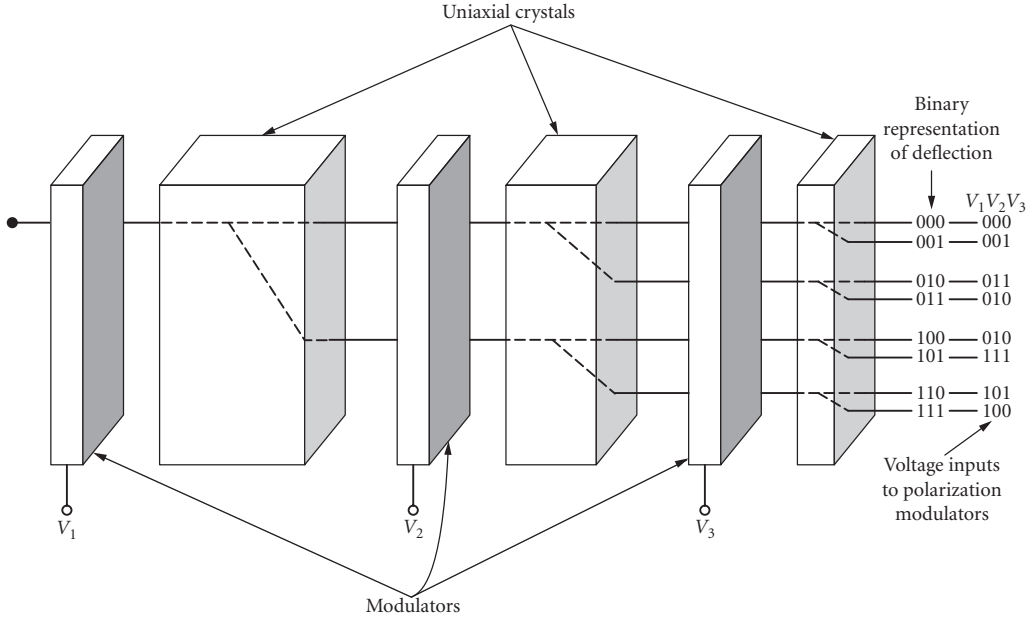


FIGURE 17 A three-stage digital scanner produces 2^3 possible deflected positions at the output.⁵⁰

magnitude is set to produce a retardation of π , thereby rotating the polarization by 90° . The vertical polarization then enters the birefringent crystal which deflects it.^{47,51}

The number of binary units n would produce a deflector of 2^n possible deflected positions at the output. An example of a three-stage deflector is shown in Fig. 17.⁵⁰⁻⁵² The on-off states of the three voltages determine at what position the output beam would be deflected. For example, if all three voltages are off, the input polarization remains horizontal through the system and is undeflected at the output. However, if all three voltages are on, the horizontal input is rotated 90° , becoming vertical after the first polarization modulator, and is deflected. The polarization is rotated 90° after the second modulator, becoming horizontal, and therefore, propagates straight through the birefringent crystal. Finally, the polarization is rotated 90° after the third modulator, becoming vertical, and is deflected. The corresponding output represents a binary five, that is, 101.

Traveling Wave Modulators The bandwidth of bulk modulators is limited by the transit time of the light through the crystal compared to the modulation frequency. The transit time of the light is the time for it to pass through the crystal

$$\tau = \frac{nL}{c} \quad (58)$$

where n is the index seen by the light. This parameter has no relevance for modulation frequencies $\omega_m \ll 2\pi/\tau$; the modulation field appears uniform in the crystal at very low frequencies. A rule of thumb for defining the limiting frequency such that τ can be neglected is that the length of the crystal be less than one-tenth the wavelength of the modulating field⁵³ or $L \ll 2\pi c/\omega_m \sqrt{\epsilon}$.⁴¹ The electro-optic crystal is modeled as a lumped capacitor at low frequencies.

As the modulating frequency becomes larger, the transit time must be taken into account in evaluating modulator performance. The modulation electric field has the form $E = E_m \sin \omega_m t$, and the optical phase can no longer follow the time-varying index of refraction adiabatically. The result is a

reduction in the modulation index parameters, δ for phase modulation and Γ_m for amplitude modulation, by a factor^{16,41,44}

$$\rho = \frac{\sin\left(\frac{1}{2}\omega_m \tau\right)}{\frac{1}{2}\omega_m \tau} \quad (59)$$

Therefore, the phase modulation index at high frequencies becomes

$$\delta_{\text{RF}} = \delta \cdot \rho = \delta \cdot \left[\frac{\sin\left(\frac{1}{2}\omega_m \tau\right)}{\frac{1}{2}\omega_m \tau} \right] \quad (60)$$

and the amplitude modulation index at high frequencies becomes

$$\Gamma_{m_{\text{RF}}} = \Gamma_m \cdot \rho = \Gamma_m \cdot \left[\frac{\sin\left(\frac{1}{2}\omega_m \tau\right)}{\frac{1}{2}\omega_m \tau} \right] \quad (61)$$

If $\tau = 2\pi/\omega_m$ such that the transit time of the light is equal to the time period of the modulation signal, then there is no retardation; the retardation produced in the first half of the crystal is exactly canceled by the retardation produced in the second half.⁴¹ The maximum modulation frequency for a given crystal length L is determined by the allowable ρ parameter set by the designer.

The limitation of the transit time on the bandwidth of the modulator can be overcome by applying the voltage as a traveling wave, propagating collinearly with the optical wave. Figure 18 illustrates a transverse traveling wave configuration. The electrode is designed to be an extension of the driving transmission line to eliminate electrode charging time effects on the bandwidth. Therefore, the transit time problem is addressed by adjusting the phase velocity of the modulation signal to be equal to the phase velocity of the optical signal.^{16,41,44}

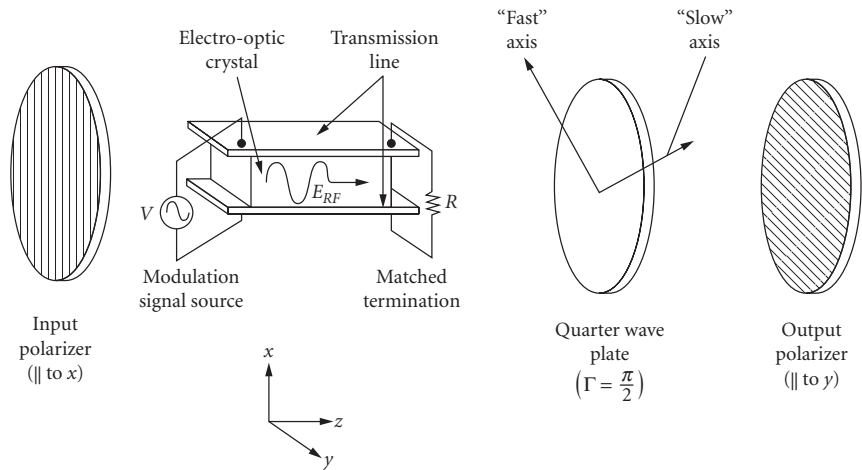


FIGURE 18 A transverse traveling wave modulator has a modulating voltage polarized in the same orientation as the input light and propagating collinearly.¹⁶

The applied modulation electric field has the form

$$E_{\text{RF}}(t, z) = E_m \sin(\omega_m t - k_m z) \quad (62)$$

for propagation along the z axis of a crystal. The direction of field vibration is the direction of the applied field, not the direction it is traveling, and is along x in Fig. 18. The parameter k_m is the wave-vector magnitude of the modulation field and is

$$k_m = \frac{\omega_m}{v_m} = \frac{\omega_m n_m}{c} \quad (63)$$

where v_m and $n_m = \sqrt{\epsilon}$ are the phase velocity and index of refraction of the modulating signal. A mismatch in the phase velocities of the modulating signal and optical wave will produce a reduction in the modulation index δ or Γ_m , by a factor

$$\rho_{\text{tw}} = \frac{\sin\left[\frac{\omega_m(n - n_m)L}{2c}\right]}{\frac{\omega_m(n - n_m)L}{2c}} = \frac{\sin(\Delta L)}{\Delta L} \quad (64)$$

In the case of amplitude modulation, Eq. (64) holds only if there is no natural birefringence in the cross section of the crystal. The eigenpolarization magnitudes are functions of time and space. Therefore, they satisfy the coupled wave equations, which are more complicated when birefringence is present.¹⁶

With no phase velocity mismatch, the phase modulation index is $\delta_{\text{RF}} = \delta$ and is linearly proportional to the crystal length L . Likewise, $\Gamma_{\text{RF}} = \Gamma_m$ for amplitude modulation. With a mismatch, the maximum possible phase modulation index is $\delta_{\text{RF,max}} = \delta/(\Delta L)$ likewise for the amplitude modulation index $\Gamma_{\text{RF,max}}$. The modulation index becomes a sinusoidal function of L . This maximum index can be achieved for crystal lengths $L = \pi/2\Delta, 3\pi/2\Delta$, and so on. The ratio n/n_m is approximately 1/2 for LiNbO_3 , producing a walk-off between the optical and modulation waves.⁴⁴ Therefore, for a given length L the modulation frequency is greatly affected by the velocity mismatch.

Waveguide or Integrated-Optic Modulators Electro-optic modulators are often fabricated as integrated optical devices. Because of the small dimensions of the waveguides, these modulators require lower drive voltages and operate at higher frequencies than bulk modulators. However, the disadvantages of integrated-optic modulators are that they have lower optical power handling capabilities; they present coupling difficulties when interfaced with free-space beams; and they are designed to operate at a specific wavelength or narrow spectral window rather than working with broad optical spectra. On the other hand, because of their size, they are compatible with single mode optical fibers, and as a result, they have globally transformed communications with optical data rates up to 40 Gbps. Integrated-optic modulators can also find usage as sensors; such as gyroscopes, voltage sensors, or other phase-sensitive applications.

Most integrated optic waveguide modulators are formed on crystalline wafer surfaces, such as LiNbO_3 , using standard photolithography techniques. The waveguides are formed by changing the refractive index in the guiding region using methods such as annealed proton exchange (APE) or titanium in-diffusion, and then depositing electrodes on the wafer surface. Integrated optic modulators have also been made by applying organic electro-optic material over glass waveguides, such that the application of the RF signal voltage modulates the waveguide cladding, thereby changing the effective propagation constant of the guided mode.⁵⁴

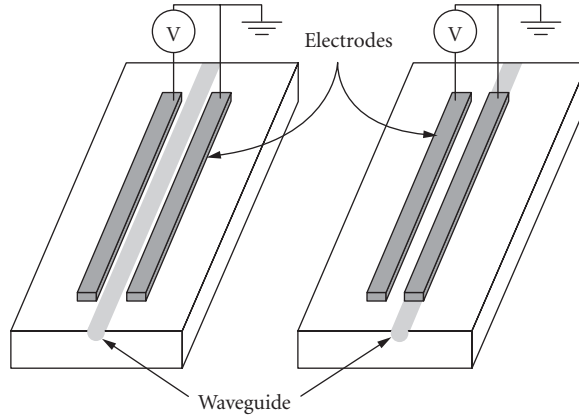


FIGURE 19 Schematic cross sections of two typical configurations of imbedded optical waveguides.

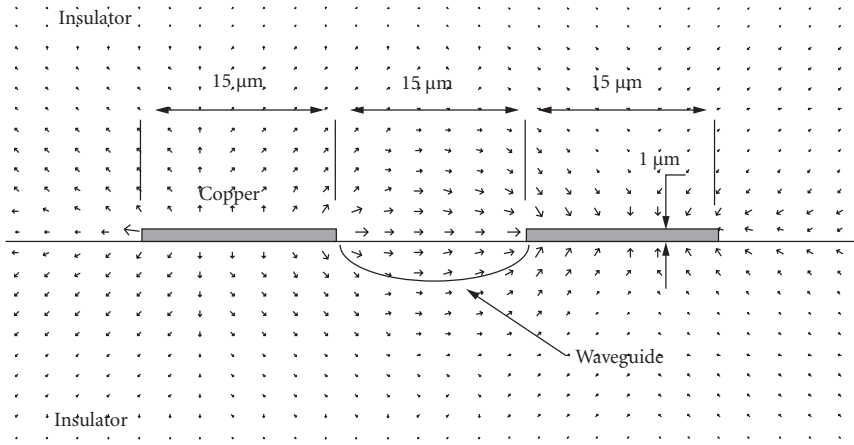


FIGURE 20 Electric field lines from electrodes across the embedded electro-optic waveguide. The electric field distribution can not be assumed to be uniform, as in bulk modulators.

The basic operation principles of waveguide modulators are the same as for bulk modulators; except that the spatial electrical and optical fields may not overlap completely. Figure 19 illustrates schematic cross sections of two typical configurations of embedded optical waveguides, and Fig. 20 shows that the field lines can not be assumed uniform as in the case of bulk modulators. Thus, the equation describing the electro-optically induced change in refractive index for a given configuration must include an overlap correction factor, Γ_{wg} .⁵⁵ For example, with a waveguide modulator, Eq. (34) would become

$$\Delta n'_x = \Delta\beta/k = -n_o^3 r_{13} \Gamma_{wg} E_z / 2 \quad (65)$$

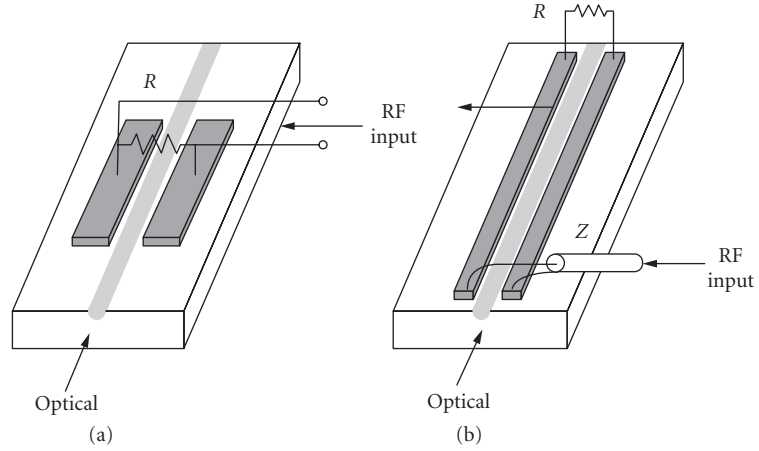


FIGURE 21 Electrode configurations on waveguides create either (a) lumped circuit element modulators or (b) traveling wave modulators.

where β is the propagation constant of the guided optical wave for a particular waveguide geometry. The overlap correction factor must be derived numerically. The total phase shift resulting from Δn over the propagation length L is again the sum of the natural phase shift and an electrically induced phase shift, but now includes a correction factor according to

$$\phi = \beta L + \Delta\beta L = \frac{2\pi}{\lambda}(n'_x + \Delta n'_x)L = \frac{2\pi}{\lambda}(n'_x - n_o^3 r_{13} \Gamma_{wg} E_z / 2)L \quad (66)$$

Waveguide modulators can be either lumped circuit element modulators or traveling wave modulators, based on their electrode configurations, which are illustrated by the phase modulators shown in Fig. 21. Similar to low frequency bulk modulators, lumped circuit element modulators have a modulation bandwidth that is limited by the parallel load resistance and electrode capacitance, which is proportional to the electrode length and gap. Decreasing the electrode length or increasing the electrode gap will decrease capacitance, and hence the RC time constant, but at the expense of requiring a higher drive voltage for a given voltage-L product. Traveling wave modulators, on the other hand, are bandwidth limited by the velocity mismatch between the propagating optical wave and the traveling RF or microwave signal.

Figure 21 illustrates phase modulators; however, amplitude (intensity) modulators are prevalent in communications applications, where they have found widespread use in fiber optical telecommunications and optical transmission of cable TV over fiber. Figure 22 shows a variety of electrode and waveguide configurations that implement intensity modulators. The Mach-Zehnder interferometer in Fig. 22a consists of traveling wave electrodes that modulate one or both arms and incur a phase difference between them. The phase-modulated optical signal recombines at the output Y-junction either constructively or destructively, so that intensity modulation results. There is an inherent 3 dB loss at the recombining Y-junction, due to an antisymmetric radiation mode that can not remain guided, which is a characteristic that does not exist for the bulk interferometric modulator. The directional coupler in Fig. 22b is designed to bring the two arms close to each other such that coupling occurs between the respective guided modes. Application of modulation controls the amount of cross coupling between the two guides and intensity modulation again results. In Fig. 22c, a device is shown that consists of a silicon substrate, a silica undercladding, and a doped-silica waveguide covered with an electro-optic polymer that has a refractive index slightly less than that of the core. When a voltage is applied to the electrodes, an electric field is produced through the electro-optic polymer and the passive optical waveguide, thus affecting the phase of the propagating light.

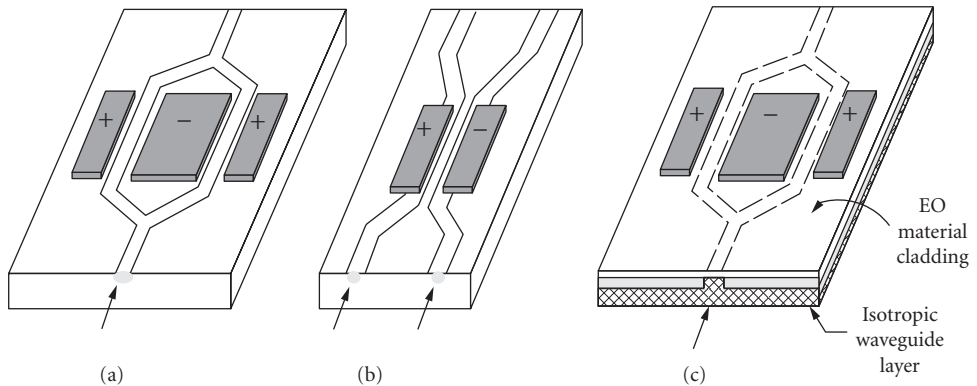


FIGURE 22 (a) Mach-Zehnder interferometer, (b) directional coupler, and (c) electro-optic cladding modulator.

Materials and Design Considerations

Two fundamental considerations to design an electro-optic modulator are signal voltage (and V_π) and frequency response. The goal is to achieve a desired depth of modulation with the smallest amount of drive power and at the same time obtain the largest possible bandwidth. In general, when designing bulk modulators, the field of the modulation signal and optical wave are assumed to have 100 percent overlap, whereas for waveguide modulators the overlap is not complete or uniform and a correction factor is needed. Bulk modulators are also capable of handling higher optical powers than waveguides, but at the expense of higher drive voltages and generally less modulation bandwidth. These are the trades that must be considered in choosing an electro-optic modulator.

This section discusses the various materials used to make electro-optic modulators and how the properties of electro-optic materials can affect the device performance. The various criteria and parameters that describe device performance are summarized.

Choice of Material An electro-optic material must be selected that would perform ideally at a given wavelength or a broad spectral range, and in a given environment. The material should have a large electro-optic coefficient, good homogeneity, low absorption, low dispersion, high resistivity; and thermal, mechanical, and photochemical stability. The fabricated modulator should have good ohmic contacts, and the fabrication techniques should be economically and technically feasible.

From a practical point of view, there are several limits on the performance of electro-optic modulators that must be considered. For example, some materials with natural birefringence, such as KDP and ferroelectric LiTaO_3 and LiNbO_3 , are sensitive to temperature variations and would induce a temperature-dependent phase shift in the light, called thermal drift. Some of these materials also have acoustic resonances due to the piezo-electric effect that cause undesirable peaks in the frequency response.^{24,45,52,56} The optical power handling capability, which is wavelength dependent, should be known so as not to induce photorefractive damage to the material.

Electro-optic materials include inorganic crystals, organic crystals, semiconductors, macroscopically oriented amorphous organics (dye-doped polymers), and liquids. Each of these materials are suited to particular applications based on their material properties and are briefly summarized as follows:

1. *Inorganic crystals* have high optical power handling ability and superior optical quality. Lithium niobate, LiNbO_3 , continues to be the “workhorse” of electro-optic modulators, both as a bulk modulator and especially as a waveguide. For laser applications, such as Pockels cells, crystals with high optical power handling capabilities are used. KDP and BBO are typical Pockels cell crystals for visible to near infrared applications, while CdTe is used for longer wavelengths.
2. *Organic crystals* have been explored for their large electro-optic coefficients that exceed those of inorganic crystals and their fast response time. The organic crystal 4-dimethylamino-4-stilbazolium tosylate

(DAST) presently has the highest known electro-optic coefficient and is commercially grown. While the electro-optic coefficients are large, organic crystals cannot be grown into large single crystals like inorganics, and they generally have lower power handling capabilities.

3. *Semiconductors* have been investigated as electro-optic materials because of the potential for these materials to be integrated with electronic integrated circuits and because of their transparency in the infrared; however, the effect is typically weak. Materials that have been studied include GaAs, ZnSe, ZnTe, CdS, CdTe:InP, CuCl, InSb, InGaAsP, and CdHgTe.⁵⁷
4. *Ordered dye-doped polymers* are a class of amorphous organic materials whose individual molecules display extremely large electro-optic properties. The electronic properties are based on field-induced delocalized π -electrons and additional donor and acceptor groups on opposite sides of the molecules. A mixture of these molecules does not exhibit a macroscopic electro-optic effect until the molecules have generally been oriented in the same direction. This is accomplished by several means: an electric field is applied to the assembly of molecules while above the glass transition temperature, then the material and orientation is solidified; or the orientation may be obtained by self-assembly techniques such as Langmuir-Blodgett. While the electro-optic effect can be large in these materials, over time the ordering of assembly of molecules can degrade, causing the effect to weaken. Nevertheless, research in this area remains active.
5. *Liquids* can exhibit the Pockels effect under applied electric field if the molecules are dipolar with donor acceptor groups, similar to dye-doped polymers. Some liquids also have very large Kerr coefficients and are the basis of Kerr cells.

Performance Criteria

The following parameters are indicators of modulator performance. Basically, the tradeoffs occur between the aspect ratio L/b , the drive voltage, and the electrode configuration (lumped or traveling wave) to achieve a desired depth of modulation with the smallest amount of power and at the same time obtain the largest possible bandwidth.

Modulation Bandwidth The bandwidth is the difference between the two closest frequencies at which the modulation index, δ or Γ , falls to 50 percent of the maximum value.^{55,58} Therefore, the 3 dB optical bandwidth is determined by setting the modulation reduction factor ρ or ρ_{tw} to 0.5.^{44,55,58} Modulation speed depends on the electrode type, lumped or traveling wave, and the modulator capacitance per length, which depends on the RF dielectric constant ϵ and the geometry of the electrodes. For a lumped modulator the bandwidth is limited by the optical (c/Ln) or electrical ($c/L\sqrt{\epsilon}$) transit time, whichever is smaller, or the time constant of the lumped-circuit parameters ($1/RC$), where R is the resistance in the circuit and C is the capacitance of the modulator. For a traveling wave modulator the bandwidth ν_{tw} is limited by the velocity mismatch between the optical and modulation waves, $\nu_{tw} = c/Ln(1 - \sqrt{3/\eta})$.¹⁶

Power per Unit Bandwidth (Specific Energy) and Resonant Circuits To apply the modulating signal efficiently to a *lumped* electro-optic modulator, a resonant *RLC* circuit can be created by adding an inductance and resistance in parallel to the modulator, which is modeled by a capacitor C .^{16,41} The terminating resistance R ensures that more of the voltage drop occurs over the modulator rather than the internal resistance of the source R_s .

The impedance of the resonant circuit is high over a bandwidth (determined by the RC time constant) of $\Delta\nu = \Delta\omega_m/2\pi \approx 1/2\pi RC$ centered at the resonant frequency of $\omega_o = 1/\sqrt{LC}$, where the impedance of the parallel *RLC* circuit is exactly equal to R . A peak voltage V_m must be applied to achieve a desired peak retardation $\Gamma_m = \pi V_m/V_\pi$. Therefore, $V_m = \Gamma_m(V_\pi/\pi)$. The power required to achieve Γ_m is $P = V_m^2/2R$, where $1/2R = \pi C\Delta\nu$, giving

$$P/\Delta\nu = \frac{1}{\pi} \Gamma_m^2 V_\pi^2 C \quad (67)$$

The power per unit bandwidth (mW/MHz) depends on the modulator capacitance C and the peak modulation voltage V_m to give the desired depth of modulation Γ_m . The required drive power increases with modulation frequency.^{41,55} Since the capacitance is a function of the modulator active area dimensions, the required power is also a function of the modulator dimensions.⁵⁵ Furthermore, since V_π is directly proportional to the wavelength λ , a higher input power (voltage) is required at longer wavelengths for a given peak retardation.

Extinction Ratio The extinction ratio is the maximum depth of intensity modulation for a time-varying voltage at the output when the optical bias is adjusted properly.^{56,58} If for no voltage the output intensity is I_o and for maximum applied voltage the output intensity is I_m , then the extinction ratio is defined as⁵⁸

$$\eta_m = \frac{|I_m - I_o|}{I_o} \quad I_m \leq I_o$$

$$\eta_m = \frac{|I_m - I_o|}{I_m} \quad I_m \geq I_o \quad (68)$$

Material effects, such as crystal imperfections, temperature sensitivities, and birefringence, can degrade the extinction ratio,^{16,38,45} thereby affecting the signal-to-noise ratio at the detector.⁵⁸ Another definition is in terms of the transmission T , $\eta_m = T_{\max}/T_{\min}$.^{42,59} In general, $T_{\max} < 1$ due to absorption, reflection, and scattering losses, and $T_{\min} > 0$ due to beam-divergence angle, residual crystal birefringence, crystal inhomogeneity, electric field uniformity, background scattered light, and polarizer-analyzer alignment.⁴⁰ Extinction ratio also can be applied to phase modulation, since phase changes can be related to equivalent changes in intensity.⁵⁸

Maximum Frequency Deviation A similar figure of merit as exists for intensity modulators likewise exists for frequency modulators. The maximum deviation of a frequency modulator is defined as⁵⁵

$$D_{\max} = \frac{|\omega_d - \omega|}{\omega} \quad (69)$$

where ω_d is the frequency shift when the maximum voltage is applied.

Percent Modulation Percent modulation is an intensity modulation parameter. Basically, it is the transmission $T = I_o/I_i$ times 100 percent at a specific wavelength. For a device with total retardation of $\Gamma = \pi/2$ radians at no voltage, the transmission $T = 0.5$. Then 70 percent modulation is achieved with an analog signal if a voltage is applied such that $\Gamma = 2$ radians; that is, $\sin^2(1) = 0.70$.⁵⁸ Hundred percent intensity modulation is the output intensity varying between the input intensity and zero intensity.⁴¹ A peak voltage of $V = V_\pi/2$ is required to achieve this level of performance for a linear modulator.

Another definition of percent modulation is the ratio of the peak output intensity at the modulation frequency to the maximum output intensity at dc voltage.⁵⁶ Reference 45 defines percent modulation as the peak modulation voltage output per dc voltage output, assuming no temperature variations.

Degree of Modulation For a sinusoidal reference of an optical electric field, $E(t) = E_i \sin \omega t$, an amplitude-modulated signal is typically expressed as

$$E(t) = E_i(1 + m \sin \omega_m t) \sin \omega t \quad (70)$$

where m is the degree of modulation at a specific wavelength.¹⁷ For $m \ll 1$ the intensity is

$$I = \frac{1}{2} E_i^2 (1 + 2m \sin \omega_m t) \quad (71)$$

The amplitude modulation index is then $\Gamma_m = 2m$, a function of the degree of modulation. The degree of modulation is often referred to as percent modulation for intensity modulation (70 to 100 percent nominal) and as modulation index for phase modulation (1 radian nominal).³⁸

Modulation Efficiency Modulation efficiency is defined as the percentage of total power which conveys information.^{60,61} The total power of an amplitude-modulated optical carrier is a function of the modulation frequency and is proportional to $1 + m^2 \langle \sin^2 \omega_m t \rangle = 1 + (1/2)m^2$. Therefore, the modulation efficiency is⁶⁰

$$\xi = \frac{\frac{1}{2}m^2}{1 + \frac{1}{2}m^2} \quad (72)$$

The maximum efficiency is achieved when m is maximum; that is, $m = (1/2)\Gamma_m = \pi V/2V_\pi$. In terms of the input power P the modulation efficiency is⁶¹

$$\xi = \frac{\Gamma^2}{P} \quad (73)$$

where $\Gamma = \pi V/V_\pi$ for maximum efficiency.

Optical Insertion Loss For an external modulator, in particular, the optical insertion loss must be minimized when coupling light into and out of the device. For an input light intensity I_{in} the insertion loss IL is⁵⁸

$$\begin{aligned} IL &= 1 - \frac{I_m}{I_{in}} & I_m &\geq I_o \\ IL &= 1 - \frac{I_o}{I_{in}} & I_m &\leq I_o \end{aligned} \quad (74)$$

where I_o and I_m are the output intensities at no voltage and maximum voltage, respectively. If a beam has a large cross-sectional area, the modulator must have a large acceptance area to couple all or most of the light into the device. A longitudinal modulator typically has a large acceptance area. Minimizing insertion loss in a transverse modulator is more of a challenge due to its narrow width, particularly for wide beams. However, a transverse modulator requires less voltage for a desired degree of modulation.

7.6 APPLICATIONS

Section 7.5 have discussed various device geometries and how the geometries translate into phase, amplitude, frequency, or displacement modulators. In this section, several applications of these modulators are discussed and examples are given. Waveguide modulators are discussed in another chapter of this *Handbook* and will not be further discussed here. Instead, high-speed electro-optic sampling, sensors applications, and laser mode-locking are discussed as classical application of both Pockels and Kerr electro-optic modulation.

Electro-Optic Sampling

Electro-optic sampling is a technique that can be used to sample ultrashort temporal phenomena that cannot be otherwise measured by conventional means. The sensing of terahertz radiation to

perform spectroscopy or to measure other short electrical phenomena in electrical circuits are two examples. The technique has been enabled by the advent of pulsed laser systems, creating picosecond to femtosecond optical pulses of duration shorter than the electrical pulse to be measured. It consists of a “probe” pulsed optical light beam that is modulated by the temporary short electric field, both of which are coincident within an electro-optic crystal or material. Because the probe duration is less than the modulating electrical pulse (which may be a terahertz pulse or other electrical transient), the probe polarization is modified through the electro-optic effect, the degree to which is based on the strength of the electric field at that instant of time. A polarizer is placed after the electro-optic crystal, so that the change in intensity of one particular polarization can be measured. The temporal position of the probe relative to the electrical signal is slowly varied. However, at a particular instant of time, many optical pulses are perturbed, thus permitting the use of slower speed detectors to measure the variation of the optical power.

Figure 23a illustrates a typical terahertz time-domain spectroscopy (TDS) system that utilizes a pump-probe method of electro-sampling to measure terahertz pulses. A ZnTe crystal or an organic DAST crystal is used as the electric field sampling sensor. However, since these both are noncentrosymmetric crystals, they can also be used to generate the terahertz pulse by the second-order process of optical rectification. Figure 23b shows the reconstructed terahertz pulse.

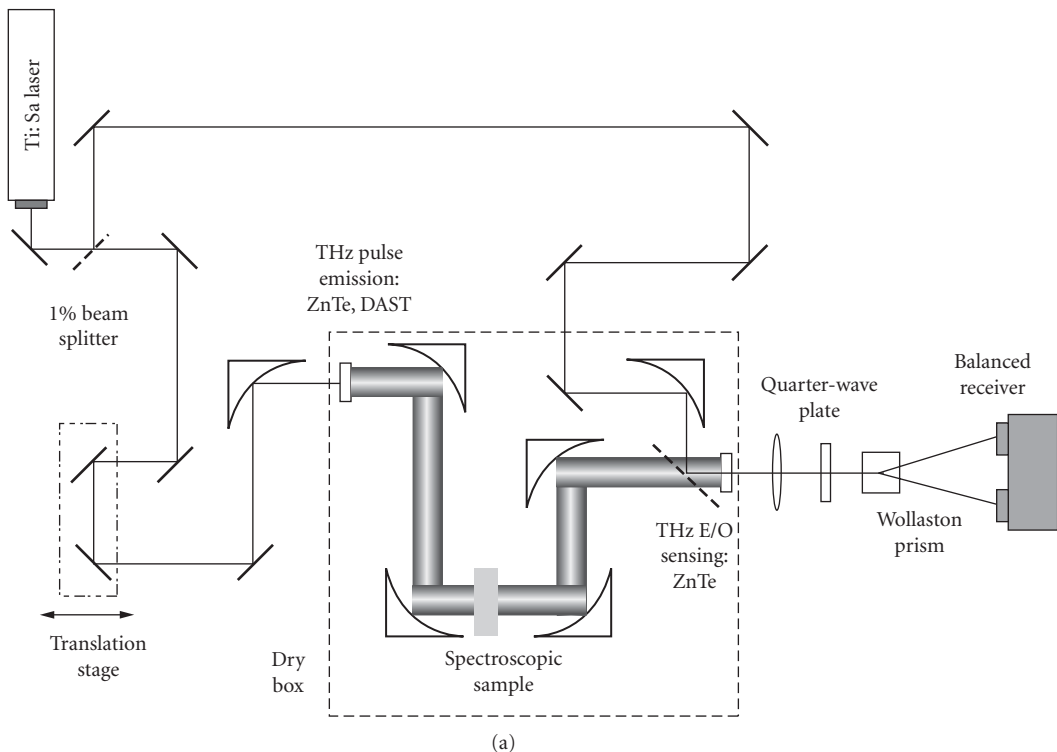


FIGURE 23 (a) A terahertz TDS system uses a Ti:Sapphire laser to emit ultrashort optical pulses, which are split into a pump beam to generate terahertz pulses, and a probe beam used to sample the terahertz pulses. An electro-optic crystal is used as a sensor. (b) The terahertz pulse reconstructed by sampling the electric field strength at different points in time.

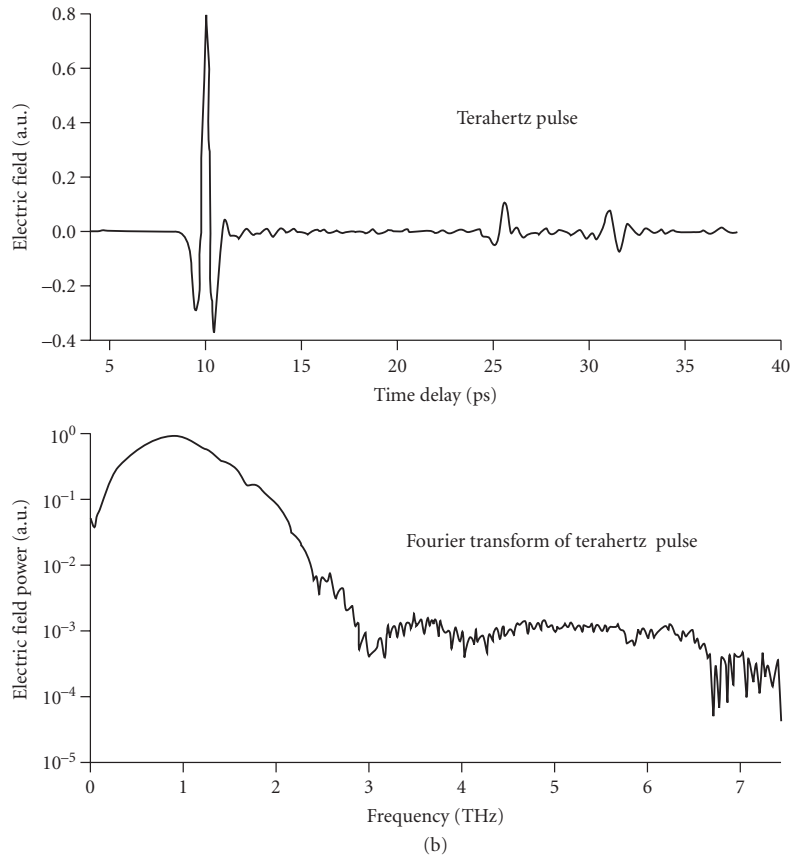


FIGURE 23 (Continued)

Sensors

A practical commercial sensor based on the electro-optic effect is a high-voltage sensor. This sensor uses the Pockels effect to sense high voltages without contact with a conductor or a voltage source. It finds application in improving reliability of electric utility systems. In contrast to other modulators discussed in this chapter, wherein a low-drive voltage was desired, high-voltage sensors accurately measure voltages as high as 765 kV. The electro-optic material for the sensor is being used as a bulk modulator, and the excitation voltage is of low frequency.

In contrast to the low-frequency high-voltage sensor just discussed, an integrated optic Mach Zehnder interferometer (MZI) can be used as a high-power microwave sensor or RF sensor. This device has been shown to operate over a wide frequency range of 200 MHz to 12 GHz.⁶³

Laser Mode-Locking

Mode-locking is a resonant phenomenon that results in ultrashort laser pulses. Mode-locking occurs by a relatively weak modulation of the radiation passing through the resonator, which results in initiating a pulse that becomes shorter with each pass through the resonator. The pulse compression technique uses the Kerr nonlinearity of an optical medium. The pulses propagating

through the medium experience nonlinear phase shifts that lead to spectral broadening, resulting in a spread of frequencies, or chirp. The different frequency components are superimposed by propagation in a dispersive medium, or by reflection from a dispersive element. The effects of a fast saturable absorber to produce mode-locking can be simulated by Kerr focusing. The Kerr effect within the resonator causes the high-intensity part of the optical beam to be focused, whereas the low-intensity components remain unfocused. If an aperture is placed in the beam path, the low-intensity parts are attenuated, and the pulse is shortened. This effect is now called Kerr-lens mode-locking (KLM).²⁸ The KLM technique is commonly used for mode-locking of Ti:Sapphire lasers.

7.7 APPENDIX: EULER ANGLES

Euler angles represent a set of three independent parameters which specify the orientation of a rigid body, in this case the index ellipsoid. Orthogonal transformations are used to convert from one set of axes (x, y, z) to another set (x', y', z'). The two sets of axes are related to each other by nine direction cosines. An orthogonal transformation matrix consisting of direction cosines and having a determinant of +1 corresponds to defining the orientation of a rigid body.

The transformation matrix is developed by a specific sequence of three plane rotations (not in principal planes) in a defined order. There are 12 possible conventions for defining a set of Euler angles in a right-handed coordinate system.³⁶ One convention that has been proposed as a standard is the y convention.⁶⁴ The transformation evolves by an initial counterclockwise rotation ζ about the z axis, followed by a counterclockwise rotation η about the intermediate y' axis, and finally a counterclockwise rotation ω about z'' . The resulting transformation matrix is

$$\bar{a} = \begin{pmatrix} \cos \omega & \sin \omega & 0 \\ -\sin \omega & \cos \omega & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \eta & 0 & -\sin \eta \\ 0 & 1 & 0 \\ \sin \eta & 0 & \cos \eta \end{pmatrix} \begin{pmatrix} \cos \zeta & \sin \zeta & 0 \\ -\sin \zeta & \cos \zeta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (\text{A.1})$$

To find the Euler angles (ζ, η, ω), Eq. (A.1) is rearranged as³⁷

$$\begin{pmatrix} \cos \omega & -\sin \omega & 0 \\ \sin \omega & \cos \omega & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} \cos \eta & 0 & -\sin \eta \\ 0 & 1 & 0 \\ \sin \eta & 0 & \cos \eta \end{pmatrix} \begin{pmatrix} \cos \zeta & \sin \zeta & 0 \\ -\sin \zeta & \cos \zeta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (\text{A.2})$$

where $\bar{a}$ is the cumulative transformation matrix of the normalized eigenvectors. Multiplying the matrices, the Euler angles are related to the elements of $\bar{a}$:

$$a_{11} \cos \omega - a_{21} \sin \omega = \cos \eta \cos \zeta \quad (\text{A.3a})$$

$$a_{12} \cos \omega - a_{22} \sin \omega = \cos \eta \sin \zeta \quad (\text{A.3b})$$

$$a_{13} \cos \omega - a_{23} \sin \omega = -\sin \eta \quad (\text{A.3c})$$

$$a_{11} \sin \omega + a_{21} \cos \omega = -\sin \zeta \quad (\text{A.3d})$$

$$a_{12} \sin \omega + a_{22} \cos \omega = \cos \zeta \quad (\text{A.3e})$$

$$a_{13} \sin \omega + a_{23} \cos \omega = 0 \quad (\text{A.3f})$$

$$a_{31} = \sin \eta \cos \zeta \quad (\text{A.3g})$$

$$a_{32} = \sin \eta \sin \zeta \quad (\text{A.3h})$$

$$a_{33} = \cos \eta \quad (\text{A.3i})$$

From Eq. (A.3f), the angle ω is $\omega = \tan^{-1}(-a_{23}/a_{13})$. From Eq. (A.3c), the angle η is $\eta = \sin^{-1}(a_{23} \sin \omega - a_{13} \cos \omega)$. From Eq. (A.3d), the angle ζ is $\zeta = \sin^{-1}(-a_{11} \sin \omega - a_{21} \cos \omega)$.³⁷

7.8 REFERENCES

1. W. J. Tomlinson and C. A. Brackett, "Telecommunications Applications of Integrated Optics and Optoelectronics," *Proc. IEEE* **75**(11):1512–1523 (1987).
2. E. Vogues and A. Neyer, "Integrated-Optic Devices on LiNbO₃ for Optical Communication," *IEEE/OSA J. Lightwave Technol.* **LT-5**:1229–1238 (1987).
3. L. Thylen, "Integrated Optics in LiNbO₃: Recent Developments in Devices in Telecommunications," *IEEE/OSA J. Lightwave Technol.* **6**(6):847–861 (1988).
4. R. C. Alferness, "Guided-Wave Devices for Optical Communication," *IEEE J. Quantum Electron.* **QE-17**(6): 946–959 (1981).
5. H. F. Taylor, "Application of Guided-Wave Optics in Signal Processing and Sensing," *Proc. IEEE* **75**(11):1524–1535 (1987).
6. See, for example, "Special Issue on Optical Computing," *Proc. IEEE* **72** (1984).
7. See, for example, "Special Feature on Integrated Optics: Evolution and Prospects," *Opt. News* **14** (1988).
8. T. K. Gaylord and E. I. Verriest, "Matrix Triangularization Using Arrays of Integrated Optical Givens Rotation Devices," *Computer* **20**:59–66 (1987).
9. C. M. Verber, R. P. Kenan, and J. R. Busch, "Design and Performance of an Integrated Optical Digital Correlator," *IEEE/OSA J. Lightwave Technol.* **LT-1**:256–261 (1983).
10. C. L. Chang and C. S. Tsai, "Electro-Optic Analog-to-Digital Conversion Using Channel Waveguide Fabry-Perot Modulator Array," *Appl. Phys. Lett.* **43**:22 (1983).
11. C. M. Verber "Integrated-Optical Approaches to Numerical Optical Processing," *Proc. IEEE* **72**:942–953 (1984).
12. X. Zheng, S. Wu, and R. Sobolewski, "Electro-Optic Sampling System with a Single-Crystal 4-N, N-Dimethylamino-4'-N'-Methyl-4-Stibazolium Tosylate Sensor," *Appl. Phys. Lett.* **82**(15):2383–2385 (2003).
13. J. Ruan, H. Edwards, C.-Y. Tan, R. Thurman-Keup, and V. Scarpine, "Design of an Electro-Optic Sampling Experiment at the AWA Facility," *Proc. IEEE PAC'07*, Albuquerque, NM, 3901–3903 (2007).
14. T. A. Maldonado and T. K. Gaylord, "Light Propagation Characteristics for Arbitrary Wavevector Directions in Biaxial Crystals by a Coordinate-Free Approach," *Appl. Opt.* **30**:2465–2480 (1991).
15. M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon Press, Oxford, UK, 1980.
16. A. Yariv and P. Yeh, *Optical Waves in Crystals*, Wiley, New York, 1984.
17. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon, London, 1960.
18. A. I. Borisenko and I. E. Tarapov, in *Vector and Tensor Analysis with Applications*, R. A. Silverman (ed.), Prentice-Hall, Englewood Cliffs, NJ, 1968.
19. I. P. Kaminow, *An Introduction to Electrooptic Devices*, Academic Press, New York, 1974.
20. T. C. Phemister, "Fletcher's Indicatrix and the Electromagnetic Theory of Light," *Am. Mineralogist* **39**:173–192 (1954).
21. T. A. Maldonado and T. K. Gaylord, "Electro-Optic Effect Calculations: Simplified Procedure for Arbitrary Cases," *Appl. Opt.* **27**:5051–5066 (1988).
22. J. F. Nye, *Physical Properties of Crystals, Their Representation by Tensors and Matrices*, Oxford Univ. Press, New York, 1985.
23. I. P. Kaminow, in *Handbook of Laser Science and Technology*, vol. IV, part 2, M. J. Weber (ed.), CRC Press, Boca Raton, FL, 1986, pp. 253–278.
24. I. P. Kaminow and E. H. Turner, "Electro-Optic Light Modulators," *Proc. IEEE* **54**(10):1374–1390 (1966).
25. J. M. Bennett and H. E. Bennett, "Polarization," in *Handbook of Optics*, W. G. Driscoll and W. Vaughan (eds.), McGraw-Hill, New York, 1978, chap. 10.
26. A. Yariv, *Optical Electronics*, Holt, Rinehart, and Winston, New York, 1976.
27. D. C. Cronmeyer and L. R. Spanberger, "Infrared Transmittance and Kerr Effect of Nitrobenzene," *J. Opt. Soc. Am.* **51**:1061–1066 (1961).
28. H. A. Haus, "Mode-Locking of Lasers," *IEEE J. Select. Topics Quantum Electronics* **66**(6):1173–1185 (2000).
29. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford Univ. Press, London, 1965.

30. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterlirig, *Numerical Recipes*, Cambridge Univ. Press, New York, 1986.
31. J. H. Wilkinson and C. Reinsch, *Handbook for Automatic Computation*, Springer-Verlag, New York, 1971.
32. G. H. Golub and C. F. Van Loan, *Matrix Computations*, John Hopkins Univ. Press, Baltimore, 1983.
33. D. E. Sands, *Vectors and Tensors in Crystallography*, Addison-Wesley, Reading, MA, 1982.
34. D. R. Hartree, *Numerical Analysis*, Clarendon Press, Oxford, 1952.
35. B. N. Parlett, *The Symmetric Eigen Value Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
36. H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, MA, 1981.
37. R. P. Paul, *Robot Manipulators*, The MIT Press, Cambridge, MA, 1981.
38. I. P. Kaminow and T. Li, "Modulation Techniques," in *Optical Fiber Telecommunications*, S. E. Miller and A. G. Chynoweth (eds.), Academic Press, New York, 1979, chap. 17.
39. D. F. Nelson, "The Modulation of Laser Light," *Scientific Am.* **218**(6):17–23 (1968).
40. E. Hartfield and B. J. Thompson, "Optical Modulators," in *Handbook of Optics*, W. G. Driscoll and W. Vaughan (eds.), McGraw-Hill, New York, 1978, chap. 17.
41. A. Ghatak and K. Thyagarajan, *Optical Electronics*, Cambridge Univ. Press, New York, 1989.
42. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, Wiley, New York, 1991, chap. 18.
43. E. Hecht, *Optics*, 2nd ed., Addison-Wesley, Reading, MA, 1990.
44. R. C. Alferness, in *Guided-Wave Optoelectronics*, T. Tamir (ed.), Springer-Verlag, New York, 1990, chap. 4.
45. W. H. Steier, "A Push-Pull Optical Amplitude Modulator," *IEEE J. Quantum Electron.* **QE-3**(12):664–667 (1967).
46. C. F. Buhrer, D. Baird, and E. M. Conwell, "Optical Frequency Shifting by Electro-Optic Effect," *Appl. Phys. Lett.* **1**(2):46–49 (1962).
47. V. J. Fowler and J. Schlafer, "A Survey of Laser Beam Deflection Techniques," *Appl. Opt.* **5**:1675–1682 (1966).
48. F. S. Chen, et al., "Light Modulation and Beam Deflection with Potassium Tantalate Niobate Crystals," *J. Appl. Phys.* **37**:388–398 (1966).
49. V. J. Fowler, C. F. Buhrer, and L. R. Bloom, "Electro-Optic Light Beam Deflector," *Proc. IEEE* (correspondence) **52**:193–194 (1964).
50. T. J. Nelson, "Digital Light Deflection," *B.S.T.I.* 821–845 (1964).
51. W. Kulcke, et al., "A Fast, Digital-Indexed Light Deflector," *IBM J.* **8**:64–67 (1964).
52. M. Gottlieb, C. L. M. Ireland, and J. M. Ley, *Electro-Optic and Acousto-Optic Scanning and Deflection*, Marcel Dekker, New York, 1983.
53. D. M. Pozar, *Microwave Engineering*, Addison-Wesley, Reading, MA, 1990.
54. R. W. Ridgway, V. McGinniss, D. W. Nippa, and S. M. Risser, "Electrooptic Control of Functionally-Clad Silica Waveguides," in *Integrated Photonics Research*, A. Sawchuk (ed.), vol. 91 of OSA Trends in Optics and Photonics (Optical Society of America, 2003), paper ITuH5.
55. R. G. Hunsperger, *Integrated Optics: Theory and Technology*, 2nd ed., Springer-Verlag, Berlin, 1983.
56. R. T. Denton, et al., "Lithium Tantalate Light Modulators," *Appl. Phys.* **38**(4):1611–1617 (1967).
57. T. G. Brown, K. Creath, H. Kogelnik, M. A. Kriss, J. Schmit, M. J. Weber (eds.), *The Optics Encyclopedia: Basic Foundations and Practical Applications*, vol. 1, Electro-Optic Devices, Wiley, 2004.
58. J. M. Hammer, in *Integrated Optics*, T. Tamir (ed.), Springer-Verlag, Berlin, 1979, chap. 4, pp. 140–200.
59. R. Simon, *Optical Control of Microwave Devices*, Artech House, Boston, MA, 1990.
60. R. E. Ziemer and W. H. Tranter, *Principles of Communications*, Houghton Mifflin Co., Boston, MA, 1976.
61. I. P. Kaminow and J. Liu, "Propagation Characteristics of Partially Loaded Two-Conductor Transmission Line for Broadband Light Modulators," *Proc. IEEE* **51**(1):132–136 (1963).
62. R. Forber, W. C. Wang, and De-Yu Zang, "Dielectric EM Field Probes for HPM Test and Evaluation," 2006 Annual ITEA Technology Review, August 7–10, Cambridge, MA 1 Non-Intrusive Instrumentation Technologies, 2006.
63. W. L. Bond "The Mathematics of the Physical Properties of Crystals," *Bell Sys. Tech. J.* **23**:1–72 (1943).

This page intentionally left blank

LIQUID CRYSTALS

Sebastian Gauza and Shin-Tson Wu

*CREOL, The College of Optics and Photonics
University of Central Florida
Orlando, Florida*

ABSTRACT

This chapter introduces the basic properties of liquid crystal materials, electro-optic properties of liquid crystal cells, and operation principles of liquid crystal display (LCD) devices, including the large screen thin-film-transistor LCD TVs and transfective LCDs for mobile displays.

8.1 GLOSSARY

C_{ii}	reduced elastic constants
d	liquid crystal thickness
d_R	reflective region cell gap
d_T	transmissive region cell gap
E	activation energy
F	Onsager reaction field
f_i	oscillator strength
f_c	crossover frequency
G	isotropic intermolecular attraction energy
H_f	heat fusion enthalpy
h	cavity field factor
I	moment of inertia
J	mean-field coupling constant
K_{ii}	elastic constants
k	Boltzmann constant
L	length of the molecule
l	aspect ratio

n	liquid crystal director
$n_{e,o}$	refractive indices
N	molecular packing density
P_4	order parameter of the fourth rank
P_s	spontaneous polarization
R	gas constant
p	molecular length-to-width ratio
S	order parameter of the second rank ($= P_2$)
T_c	clearing temperature
V_n	mole volume
V_b	bias voltage
V_{th}	threshold voltage
Z	number of electrons
$\alpha_{l,t}$	molecular polarizability
α_i	Leslie viscosity coefficients
β	angle between dipole and molecular axis
γ_1	rotational viscosity
δ	phase retardation
ϵ	dielectric constant
ϵ_0	vacuum permittivity
$\Delta\epsilon$	dielectric anisotropy
Δn	birefringence
μ	dipole moment
η_i	Miesowicz viscosity coefficient
θ	twist angle
Λ	friction coefficient
Φ	volume fraction of the molecules
λ	wavelength
ϕ	tilt angle
ϕ_s	pretilt angle
τ	response time
ω	frequency

8.2 INTRODUCTION TO LIQUID CRYSTALS

Liquid crystal (LC) was discovered in 1888 when an Austrian botanist named Friedrich Reinitzer observed that a material known as cholesteryl benzoate exhibiting two distinct melting points.¹ In his experiments, Reinitzer increased the temperature of a solid sample and observed that the crystal changes into a hazy liquid. As he continued to increase the temperature further, the hazy liquid turned into a clear state. Because of this early work, Reinitzer is often credited as the discoverer of a new phase of matter—the LC phase which is also known as liquid crystalline or mesophase. Most people are familiar with the fact that matter can exist in three different states: solid, liquid, and gas (vapor). However, this is a simplification, and under extreme conditions other forms of matter can exist, for example, plasma at very high temperatures or superfluid helium at very low temperatures. The difference between these states of matter is the degree of order in the material, which is directly related to the surrounding temperature and pressure. If the temperature is raised,

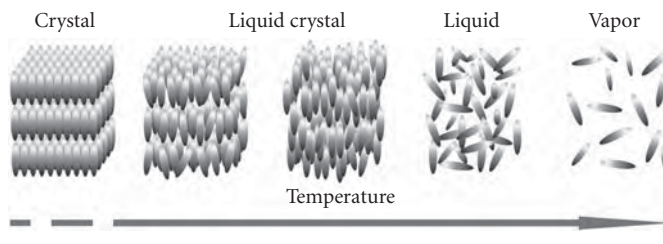


FIGURE 1 Schematic arrangement of the molecules in different phases.

more energy is transferred into the system, leading to increasingly stronger vibrations. Finally, at the transition temperatures between the solid and liquid states, the long range positional order is broken and the constituents may move in a random fashion (Fig. 1), constantly bumping into one another and abruptly changing the direction of motion. However, the thermal energy is not high enough to completely overcome the attractive forces between the constituents, so there is still some positional order at short range. Because of the remaining cohesion, the density of the liquid is constant even though the liquid takes the shape of its container, as opposed to a solid. The liquid and solid phases are called condensed phases. If we keep on raising the temperature until the next phase change, as shown in Fig. 1, the substance enters its gas (or vapor) state and its constituents are no longer bounded to each other.

The molecular order of an LC lies between those of the isotropic liquid and the crystal. The classification of LC materials is based on the degree of orientational and positional order. When the LC molecules are sandwiched between two glass substrates with their inner surfaces coated by an alignment agent, such as rubbed polyimide, the LC molecules tend to form a statistically preferred direction called the director, denoted by the vector $\mathbf{n}$. To specify the amount of orientational order in an LC phase, an order parameter is defined as

$$S = \langle P_2(\cos \theta) \rangle = \frac{2}{3} \langle \cos^2 \theta - 1 \rangle \quad (1)$$

where $\langle \rangle$ denotes a thermal averaging and θ is the angle between each molecule and the director $\mathbf{n}$. If the molecules are perfectly oriented (crystal state), that is, if $\theta = 0$ with the director, then $S = 1$. On the contrary, if the molecules are randomly oriented about $\mathbf{n}$, that is, an isotropic state, then $S = 0$ because there is no orientational order. So the higher the order parameter, the more ordered the liquid crystal phase is. In a typical LC system, order parameter decreases as the temperature increases. Temperature dependence of the order parameter S is shown in Fig. 2. Most common values

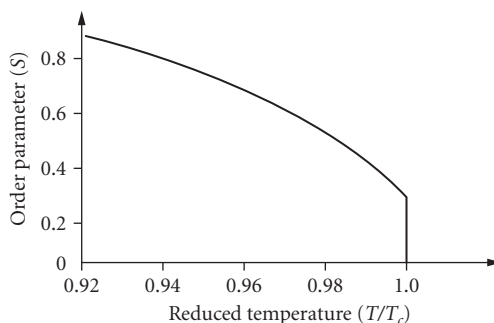


FIGURE 2 Temperature dependent order parameter of the liquid crystal phase.

of S are between 0.3 and 0.8. The order parameter is typically determined by the measured macroscopic properties like optical birefringence or diamagnetism. At certain temperatures, the LC material may gain a certain amount of positional order. When it happens, the center of mass of the LC molecules, although still forming a fluid, prefers to lie, on average, in layers as Fig. 1 depicts. This positional ordering may be described in terms of the density of the center of the mass of the molecules as

$$\rho(z) = \rho_0 \left[1 + \Psi \cos\left(\frac{2\pi z}{d}\right) \right] \quad (2)$$

where z is the coordinate parallel to the layer normal, the average density of the fluid is ρ_0 , d is the distance between layers, and Ψ is the order parameter. The modulus of Ψ , $|\Psi|$, represents the amplitude of the oscillation of the density. When $|\Psi| = 0$, there is no layering, but if $|\Psi| > 0$ then some amount of sinusoidal layering exists. Different LC phases are formed as a consequence, which has been described later in this chapter.

From a basic physics point of view, LC materials are of great interest and have contributed to the modern understanding of phase transitions and order phenomena in one, two, and three dimensions. To most of the people today, LC is almost synonymous to flat panel displays (LCDs) for computers, mobile phones, and other electronic equipments. However, there is a rapid development of the other types of applications. For example, there have been developments in telecommunications, pattern recognition, real time holography, light shutters, nonmechanical beam steering, and so on.² Liquid crystals constitute a unique form of soft matter and are becoming increasingly more important in pure material science in the development of polymer materials and biomaterials.

8.3 TYPES OF LIQUID CRYSTALS

Considering the geometrical structure of the mesogenic molecules, the liquid crystals can be classified into several types. The widespread LCDs are using rod-shaped nematic molecules, as shown in Fig. 3, where one molecular axis is much longer than the other two. These molecules are called calamitic liquid crystals. Molecules have to be rigid in one part (rigid core) and flexible in another (terminal flexible hydrocarbon chain). Different physical properties may be affected by exchanging one of the terminal chains with a group having different polarities. Liquid crystals formed by disc-shaped molecules with one molecular axis much shorter than the other two are called discotic liquid crystals, as shown in Fig. 3. In this case, the rigid part of the molecule is disc-shaped, typically having multiple aromatic rings. There are some possible intermediates between rod-shaped and disc-shaped molecules known as lath-like LC molecules. Transition to the mesophase (liquid crystalline phase) may be caused by either temperature or influence of solvents. The LCs obtained from the former method are called thermotropic. If the influence of solvents induces the transition, the LCs are called lyotropic. The LC materials capable of forming thermotropic as well as lyotropic mesophases are called amphotropic liquid crystals, as Fig. 4 shows. The lyotropic liquid crystal phases are formed by dissolving amphiphilic molecules in a suitable solvent. Amphiphilic molecules consist of

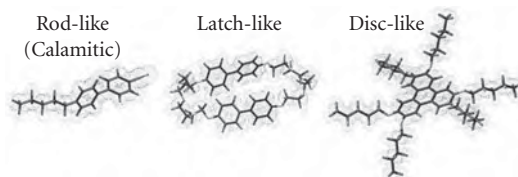


FIGURE 3 Molecular geometry of different types of LC.

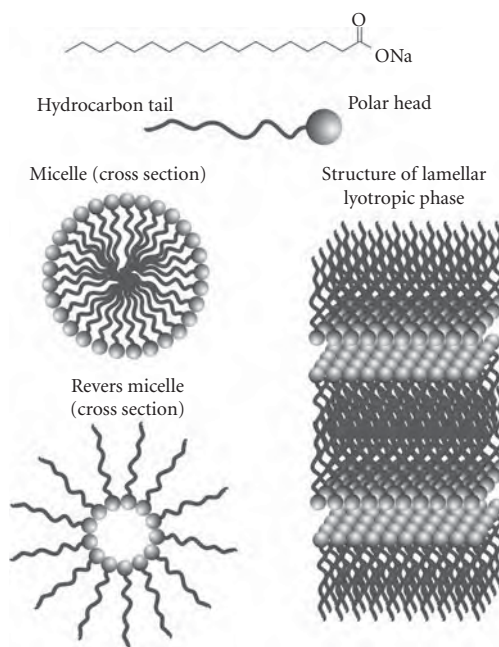
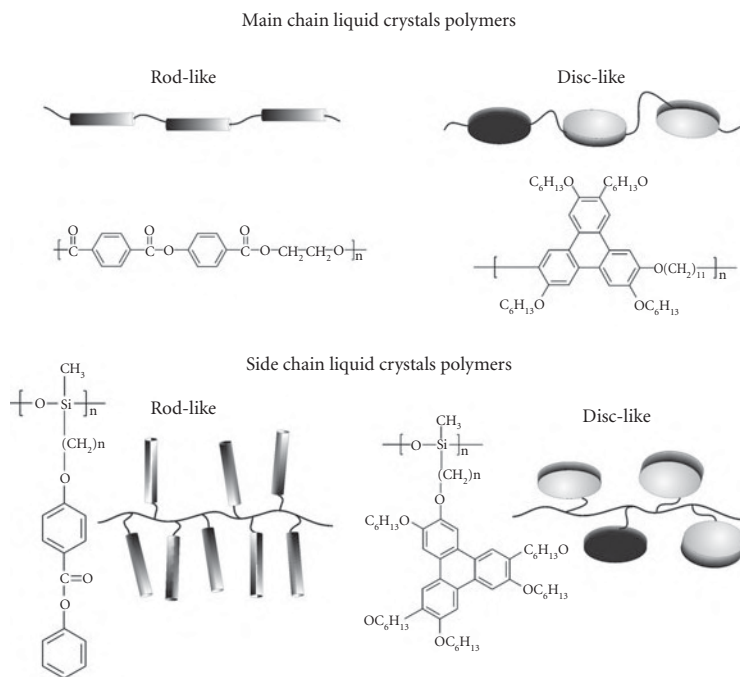


FIGURE 4 Examples of the lyotropic LCs.

hydrophobic group at one end and hydrophilic group at another. Such molecules form self-organized structures in solvents with different polarities. A very good example is soap. When dissolved in water (polar solvent), hydrophobic tails face each other and hydrophilic parts face solvent, forming a micelle. Another type, which is biologically important, is phospholipids which form bilayers. Such aggregations are the building units of biological membranes. Another important structure is liquid crystal polymers. As thermotropic mesogenes, these structures consist of mesogenic subunits (rod-like or disc-like) which are linked together with flexible linkage units, as shown in Fig. 5. If rigid mesogenic units are linked directly by flexible links, the main-chain polymer LC is formed. Mesogenic subunits can also be attached to the polymer chain as a side group which are known as side-chain polymer liquid crystals. Merged structure, called combined polymer liquid crystal has build-in main-chain as well as side-chain mesogenic units. Since the discovery of the first LC substances, there have been many research activities to determine what kind of structure forms desired liquid crystalline phase. Theory suggests that mesophases can be achieved when the molecules have elongated shape and some flexibility. Typically mesogenic molecules consist of several building blocks. If we concentrate on the most common calamitic LC block systems, it will contain elements shown in Fig. 6. As shown in Fig. 6, calamitic structures consist of a rigid rod formed by two (or more) ring systems (R1 and R2). These are connected together by a single bond or a linking group known as central bridge group. Molecular constituents at the ends of the rigid core (*para* position to the central group) are called terminal groups or chains. Usually at least one of the terminal groups must be a flexible carbon chain. Tables 1 to 3 show the most popular choices for the central bridge, ring systems, and terminal substituents.

The introduction of a linking group into a mesogenic molecule often determines linearity, increases overall molecular length, and changes polarizability anisotropy, thus influences mesomorphic and physical properties in general. The presence of a central linking bridge often widens the temperature range of the mesophase by reducing the melting point. However, sometimes, this may also affect the thermal and photochemical stability by increasing π -electron conjugation of the molecules. In some cases, linking groups may induce coloration.

**FIGURE 5** Types of LC polymeric materials.**FIGURE 6** Building blocks of calamitic LCs. A-R1-Z-R2-B.**TABLE 1** Typical Central Bridge Groups

Central Bridge Group	Central Bridge Group
$\text{—}[\text{CH}_2]\text{—}_n$ alkane	$\text{—CH}_2\text{—}$ ether
—C=C— alkene (olefin)	—N=N— azo
$\text{—C}\equiv\text{C—}$ alkyne (tolane, acetylene)	—N=N— ↓ O azoxy
—C(=O)O— ester	$\text{—C}\equiv\text{C—C}\equiv\text{C—}$ diacetylene

TABLE 2 Typical Ring Systems

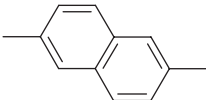
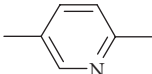
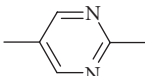
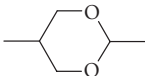
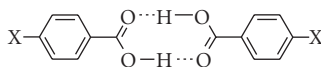
Ring System	Ring System
 $n = 1$ benzene $n = 2$ biphenyl $n = 3$ terphenyl	 bicyclooctane
 naphthalene	 pyridine
 phenanthrene	 pyrimidine
 cyclohexane	 dioxane

TABLE 3 Terminal Units

Terminal Unit A	Terminal Unit B
C_nH_{2n+1} Alkyl	-F, -Cl, -Br, -I Single elements
OC_nH_{2n+1} Alkoxy	-CN, -NO ₂ , -NCS Highly polar moieties

Most calamitic liquid crystals possess aromatic and alicyclic rings in the molecular structures. These include single benzene ring (commonly 1,4-phenyl), alicyclic ring (commonly cyclohexane), heterocyclic ring (e.g., 1,3-pyrimidine), and a wide variety of the combinations. Individual benzene rings do not yield good mesogens. At least two rings are needed to create a rod-like LC. The exception is the 4-alkyl-benzoic acid, which forms a dimer by hydrogen bond, as depicted in Fig. 7. Beside the heterocyclic rings, ring systems of the rigid core could be laterally substituted by elements other than hydrogen (most popular are F and Cl) or groups of elements such as CH₃, NH₂, CN, and so on. This

**FIGURE 7** Dimeric structure formed by 4-alkyl-benzoic acid.

will reduce the melting point of a mesogen, however, increase its viscosity. Additionally, optical and dielectric anisotropies will be affected. In general, the core of the mesogen determines the mesogenic properties by establishing the primary shape of the molecule and its rigidity. Many terminal units (A, B) have been employed in the creation of mesogenic molecules. The most successful route is to use either small polar substituents or a fairly long, straight, hydrocarbon chain. The role of these groups is to act as either a flexible extension to the rigid core or as a dipolar moiety to introduce anisotropy in physical properties. Molecules can obtain chirality from the flexible chain if it is branched and chiral. Further, the terminal moieties are believed to be responsible for stabilizing the molecular order essential for the mesophase generation. All physical properties strongly depend on the terminal units employed in the molecular system of mesogenic molecules. Some of the typical calamitic LC single compounds are listed in Table 4.³

8.4 LIQUID CRYSTALS PHASES

In this section, we focus on the calamitic liquid crystals which have been widely used in display industries. Other LC materials, such as discotic LCs, polymeric LCs, and lyotropic LCs which have great scientific values will not be discussed further. By doing so, we do not intend to minimize the importance of these materials. For example, Kevlar reveals the importance of lyotropic LC polymer material, while others are important in the biological sciences. The readers can find more precise information about these groups of LCs in Ref. 4.

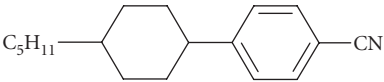
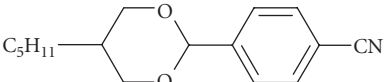
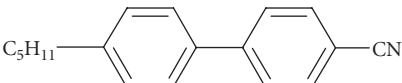
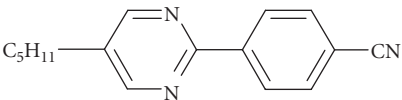
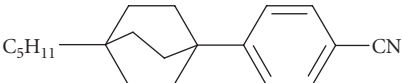
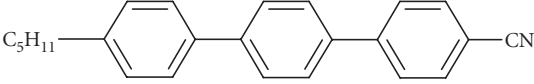
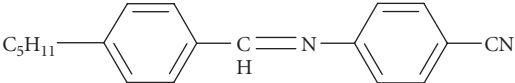
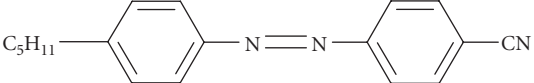
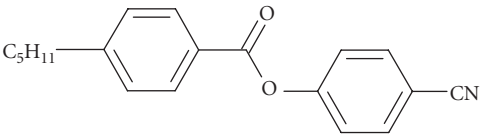
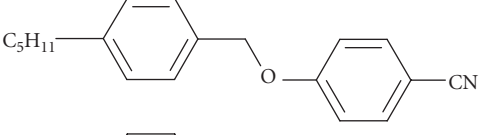
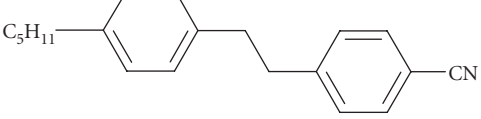
Calamitic Liquid Crystals

When classifying LC phases after G. Friedel,⁵ we first distinguish between two main types: one with nematic order and the other with smectic order (Fig. 8). In the nematic phase (of which tilted nematic is known as the cholesteric phase and is a special case) the molecules are free to move in all directions (e.g., there is no positional order of the center of mass). But on average, they keep their long axes locally parallel. In a smectic state, a number of structural variations exist and there is a positional order along one dimension (some smectic phases have positional order in more than one dimensions). A smectic LC is a layered structure with the molecules oriented parallel or tilted relative to the layer normal. Two smectic phases, called smectic *A* and smectic *C*, have acquired a special importance and are now relatively well understood. They are characterized by an absence of positional order within the layers. The molecules have some freedom to move within the layers, as in all smectic phases, but are much less free to move between the layers. These smectics can therefore be described as stacks of two-dimensional fluids, but behave as crystalline across the layers. The absence of in-layer order contributes to their high potential for future electro-optic applications. There are several smectic phases different from one another in areas such as the tilt angle of the director with regard to the layer normal, and the arrangement of molecules within each layer. The simplest is the smectic *A* phase (Fig. 8) characterized by a director parallel to the layer normal and a random positional order within the plane. Substances featuring the *A* phase often exhibit the smectic *C* phase at a lower temperature (Fig. 8). In this phase, the molecules have the same random order within the layer but tilt relative to the layer normal. The tilt angle normally increases with decreasing temperature. The other smectic phases are even more crystalline as they also feature some positional order within the layers. They may, for instance, exhibit hexagonal packing of the molecules.

Chiral Liquid Crystals

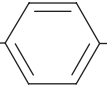
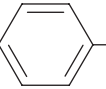
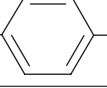
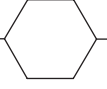
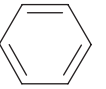
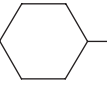
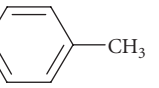
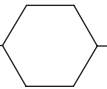
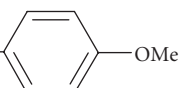
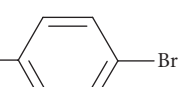
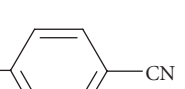
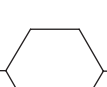
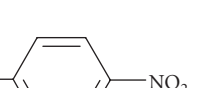
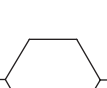
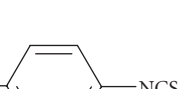
Molecules which are not identical to their mirror image are said to be chiral. A depiction of the simple concept of a chiral molecule is shown in Fig. 9. Another example is the human hand. Chiral molecules are able to form liquid crystals with structures related to those of nonchiral materials but with

TABLE 4 Physical Properties of Popular LC Compounds³

Compound	Phase Transition Temperature(°C)
Rigid core	
	cr 31 N 55 is
	cr 56 (N 52) is
	cr 31 N 55 is
	cr 71 (N 52) is
	cr 62 N 100 is
	cr 130 N 239 is
Central bridge link	
	cr 46 N 75 is
	cr 89 (N 87) is
	cr 65 (N 56) is
	cr 49 (N-20) is
	cr 62 (N -24) is

(Continued)

TABLE 4 Physical Properties of Popular LC Compounds³ (*Continued*)

Compound	Phase Transition Temperature (°C)
C_5H_{11} —  —C≡C—  —CN	cr 80 (N71) is
C_5H_{11} —  —CH=CH—  —CN	cr 55 N 101 is
Terminal moiety	
C_5H_{11} —  —C(=O)O— 	cr 30 is
C_5H_{11} —  —C(=O)O— 	cr 48 (N 45) is
C_5H_{11} —  —C(=O)O— 	cr 41 N 71 is
C_5H_{11} —  —C(=O)O— 	cr 78 (N 48) is
C_5H_{11} —  —C(=O)O— 	cr 47 N 79 is
C_5H_{11} —  —C(=O)O— 	cr 53 (N 38) is
C_5H_{11} —  —C(=O)O— 	cr 87 (N 86) is

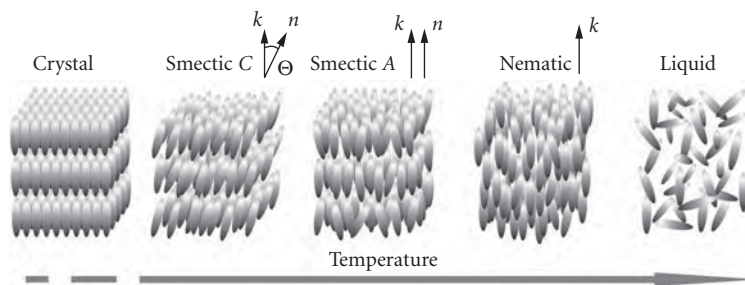


FIGURE 8 Schematic arrangement of the molecules in basic nematic and smectic phases.

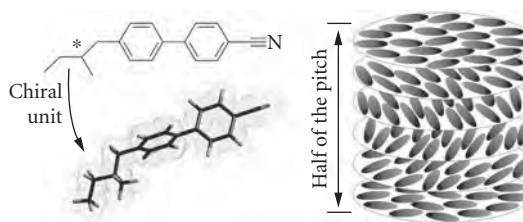


FIGURE 9 Chirality of the calamitic molecule and chiral nematic structure.

different properties. The cholesteric (or chiral nematic) liquid crystal phase is typically composed of nematic mesogenic molecules containing a chiral center which produces intermolecular forces that favor alignment between molecules at a slight angle to one another. When the molecules that make up a nematic liquid crystal are chiral, the chiral nematic phase will exist instead of the normal nematic. In this phase, the molecules prefer to lie next to each other in a slightly skewed orientation. This leads to the formation of a structure which can be visualized as a stack of very thin 2-D nematic-like layers with the director in each layer twisted with respect to those above and below, as Fig. 9 depicts. In this structure, the directors actually form a continuous helical pattern around the layer normal. The molecules shown here are merely representations of the many chiral nematic mesogens lying in the slabs of infinitesimal thickness with a distribution of orientation around the director. This is not to be confused with the planar arrangement found in smectic mesophases.

An important characteristic of the cholesteric mesophase is the pitch. The pitch, p , is defined as the distance it takes for the director to rotate one full turn in the helix as illustrated in Fig. 9. A by-product of the helical structure of the chiral nematic phase is its ability to selectively reflect light with central wavelength located $\lambda_0 = p \cdot \langle n \rangle$ and bandwidth $\Delta\lambda = p \cdot \Delta n$, where $\langle n \rangle$ represents the average refractive index and Δn , the birefringence of the LC. The cholesteric display reflects color without using color filters. Thus, its brightness is about $3\times$ higher than those using color filters. This Bragg reflection is established by the helical pitch of the cholesteric layers. The angle the LC director changes can be made larger by increasing the temperature of the sample, and thus tighten the pitch. Hence, more thermal energy will result from the increased temperature. Similarly, decreasing the temperature of the chiral sample increases the pitch length of the chiral nematic liquid crystal. Similar pitch changes are possible by applying electromagnetic field to aligned cholesteric samples. An interesting phenomenon of chiral nematic phase is that chirality can be introduced to the nonchiral nematic material by adding a small amount of chiral nematic mesogens. Not necessarily all the molecules have to be chiral. Sometimes, slightly below phase transition to the isotropic state (clearing point), some anomalous phases appear. They are known as blue phases (BPs).^{6,7} In many chiral compounds, with sufficiently

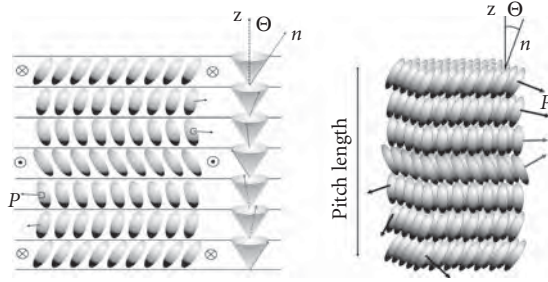


FIGURE 10 Structure of the ferroelectric smectic C^* phase.

high twist, up to three distinct blue phases appear. The two low-temperature phases, blue phase I (BP-I) and blue phase II (BP-II), have cubic symmetry. The highest temperature phase (closest to the clearing point), blue phase III (BP-III) appears to be amorphous.

Similar to chiral nematics, there are chiral forms of smectic phases. The only untilted chiral smectic phase is S_A^* . The most important feature of the chiral smectic phases with a tilted structure is ferroelectricity.^{8,9} Due to their low symmetry, chiral smectic phases are able to exhibit spontaneous polarization (P_s) that is oriented perpendicular to the director n and parallel to the smectic layer plane. Ferroelectric smectic C^* phase (S_{C^*}) is the most known tilted chiral smectic phase. The structure of the S_{C^*} has layers of molecules which are tilted in each layer at a temperature-dependent angle (θ) to the layer normal. Additionally, there is a slight and continuous change in the direction of the molecular tilt between adjacent layers as described in Fig. 10. In a macroscopic sample, without surfaces or external electric field, tilt will follow the helix and result in zero total spontaneous polarization. However, if a strong electric field is applied, the helix is unwound and nonzero spontaneous polarization can be observed. Recently, more phases closely related to the ferroelectric S_{C^*} have been discovered. In these phases, the layer spacing and the polarization direction are related in a different manner than in S_{C^*} phase. In an antiferroelectric smectic C^* phase ($S_{C^*}^{\text{anti}}$),¹⁰ the spontaneous polarization and the molecular tilt in adjacent layers are pointing in alternating directions, as Fig. 11 shows. Thus, for the macroscopic sample, both average spontaneous polarization and ideally, average molecular tilts are zero. Sufficiently strong electric fields will switch the antiferroelectric order to ferroelectric order. An important difference between ferroelectric and antiferroelectric smectic C^* phase is that antiferroelectric repeats its helical structure every 180° of rotation about the layer normal, compared to 360° for the ferroelectric phase. Many existing ferroelectric phases are different in the proportion of number of the layers with opposite directions. Similar to blue phases in the chiral nematic, the twist grain boundary (TGB) phases appear in chiral smectics.^{6,11} Just like in the blue

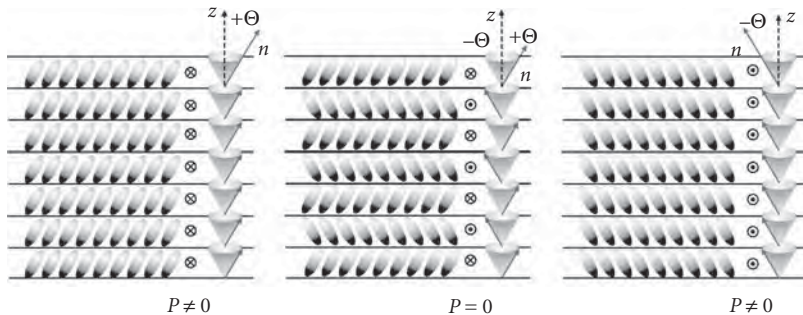


FIGURE 11 Structure of the antiferroelectric smectic C^* phase.

phases, high chirality subtly changes the energy of the system leading to a different type of structure. In this case, the free energy is minimized by introducing grain boundaries at periodic intervals. We may observe different TGB phases depending on the base of the phase from which they were developed. For example, Smectic C^* will generate a TGB $_C$ phase.^{6,12}

8.5 PHYSICAL PROPERTIES

The molecular order existing in the liquid crystalline phases induces anisotropy in the system. What it means is that all directions in the system are not equivalent to each other due to the shape of the molecules and the molecular distribution along the director $\mathbf{n}$. The anisotropy of the physical property is the most useful feature of liquid crystals and enables electro-optical application of LC materials. The physical properties can be divided into scalar and nonscalar quantities. Typical scalar properties are the thermodynamic transition parameters such as transition temperatures, transition enthalpy, and entropy changes. The dielectric, diamagnetic, optical, elastic, and viscous properties are the most important nonscalar properties. We will concentrate on the physical properties of the nematic phase because we commonly employ this phase and its properties in electro-optical applications.

Phase Transitions

The difference in the transition temperatures between melting and clearing points give the range of stability of the mesophases. For polymorphous (more than one phase) substances, the higher ordered phase exhibits the lower transition temperatures. When a material melts, a change of state occurs from solid to liquid (mesophase) and this process requires energy (endothermic) from the surroundings. If several mesophases exist, then several transitions will occur. The melting transition has typically $\sim 10\times$ larger enthalpy change (30 to 40 kJ/mol) than a transition between different mesophases (3 to 5 kJ/mol). The large enthalpy change is due to the drastic structural changes during the melting process; however, there is only a small difference between the different mesophases. One of the richest polymorphism of the single LC compound is shown by Terephthalylidene-bis-p-n-pentylaniline (TBPA) with six different mesophases, as shown in Fig. 12. A number of techniques (optical microscopy with hot stage, polarizing optical microscopy, differential thermal analysis, differential scanning calorimetry, etc.) may be used for determining the phase transition temperatures. Polarizing microscopy with hot stage and differential scanning calorimetry is particularly useful for the measurement of the phase transition temperatures. By using both methods, one can determine the number and type of mesophases and also the exact phase transitions, temperature, and enthalpy change associated with each transition. These are the crucial parameters to determine the components of eutectic mixture formulation for devices. A majority of the single liquid crystal components do not possess adequate range of required mesophase. Therefore, the

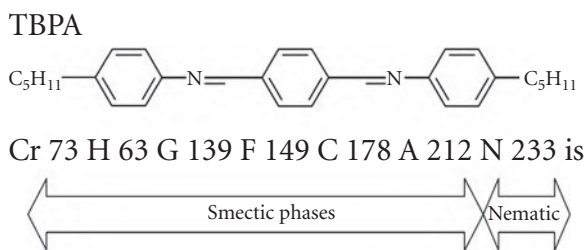


FIGURE 12 Molecular structure and mesomorphic properties of the TBPA.

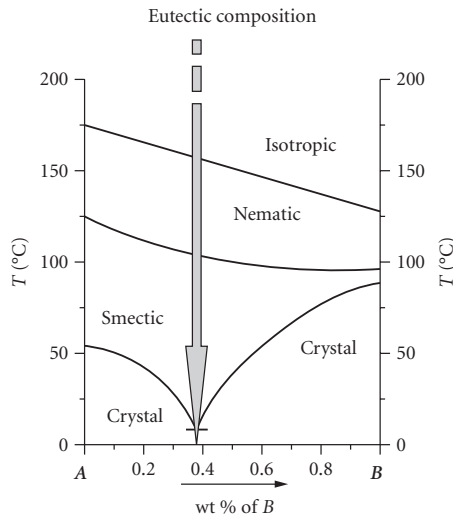


FIGURE 13 Simple phase diagram of the binary LC mixture.

eutectic compositions are required to lower the melting point of the material. It is known that the melting point of a binary (or higher number components) mixture is less than either of its constituent compounds. Figure 13 shows the phase diagram of a binary mixture. The mesogenic range for components 1 and 2 are shown in the two boundary vertical axes. The horizontal axis represents the molar concentration (X_2) of component 2. At a certain molar concentration, the melting point of the mixture will reach its minimum. Meanwhile, the clearing point of the mixture is linearly proportional to the molar concentration. The eutectic mixture calculation is based on the Schroder-Van Laar equation:^{13,14}

$$T_i = \frac{\Delta H_{fi}}{\frac{\Delta H_{fi}}{T_{fi}} - R \ln(X_i)} \quad (3)$$

where T_i is the temperature at which the pure component melts in the mixture, ΔH_{fi} is the heat fusion enthalpy of the pure component i , T_{fi} is the melting point of the pure component i , R is the gas constant (1.98 cal/mol-K), and X_i is the mole fraction of the pure component i . The clearing point (T_c) of the eutectic mixture can be estimated from the clearing points (T_{ci}) of the individual components (X_i) as

$$T_c = \sum_i X_i T_{ci} \quad (4)$$

Dielectric Properties

It was already mentioned, that liquid crystals exhibit anisotropy in many of their physical properties. It is well known that liquid crystals as dielectric and diamagnetic materials are sensitive to the external electric and magnetic fields. Dielectric studies are in general concerned with the response

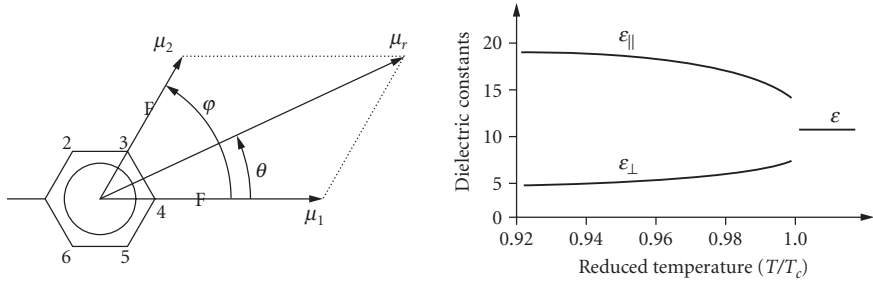


FIGURE 14 Dipole moment of the simple mesogenic structure.

of materials to the application of an electric field. Dielectric constants and their anisotropy affect the sharpness of the voltage-dependent optical transmission curve of a LC device and its threshold voltage.

The nematic liquid crystals are uniaxially symmetric to the axes of the director $\mathbf{n}$, and the dielectric constants differ in value along the preferred axis ($\epsilon_{\parallel}$) and perpendicular to this axis ($\epsilon_{\perp}$). They are mainly determined by the dipole moment, μ , its angle, θ , with respect to the principal molecular axis, and order parameter, S . When the molecule has two (or more) polar groups with dipole moment μ_1 and μ_2 , its effective dipole can be calculated by the vector addition method, see Fig. 14. In Fig. 14, the first dipole is along the principal molecular axis and the second dipole μ_2 is at an angle ϕ with respect to the principal molecular axis (and μ_1). The resultant dipole moment μ_r can be calculated from following equation:

$$\mu_r = (\mu_1^2 + \mu_2^2 + 2\mu_1\mu_2 \cos\phi)^{1/2} \quad (5)$$

Maier and Meier¹⁵ have developed a mean field theory to correlate the microscopic molecular parameters with the macroscopic dielectric constants of anisotropic LCs:

$$\epsilon_{\parallel} = NhF\{\langle\alpha_{\parallel}\rangle + (F\mu^2/3kT)[1 - (1 - 3\cos^2\theta)S]\} \quad (6a)$$

$$\epsilon_{\perp} = NhF\{\langle\alpha_{\perp}\rangle + (F\mu^2/3kT)[1 + (1 - 3\cos^2\theta)S/2]\} \quad (6b)$$

$$\Delta\epsilon = NhF\{(\langle\alpha_{\parallel}\rangle - \langle\alpha_{\perp}\rangle) - (F\mu^2/2kT)(1 - 3\cos^2\theta)S\} \quad (6c)$$

where N is the molecular packing density, $h = 3\epsilon/(2\epsilon + 1)$ is the cavity field factor, $\epsilon = (\epsilon_{\parallel} + 2\epsilon_{\perp})/3$ is the averaged dielectric constant, F is the Onsager reaction field, $\langle\alpha_{\parallel}\rangle$ and $\langle\alpha_{\perp}\rangle$ are the principal elements of the molecular polarizability tensor, β is the angle between the dipole moment μ and the principal molecular axis, and S is the order parameter of the second rank. From Eq. (6), the dielectric constants of anisotropic liquid crystals are influenced by the molecular structure, temperature, and frequency. These individual effects are discussed separately.

Structural Effect In general, $\Delta\epsilon$ depends on the molecular constituents, temperature, and frequency. For a nonpolar liquid crystal compound, its dipole moment $\mu \sim 0$. Thus, its $\Delta\epsilon$ is expected to be small and its magnitude is proportional to the differential molecular polarizability, similar to birefringence. On the other hand, for a LC molecule containing a polar group, such as cyano, isocyanate, fluoro, or chloro group, its $\Delta\epsilon$ can be positive or negative depending on the position(s) of the polar group(s). If $\beta \sim 0$, that is, the dipole moment of the polar group is along the principal molecular axis, $\Delta\epsilon$ is large and positive. Cyano-biphenyls are such examples. The $\Delta\epsilon$ of 5CB is about 10 at $T = 20^\circ\text{C}$ and $f = 1$ kHz. These positive $\Delta\epsilon$ materials are useful for parallel or twist alignment. On the contrary, if $\beta > 55^\circ$, $1 - 3\cos^2\beta > 0$ and $\Delta\epsilon$ may become negative depending on the dipole moment as indicated in Eq. (6c).

From Eq. (6c), for a nonpolar compound, $\mu \sim 0$ and its dielectric anisotropy is very small ($\Delta\epsilon < 0.5$). In this case, $\Delta\epsilon$ is determined mainly by the differential molecular polarizability, that is, the first term in Eq. (6c). For a polar compound, the dielectric anisotropy depends on the dipole moment, angle θ , temperature (T) and applied frequency. If a LC has more than one dipole, then the resultant dipole moment is their vector summation. In a phenyl ring, the position of the dipole is defined as

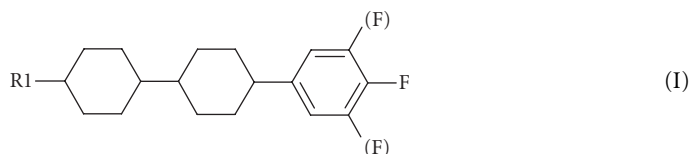


From Eq. (6c), if a polar compound has an effective dipole at $\theta < 55^\circ$, then its $\Delta\epsilon$ is positive. On the other hand, $\Delta\epsilon$ becomes negative if $\theta > 55^\circ$.

Fluoro (F)¹⁶, cyano (CN)¹⁷, and isothiocyanato (NCS)¹⁸ are the three commonly employed polar groups. Among them, fluoro group possess a modest dipole moment ($\mu \sim 1.5$ Debye), high resistivity, and low viscosity. However, its strong negativity compresses the electron clouds and, subsequently, lowers the compound's birefringence. For direct-view LCDs, the required birefringence is around 0.1, depending on the LC alignment and cell gap (d) employed. On the other hand, the cyano and isothiocyanato groups not only exhibit a large dipole moment ($\mu \sim 3.9$ Debye for $C\equiv N$ and ~ 3.7 Debye for $N=C=S$) but also contribute to lengthen the π -electron conjugation. As a result, their birefringence is much higher than their fluorinated counterpart. High birefringence is favorable for long wavelength, such as infrared applications in order to use a thin cell gap to achieve the same phase change while keeping a fast response time. Under strong anchoring condition, the LC response time is proportional to d^2 . However, the CN compounds are more viscous than the corresponding NCS and fluoro compounds. Therefore, their major applications are in low-end displays such as wrist watches and calculators where response time is not crucial.

Example 1: Positive $\Delta\epsilon$ LCs Positive $\Delta\epsilon$ LCs have been used in twisted nematic (TN)¹⁹ and in-plane switching (IPS)^{20,21} displays, although IPS can also use negative $\Delta\epsilon$ LCs. For TFT LCD, the employed LC material must also possess a high resistivity ($>10^{13} \Omega\text{-cm}$) in order to steadily hold the charges and avoid image flickering.²² The resistivity of a LC mixture depends heavily on the impurity contents, for example, ions. Purification process plays an important role in removing the ions for achieving high resistivity. Fluorinated compounds exhibit a high resistivity and are the natural choices for TFT LCDs.^{23,24}

A typical fluorinated LC structure is shown below:



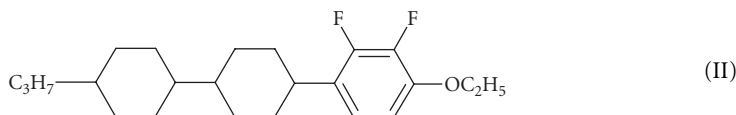
Most liquid crystal compounds discovered so far possess at least two rings, either cyclohexane-cyclohexane, cyclohexane-phenyl or phenyl-phenyl, and a flexible alkyl or alkoxy chain. The compound shown in structure (I) has two cyclohexane and one phenyl rings. The R_1 group represents a terminal alkyl chain, and a single or multiple fluoro substitutions take place in the phenyl ring. For multiple dipoles, the net dipole moment can be calculated from their vector sum. From Eq. (6c), to obtain the largest $\Delta\epsilon$ for a given dipole, the best position for the fluoro substitution is along the principal molecular axis, that is, in the fourth position. The single fluoro compound should have $\Delta\epsilon \sim 5$. To further increase $\Delta\epsilon$, more fluoro groups can be added. For example, compound (I) has two more fluoro groups in the third and fifth positions.²⁴ Its $\Delta\epsilon$ is about 10, but its birefringence would slightly decrease (because of the lower molecular packing density) and viscosity increases substantially (because of the higher moment of inertia). The birefringence of compound (I) is around 0.07. If a higher birefringence is needed, the middle cyclohexane ring can be replaced by a phenyl ring. The elongated electron cloud will enhance the birefringence to approximately 0.12 without increasing the viscosity noticeably.

The phase transition temperatures of a LC compound are difficult to predict before the compound is synthesized. In general, the lateral fluoro substitution lowers the melting temperature of the parent

compound because the increased intermolecular separation leads to a weaker molecular association. Thus, a smaller thermal energy is able to separate the molecules which implies to a lower melting point. A drawback of the lateral substitution is the increased viscosity.

Example 2: Negative $\Delta\epsilon$ LCs From Eq. (6c), in order to obtain a negative dielectric anisotropy, the dipoles should be in the lateral (2,3) positions. For the interest of obtaining high resistivity, lateral difluoro group is a favorable choice. The negative $\Delta\epsilon$ LCs are useful for vertical alignment.²⁵ The VA cell exhibits an unprecedented contrast ratio when viewed at normal direction between two crossed linear polarizers.^{26,27} However, a single domain VA cell has a relatively narrow viewing angle and is only useful for projection displays. For wide-view LCDs, a multidomain (4 domains) vertical alignment (MVA) cell is required.²⁸

The following structure is an example of the negative $\Delta\epsilon$ LC:²⁹



Compound (II) has two lateral fluoro groups in the (2,3) positions so that their dipoles in the horizontal components are perfectly cancelled whereas the vertical components add up. Thus, the net $\Delta\epsilon$ is negative. A typical $\Delta\epsilon$ of lateral difluoro compounds is -4 . The neighboring alkoxy group also has a dipole in the vertical direction. Therefore, it contributes to enlarge the dielectric anisotropy ($\Delta\epsilon \sim -6$). However, the alkoxy group has a higher viscosity than its alkyl counterpart and it also increases the melting point by $\sim 20^\circ$.

Temperature Effect In general, as temperature rises $\epsilon_{\parallel}$ decreases but $\epsilon_{\perp}$ increases gradually resulting in a decreasing $\Delta\epsilon$. From Eq. (6c) the temperature dependence of $\Delta\epsilon$ is proportional to S for the nonpolar LCs and S/T for the polar LCs. At $T > T_c$, the isotropic phase is reached and dielectric anisotropy vanishes, as Fig. 14 shows.

Frequency Effect From Eq. (6), two types of polarizations contribute to the dielectric constant: (1) induced polarization (the first term), and (2) orientation polarization (the dipole moment term). The field-induced polarization has a very fast response time, and it follows the alternating external field. But the permanent dipole moment associated orientation polarization exhibits a longer decay time, τ . If the external electric field frequency is comparable to $1/\tau$, the time lag between the average orientation of the dipole moments and the alternating field becomes noticeable. At a frequency $\omega (=2\pi f)$ which is much higher than $1/\tau$, the orientation polarization cannot follow the variations of the external field any longer. Thus, the dielectric constant drops to $\epsilon_{\parallel}$ which is contributed solely by the induced polarization:³⁰

$$\epsilon_{\parallel}(\omega) = \epsilon_{\infty} + \frac{\epsilon_{\parallel} - \epsilon_{\infty}}{1 + \omega^2 \tau^2} \quad (7)$$

where $\epsilon_{\parallel} = \epsilon_{\parallel}(\omega = 0)$ and $\epsilon_{\infty} = \epsilon_{\parallel}(\omega = \infty)$ are the parallel component of the dielectric constant at static and high frequencies, respectively.

In an aligned LC, the molecular rotation around their short axis is strongly hindered. Thus, the frequency dispersion occurs mainly at $\epsilon_{\parallel}$, while $\epsilon_{\perp}$ remains almost constant up to mega-Hertz region. Figure 15 shows the frequency dependent dielectric constants of the M1 LC mixture (from Roche) at various temperatures.³¹ As the frequency increases $\epsilon_{\parallel}$ decreases and beyond the crossover frequency f_c , $\Delta\epsilon$ changes sign. The dielectric anisotropies of M1 are fairly symmetric at low and high frequencies. The crossover frequency is sensitive to the temperature. As temperature rises, the $\epsilon_{\parallel}$ and $\epsilon_{\perp}$ of M1 both decrease slightly. However, the frequency-dependent $\epsilon_{\parallel}$ is strongly dependent on the temperature, but $\epsilon_{\perp}$ is inert. Thus, the crossover frequency increases exponentially with the temperature as: $f_c \sim \exp(-E/kT)$, where E is the activation energy. For M1 mixture, $E = 0.96$ eV.³¹

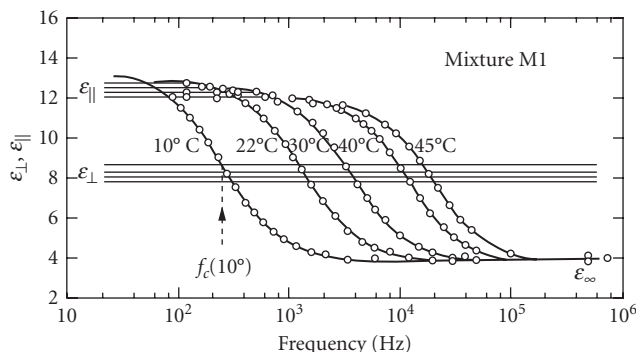


FIGURE 15 The frequency dependent dielectric constants of the M1 LC mixture (from Roche). (Redrawn from Ref. 31.)

Dual frequency effect is a useful technique for improving the response times of a LC device.³² In the dual frequency effect, a low frequency ($f < f_c$, where $\Delta\epsilon > 0$) an electric field is used to drive the device to its ON state, and during the decay period a high frequency ($f > f_c$, where $\Delta\epsilon < 0$) electric field is applied to speed up the relaxation process. From material standpoint, a LC mixture with low f_c and large $|\Delta\epsilon|$ at both low and high frequencies is beneficial. But for a single LC substance (such as cyanobiphenyls), its f_c is usually too high ($> 10^6$ Hz) to be practically employed.

In such a high frequency regime, the imaginary part of dielectric constant (which is responsible for absorption) becomes so significant that the dielectric heating effect is amplified and heats up the LC. The electro-optic properties of the cell are then altered. The imaginary part of dielectric constant contributes to heat. Thus, if a LC device is operated at MHz frequency region, significant heating effect due to the applied voltage will take place. This heating effect may be large enough to change all the physical properties of the LC. Dielectric heating is more severe if the crossover frequency is high.^{33–35}

Thus, liquid crystals are useful electro-optic media in the spectral range covering from UV, visible, IR to microwave. Of course, in each spectral region, an appropriate LC material has to be selected. For instance, the totally saturated LC compounds should be chosen for UV application because of photostability. For flat panel displays, LCs with a modest conjugation are appropriate. On the other hand, highly conjugated LCs are favorable for IR and microwave applications for the interest of keeping fast response time.

Optical Properties

Refractive indices and absorption are fundamentally and practically important parameters of a LC compound or mixture.³⁶ Almost all the light modulation mechanisms are involved with refractive index change. The absorption has a crucial impact on the photostability or lifetime of the liquid crystal devices. Both refractive indices and absorption are determined by the electronic structures of the liquid crystal studied.

The major absorption of a LC compound occurs in ultraviolet (UV) and infrared (IR) regions. The $\sigma \rightarrow \sigma^*$ electronic transitions take place in the vacuum UV (100 to 180 nm) region whereas the $\pi \rightarrow \pi^*$ electronic transitions occur in the UV (180 to 400 nm) region. Figure 16 shows the measured polarized UV absorption spectra of 5CB.³⁷ The λ_1 band which is centered at ~ 200 nm consists of two closely overlapped bands. The λ_2 band shifts to ~ 282 nm. The λ_2 band should occur in the vacuum UV region ($\lambda_0 \sim 120$ nm) which is not shown in the figure.

Refractive Indices Refractive index has great impact on LC devices. Almost every electro-optic effect of LC modulators, no matter amplitude or phase modulation, involves refractive index

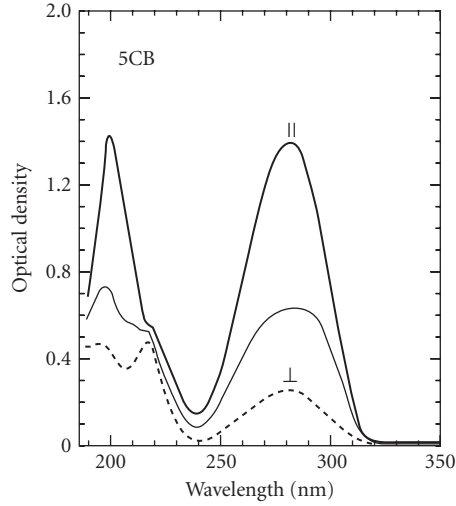


FIGURE 16 The measured polarized absorption spectra of 5CB. The middle trace is for unpolarized light. $\lambda_1 \sim 200$ nm and $\lambda_2 \sim 282$ nm.

change. An aligned LC exhibits anisotropic properties, including dielectric, elastic, and optical anisotropies. Let us take a homogeneous alignment as an example.³⁸ Assume a linearly polarized light is incident to the LC cell at normal direction. If the polarization axis is parallel to the LC alignment axis (i.e., LC director which represents an average molecular distribution axis), then the light experiences the extraordinary refractive index, n_e . If the polarization is perpendicular to the LC directors, then the light sees the ordinary refractive index n_o . The difference between n_e and n_o is called birefringence, defined as $\Delta n = n_e - n_o$. Refractive indices are dependent on the wavelength and temperature. For a full-color LCD, RGB color filters are employed. Thus, the refractive indices at these wavelengths need to be known in order to optimize the device performance. Moreover, about 50 percent of the backlight is absorbed by the polarizer. The absorbed light turns into heat and causes the LCD panel's temperature to increase. As the temperature increases, refractive indices decrease gradually. The following sections will describe how the wavelength and temperature affect the LC refractive indices.

Wavelength Effect Based on the electronic absorption, a three-band model which takes one $\sigma \rightarrow \sigma^*$ transition (the λ_0 -band) and two $\pi \rightarrow \pi^*$ transitions (the λ_1 - and λ_2 -bands) into consideration has been developed. In the three band model, the refractive indices (n_e and n_o) are expressed as follows:^{39,40}

$$n_{e,o} \cong 1 + g_{0e,o} \frac{\lambda^2 \lambda_0^2}{\lambda^2 - \lambda_0^2} + g_{1e,o} \frac{\lambda^2 \lambda_1^2}{\lambda^2 - \lambda_1^2} + g_{2e,o} \frac{\lambda^2 \lambda_2^2}{\lambda^2 - \lambda_2^2} \quad (8)$$

The three-band model clearly describes the origins of refractive indices of LC compounds. However, a commercial mixture usually consists of several compounds with different molecular structures in order to obtain a wide nematic range. The individual λ_i 's are therefore different. Under such a circumstance, Eq. (8) would have too many unknowns to quantitatively describe the refractive indices of a LC mixture.

In the off-resonance region, the right three terms in Eq. (8) can be expanded by a power series to the λ^{-4} terms to form the extended Cauchy equations for describing the wavelength-dependent refractive indices of *anisotropic* LCs:^{40,41}

$$n_{e,o} \cong A_{e,o} + \frac{B_{e,o}}{\lambda^2} + \frac{C_{e,o}}{\lambda^4} \quad (9)$$

In Eq. (9), $A_{e,o}$, $B_{e,o}$, and $C_{e,o}$ are known as Cauchy coefficients. Although Eq. (9) is derived based on a LC compound, it can be extended easily to include eutectic mixtures by taking the superposition of each compound. From Eq. (9) if we measure the refractive indices at three wavelengths, the three Cauchy coefficients ($A_{e,o}$, $B_{e,o}$, and $C_{e,o}$) can be obtained by fitting the experimental results. Once these coefficients are determined, the refractive indices at any wavelength can be calculated. From Eq. (9) both refractive indices and birefringence decrease as the wavelength increases. In the long wavelength (IR and millimeter wave) region, n_e and n_o are reduced to A_e and A_o , respectively. The coefficients A_e and A_o are constants; they are independent of wavelength, but dependent on the temperature. That means, in the IR region the refractive indices are insensitive to wavelength, except for the resonance enhancement effect near the local molecular vibration bands. This prediction is consistent with many experimental evidences.⁴²

Figure 17 depicts the wavelength-dependent refractive indices of E7 at $T = 25^\circ\text{C}$. Open squares and circles represent the n_e and n_o of E7 in the visible region while the downward- and upward-triangles stand for the measured data at $\lambda = 1.55$ and $10.6 \mu\text{m}$, respectively. Solid curves are fittings to the experimental n_e and n_o data in the visible spectrum by using the extended Cauchy equations [Eq. (9)]. The fitting parameters are listed as follows: ($A_e = 1.6933$, $B_e = 0.0078 \mu\text{m}^2$, $C_e = 0.0028 \mu\text{m}^4$) and ($A_o = 1.4994$, $B_o = 0.0070 \mu\text{m}^2$, $C_o = 0.004 \mu\text{m}^4$). In Fig. 17, the extended Cauchy model is extrapolated to the near- and far-infrared regions. The extrapolated lines almost strike through the center of the experimental data measured at $\lambda = 1.55$ and $10.6 \mu\text{m}$. The largest difference between the extrapolated and experimental data is only 0.4 percent.

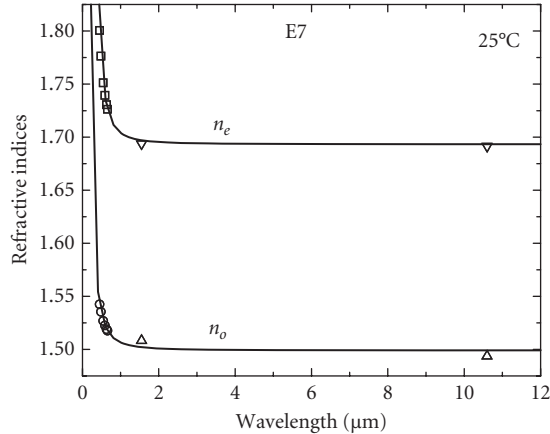


FIGURE 17 Wavelength-dependent refractive indices of E7 at $T = 25^\circ\text{C}$. Open squares and circles are the n_e and n_o measured in the visible spectrum. Solid lines are fittings to the experimental data measured in the visible spectrum by using the extended Cauchy equation [Eq. (4.9)]. The downward- and upward-triangles are the n_e and n_o measured at $T = 25^\circ\text{C}$ and $\lambda = 1.55$ and $10.6 \mu\text{m}$, respectively.

Equation (9) applies equally well to both high and low birefringence LC materials in the off-resonance region. For low birefringence ($\Delta n < 0.12$) LC mixtures, the λ^{-4} terms are insignificant and can be omitted and the extended Cauchy equations are simplified as:⁴³

$$n_{e,o} \cong A_{e,o} + \frac{B_{e,o}}{\lambda^2} \quad (10)$$

Thus, n_e and n_o each has only two fitting parameters. By measuring the refractive indices at two wavelengths, we can determine $A_{e,o}$ and $B_{e,o}$. Once these two parameters are determined, n_e and n_o can be calculated at any wavelength of interest. Because most of TFT LC mixtures have $\Delta n \sim 0.1$, the two-coefficient Cauchy model is adequate to describe the refractive index dispersions. Although the extended Cauchy equation fits experimental data well,⁴⁴ its physical origin is not clear. A better physical meaning can be obtained by the three-band model which takes three major electronic transition bands into consideration.

Temperature Effect The temperature effect is particularly important for projection displays.⁴⁵ Due to the thermal effect of the lamp, the temperature of the display panel could reach 50°C. It is important to know the LC properties at the anticipated operating temperature beforehand.

Birefringence Δn is defined as the difference between the extraordinary and ordinary refractive indices, $\Delta n = n_e - n_o$ and the average refractive indices $\langle n \rangle$ is defined as $\langle n \rangle = (n_e + 2n_o)/3$. Based on these two definitions, n_e and n_o can be rewritten as

$$n_e = \langle n \rangle + \frac{2}{3} \Delta n \quad (11)$$

$$n_o = \langle n \rangle - \frac{1}{3} \Delta n \quad (12)$$

To describe the temperature dependent birefringence, the Haller approximation can be employed when the temperature is not too close to the clearing point:

$$\Delta n(T) = (\Delta n)_o (1 - T/T_c)^\beta \quad (13)$$

In Eq. (13), $(\Delta n)_o$ is the LC birefringence in the crystalline state (or $T = 0$ K), the exponent α is a material constant, and T_c is the clearing temperature of the LC material under investigation. On the other hand, the average refractive index decreases linearly with increasing temperature as⁴⁶

$$\langle n \rangle = A - BT \quad (14)$$

because the LC density decreases with increasing temperature. By substituting Eqs. (14) and (13) back to Eqs. (11) and (12), the 4-parameter model for describing the temperature dependence of the LC refractive indices is given as⁴⁷

$$n_e(T) \approx A - BT + \frac{2(\Delta n)_o}{3} \left(1 - \frac{T}{T_c}\right)^\beta \quad (15)$$

$$n_o(T) \approx A - BT - \frac{(\Delta n)_o}{3} \left(1 - \frac{T}{T_c}\right)^\beta \quad (16)$$

The parameters $[A, B]$ and $[(\Delta n)_o, \beta]$ can be obtained separately by two-stage fittings. To obtain $[A, B]$, one can fit the average refractive index $\langle n \rangle = (n_e + 2n_o)/3$ as a function of temperature using Eq. (14). To find $[(\Delta n)_o, \beta]$, one can fit the birefringence data as a function of temperature using Eq. (13). Therefore, these two sets of parameters can be obtained separately from the same set of refractive indices but at different forms.

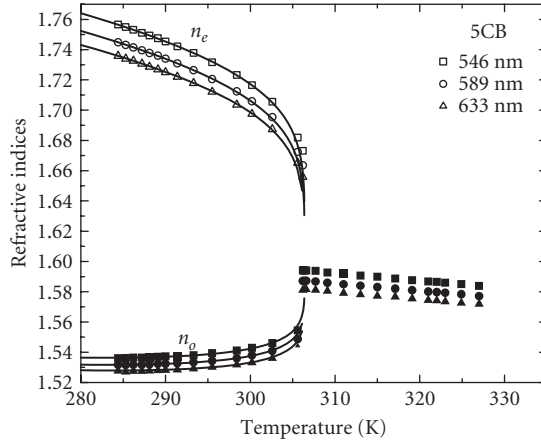


FIGURE 18 Temperature-dependent refractive indices of 5CB at $\lambda = 546, 589,$ and 633 nm. Squares, circles, and triangles are experimental data for refractive indices measured at $\lambda = 546, 589,$ and 633 nm, respectively.

Figure 18 is a plot of the temperature dependent refractive indices of 5CB at $\lambda = 546, 589,$ and 633 nm. As the temperature increases, n_e decreases, but n_o gradually increases. In the isotropic state, $n_e = n_o$ and the refractive index decreases linearly as the temperature increases. This correlates with the density effect.

Elastic Properties

The molecular order existing in liquid crystals has interesting consequences on the mechanical properties of these materials. They exhibit elastic behavior. Any attempt to deform the uniform alignments of the directors and the layered structures (in case of smectics) results in an elastic restoring force. The constants of proportionality between deformation and restoring stresses are known as elastic constants.

Elastic Constants Both threshold voltage and response time are related to the elastic constant of the LC used. There are three basic elastic constants involved in the electro-optics of liquid crystals depending on the molecular alignment of the LC cell: the splay (K_{11}), twist (K_{22}), and bend (K_{33}), as Fig. 19 shows. Elastic constants affect the liquid crystal's electro-optical cell in two aspects: the threshold voltage and the response time. The threshold voltage in the most common case of homogeneous electro-optical cell is expressed as follows:

$$V_{th} = \pi \sqrt{\frac{K_{11}}{\epsilon_0 \Delta \epsilon}} \quad (17)$$

Several molecular theories have been developed for correlating the Frank elastic constants with molecular constituents. Here we only introduce two theories: (1) the mean-field theory,^{48,49} and (2) the generalized van der Waals theory.⁵⁰

Mean-Field Theory In the mean-field theory, the three elastic constants are expressed as

$$K_{ii} = C_{ii} V_n^{-7/3} S^2 \quad (18a)$$

$$C_{ii} = (3A/2)(Lm^{-1}\chi_{ii}^{-2})^{1/3} \quad (18b)$$

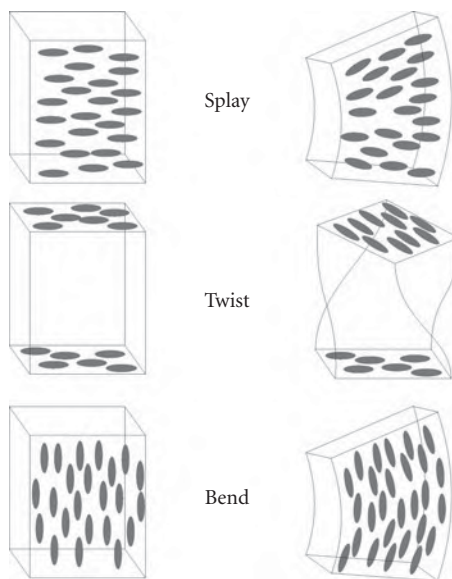


FIGURE 19 Elastic constants of the liquid crystals.

where C_{ii} is called the reduced elastic constant, V_n is the mole volume, L is the length of the molecule, m is the number of molecules in a steric unit in order to reduce the steric hindrance, $\chi_{11} = \chi_{22} = z/x$ and $\chi_{33} = (x/z)^2$, where x ($y = x$) and z are the molecular width and length, respectively, and $A = 1.3 \times 10^{-8} \text{ erg} \cdot \text{cm}^6$.

From Eq. (18), the ratio of K_{11} : K_{22} : K_{33} is equal to 1: 1: $(z/x)^2$ and the temperature dependence of elastic constants is basically proportional to S^2 . This S^2 dependence has been experimentally observed for many LCs. However, the prediction for the relative magnitude of K_{ii} is correct only to the first order. Experimental results indicate that K_{22} often has the lowest value, and the ratio of K_{33}/K_{11} can be either greater or less than unity.

Generalized van der Waals Theory Gelbart and Ben-Shaul⁵⁰ extended the generalized van der Waals theory for explaining the detailed relationship between elastic constants and molecular dimensions, polarizability, and temperature. They derived the following formula for nematic liquid crystals:

$$K_{ii} = a_i \langle P_2 \rangle \langle P_2 \rangle + b_i \langle P_2 \rangle \langle P_4 \rangle \quad (19)$$

where a_i and b_i represent sums of contributions of the energy and the entropy terms; they depend linearly on temperature, $\langle P_2 \rangle$ ($=S$) and $\langle P_4 \rangle$ are the order parameter of the second and the fourth rank, respectively. In general, the second term may not be negligible in comparison with the S^2 term depending on the value of $\langle P_4 \rangle$. As temperature increases, both S and $\langle P_4 \rangle$ decrease. If the $\langle P_4 \rangle$ of a LC is much smaller than S in its nematic range, Eq. (19) is reduced to the mean-field theory, or $K_{ii} \sim S^2$. The second term in Eq. (19) is responsible for the difference between K_{11} and K_{33} .

Viscosities

The resistance of fluid system to flow when subjected to a shear stress is known as viscosity. In liquid crystals several anisotropic viscosity coefficients may result, depending on the relative orientation

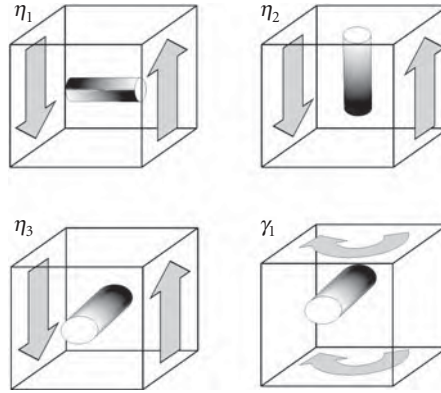


FIGURE 20 Anisotropic viscosity coefficients required to characterize a nematic.

of the director with respect to the flow of the LC material. When an oriented nematic liquid crystal is placed between two plates which are then sheared, four cases shown in Fig. 20 are to be studied. Three, are known as Miesowicz viscosity coefficients. They are η_1 —when director of LC is perpendicular to the flow pattern and parallel to the velocity gradient, η_2 —when director is parallel to the flow pattern and perpendicular to the velocity gradient and η_3 —when director is perpendicular to the flow pattern and to the velocity gradient. Viscosity, especially rotational viscosity γ_1 , plays a crucial role in the liquid crystals displays (LCD) response time. The response time of a nematic liquid crystals device is linearly proportional to γ_1 .

The rotational viscosity of an aligned LC is a complicated function of molecular shape, moment of inertia, activation energy, and temperature. Several theories, including both rigorous and semi-empirical, have been developed in an attempt to account for the origin of the LC viscosity. However, owing to the complicated anisotropic attractive and steric repulsive interactions among LC molecules, these theoretical results are not yet completely satisfactory. Some models fit certain LCs, but fail to fit others.⁵¹ In the molecular theory developed by Osipov and Terentjev,⁵² all six Leslie viscosity coefficients are expressed in terms of microscopic parameters:

$$\alpha_1 = -N\Lambda \frac{p^2 - 1}{p^2 + 1} P_4 \quad (20a)$$

$$\alpha_2 = -\frac{N\Lambda}{2} S - \frac{1}{2} \gamma_1 \equiv -\frac{N\Lambda}{2} \left[S + \frac{1}{6} \sqrt{J/kT} e^{j/kT} \right] \quad (20b)$$

$$\alpha_3 = -\frac{N\Lambda}{2} S + \frac{1}{2} \gamma_1 \equiv -\frac{N\Lambda}{2} \left[S - \frac{1}{6} \sqrt{J/kT} e^{j/kT} \right] \quad (20c)$$

$$\alpha_4 = \frac{N\Lambda}{35} \frac{p^2 - 1}{p^2 + 1} [7 - 5S - 2P_4] \quad (20d)$$

$$\alpha_5 = \frac{N\Lambda}{2} \left[\frac{p^2 - 1}{p^2 + 1} \frac{3S + 4P_4}{7} + S \right] \quad (20e)$$

$$\alpha_6 = \frac{N\Lambda}{2} \left[\frac{p^2 - 1}{p^2 + 1} \frac{3S + 4P_4}{7} - S \right] \quad (20f)$$

where N represents molecular packing density, p is the molecular length-to-width (w) ratio, $J \equiv J_\theta S$ is the Maier-Saupe mean-field coupling constant; k is the Boltzmann constant, T is the Kelvin temperature, S and P_4 are the nematic order parameters of the second and fourth order, respectively, and Λ is the friction coefficient of LC:

$$\Lambda \equiv 100 (1 - \Phi) N^2 w^6 p^{-2} [(kT)^5 / G^3] (I_\perp / kT)^{1/2} \cdot \exp [3(G + I_\theta) / kT] \quad (21)$$

where Φ is the volume fraction of molecules ($\Phi = 0.5$ to 0.6 for dense molecular liquids), $G \gg I_\theta$ is the isotropic intermolecular attraction energy, and $I_\perp$ and I_θ are the inertia tensors.

From the above analysis, the parameters affecting the rotational viscosity of a LC are⁵³

1. Activation energy ($E_a \sim 3G$): a LC molecule with low activation energy leads to a low viscosity.
2. Moment of inertia: a LC molecule with linear shape and low molecular weight would possess a small moment of inertia and exhibit a low viscosity.
3. Inter-molecular association: a LC with weak inter-molecular association, for example, not form dimer, would reduce the viscosity significantly.
4. Temperature: elevated temperature operation of a LC device may be the easiest way to lower viscosity. However, birefringence, elastic, and dielectric constants are all reduced as well.

8.6 LIQUID CRYSTAL CELLS

Three kinds of LC cells have been widely used for display applications. They are (1) twisted-nematic (TN) cell, (2) in-plane switching (IPS) cell, and multidomain vertical alignment (MVA) cell. For phase-only modulation, homogeneous cell is preferred. The TN cell dominates notebook market because of its high transmittance and low cost. However, its viewing angle is limited. For wide-view applications, for example, LCD TVs, optical film-compensated IPS and MVA are the two major camps. In this section, we will discuss the basic electro-optics of TN, IPS, and MVA cells.

Twisted-Nematic (TN) Cell

The first liquid crystal displays that became successful on the market were the small displays in digital watches. These were simple twisted nematic (TN) devices with characteristics satisfactory for such simple applications.¹⁹ The basic operation principals are shown in Fig. 21. In the TN display, each LC cell consists of a LC material sandwiched between two glass plates separated by a gap of 5 to 8 μm .

The inner surfaces of the plates are deposited with transparent electrodes made of conducting coatings of indium tin oxide (ITO). These transparent electrodes are overcoated with a thin layer of polyimide with a thickness of about 80 angstroms. The polyimide films are unidirectional rubbed with the rubbing direction of the lower substrate perpendicular to the rubbing direction of the upper surface. Thus, in the inactivated state (voltage OFF), the local director undergoes a continuous twist of 90° in the region between the plates. Sheet polarizers are laminated on the outer surfaces of the plates. The transmission axes of the polarizers are aligned parallel to the rubbing directions of the adjacent polyimide films. When light enters the cell, the first polarizer lets through only the component oscillating parallel to the LC director next to the entrance substrate. During the passage through the cell, the polarization plane is turned along with the director helix, so that when the light wave arrives at the exit polarizer, it passes unobstructed. The cell is thus transparent in the OFF state; this mode is called normally white (NW).

Figure 22 depicts the normalized voltage-dependent light transmittance ($T_\perp$) of the 90° TN cell at three primary wavelengths: $R = 650$ nm, $G = 550$ nm, and $B = 450$ nm. Since human eye has the greatest

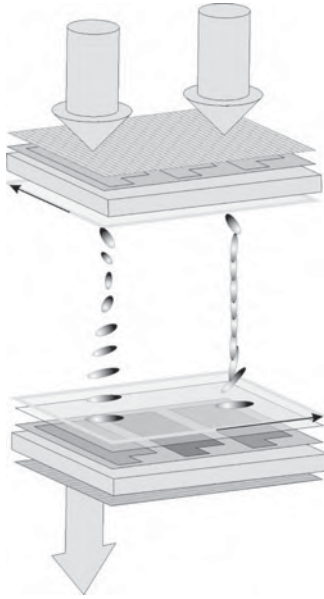


FIGURE 21 The basic operation principles of the TN cell.

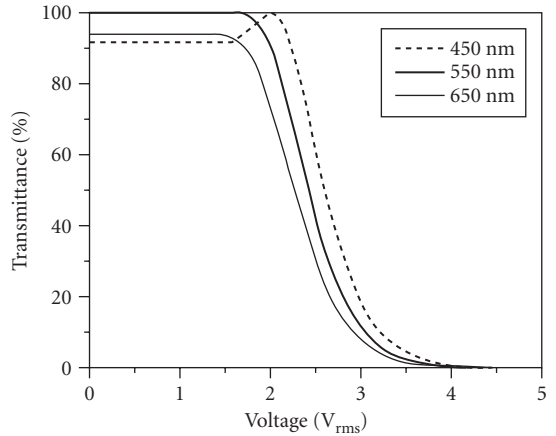


FIGURE 22 Voltage-dependent transmittance of a normally white 90° TN cell. $d\Delta n = 480$ nm.

sensitivity at green, we normally optimize the cell design at $\lambda \sim 550$ nm. To meet the Gooch-Tarry first minimum condition, for example, $d\Delta n = (\sqrt{3}/2)\lambda$, the employed cell gap is 5- μm and the LC birefringence is $\Delta n \sim 0.096$. From Fig. 22, the wavelength effect on the transmittance at $V = 0$ is within 8 percent. Therefore, the TN cell can be treated as an “achromatic” half-wave plate.

The response time of a TN LCD depends on the cell gap and the γ_1/K_{22} of the LC mixture employed. For a 5- μm cell gap, the optical response time is ~ 30 to 40 ms. At $V = 5 V_{\text{rms}}$, the contrast ratio (CR) reaches $\sim 500:1$. These performances, although not perfect, are acceptable for notebook computers. A major drawback of the TN cell is its narrow viewing angle and grayscale inversion originated from the LC directors tilting out of the plane. Because of this molecular tilt, the viewing angle in the vertical direction is narrow and asymmetric, and has grayscale inversion.⁵⁴ Despite a relatively slow switching time and limited viewing angle, the TN cell is still widely used for many applications because of its simplicity and low cost. Recently, a fast-response (~ 2 ms gray-to-gray response time) TN notebook computer has been demonstrated by using a thin cell gap (2 μm), low viscosity LC mixture, and overdrive and undershoot voltage method.

In-Plane Switching (IPS) Cell

In an IPS cell, the transmission axis of the polarizer is parallel to the LC director at the input plane.⁵⁵ The optical wave traversing through the LC cell is an extraordinary wave whose polarization state remains unchanged. As a result, a good dark state is achieved since this linearly polarized light is completely absorbed by the crossed analyzer. When an electric field is applied to the LC cell, the LC directors are reoriented toward the electric field (along the y axis). This leads to a new director distribution with a twist $f(z)$ in the xy plane as Fig. 23 shows.

Figure 24 depicts the voltage-dependent light transmittance of an IPS cell. The LC employed is MLC-6686, whose $\Delta\epsilon = 10$ and $\Delta n = 0.095$, the electrode width is 4 μm , electrode gap 8 μm , and cell

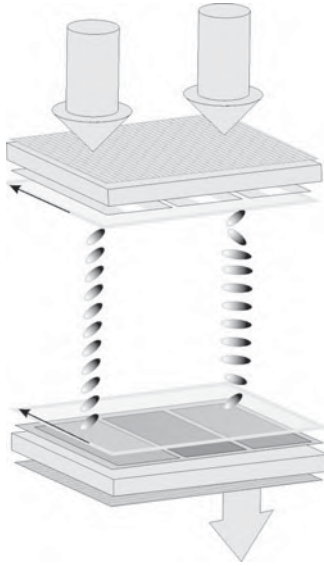


FIGURE 23 IPS mode LC display.

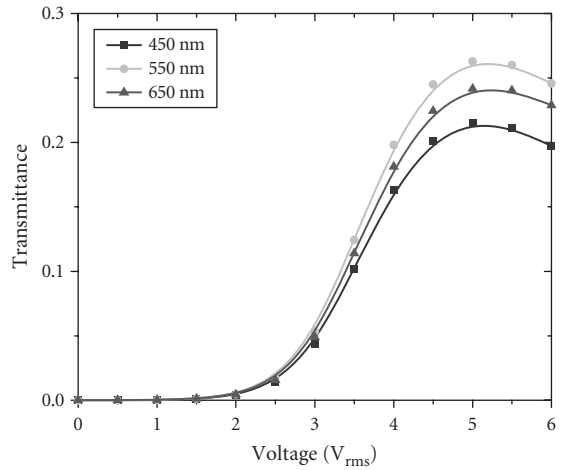


FIGURE 24 Voltage-dependent light transmittance of the IPS cell. LC: MLC-6686, $\Delta\epsilon = 10$, electrode width = $4\ \mu\text{m}$, gap = $8\ \mu\text{m}$, and cell gap $d = 3.6\ \mu\text{m}$.

gap $d = 3.6\ \mu\text{m}$. The threshold voltage occurs at $V_{th} \sim 1.5 V_{rms}$ and maximum transmittance at $\sim 5 V_{rms}$ for both wavelengths. Due to absorption, the maximum transmittance of the two polarizers (without LC cell) is 35.4, 33.7, and 31.4 percent for RGB wavelengths, respectively.

Vertical Alignment (VA) Cell

The vertically aligned LC cell exhibits an unprecedentedly high contrast ratio among all the LC modes developed.⁵⁶ Moreover, its contrast ratio is insensitive to the incident light wavelength, cell thickness, and operating temperature. In principle, in voltage-OFF state the LC directors are perpendicular to the cell substrates, as Fig. 25 shows. Thus, its contrast ratio at normal angle is limited by the crossed polarizers. Application of a voltage to the ITO electrodes causes the directors to tilt away from the normal to the glass surfaces. This introduces birefringence and subsequently light transmittance because the refractive indices for the light polarized parallel and perpendicular to the directors are different.

Figure 26 shows the voltage-dependent transmittance of a VA cell with $d\Delta n = 350\ \text{nm}$ between two crossed polarizers. Here, a single domain VA cell employing Merck high resistivity MLC-6608 LC mixture is simulated. Some physical properties of MLC-6608 are listed as follows: $n_e = 1.562$, $n_o = 1.479$ (at $\lambda = 546\ \text{nm}$ and $T = 20^\circ\text{C}$); clearing temperature $T_c = 90^\circ\text{C}$; $\Delta\epsilon = -4.2$, and rotational viscosity $\gamma_1 = 186\ \text{mPas}$ at 20°C .

From Fig. 26, an excellent dark state is obtained at normal incidence. As the applied voltage exceeds the Freederickz threshold voltage ($V_{th} \sim 2.1 V_{rms}$), LC directors are reoriented by the applied electric field resulting in light transmission from the crossed analyzer. From the figure, RGB wavelengths reach their peak at different voltages, blue at $\sim 4 V_{rms}$, and green at $\sim 6 V_{rms}$. The on-state dispersion is more forgiven than the dark state. A small light leakage in the dark state would degrade the contrast ratio significantly, but less noticeable from the bright state.

It should be mentioned that in Figs. 25 and 26 only a single domain is considered, thus, its viewing angle is quite narrow. To achieve wide view, four domains with film compensation are required. Several approaches, such as Fujitsu's MVA (multidomain VA) and Samsung's PVA (Patterned VA),

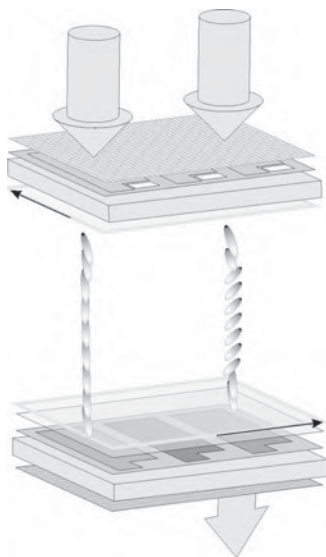


FIGURE 25 VA mode LC display.

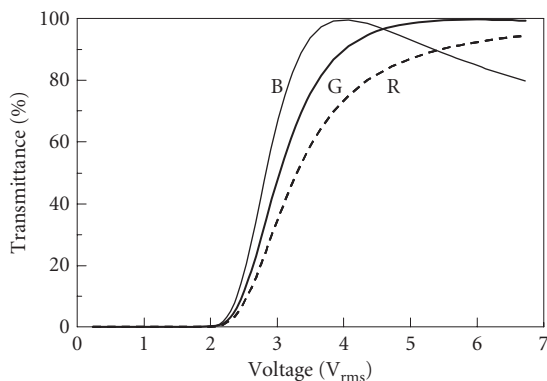


FIGURE 26 Voltage-dependent normalized transmittance of a VA cell. LC: MLC-6608. $d\Delta n = 350$ nm. $R = 650$ nm, $G = 550$ nm, and $B = 450$ nm.

have been developed for obtaining four complementary domains. Figure 27 depicts the PVA structure developed by Samsung.

As shown in Fig. 27, PVA has no pretilt angle. The four domains are induced by the fringe electric fields. On the other hand, in Fujitsu's MVA, physical protrusions are used to provide initial pretilt direction for forming four domains. The protrusions not only reduce the aperture ratio but also cause light leakage in the dark state because the LCs on the edges of protrusions are tilted so that they exhibit birefringence. It would be desirable to eliminate protrusions for MVA and create a pretilt angle in each domain for both MVA and PVA to guide the LC reorientation direction. Based on this concept, surface polymer sustained alignment (PSA) technique has been developed.⁵⁷ A very small percentage (~ 0.2 wt %) of reactive mesogen monomer and photoinitiator are mixed in a negative $\Delta\epsilon$ LC host and injected into a LCD panel. While a voltage is applied to generate four domains, a UV light is used to cure the monomers. As a result, the monomers are adsorbed onto the surfaces. These cured polymers, although in low density, will provide a pretilt angle within each domain to guide the LC reorientation. Thus, the rise time is reduced by nearly $2\times$ while the decay time remains more or less unchanged.⁵⁸

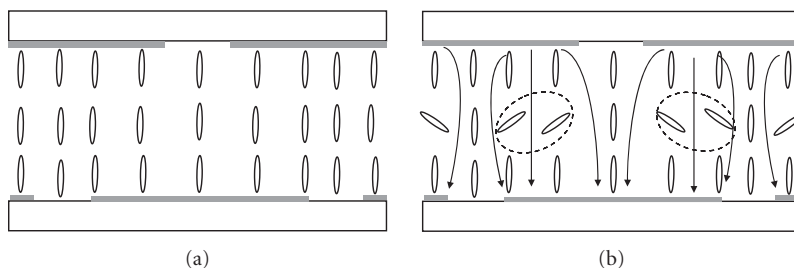


FIGURE 27 (a) LC directors of PVA at $V = 0$ and (b) LC directors of PVA at a voltage-on state. The fringe fields generated by the top and bottom slits create two opposite domains in this cross section. When the zigzag electrodes are used, four domains are generated.

8.7 LIQUID CRYSTALS DISPLAYS

The most common and well recognized applications of liquid crystals nowadays are displays. It is the most natural way to utilize extraordinary electro-optical properties of liquid crystals together with its liquid-like behavior. All the other applications are called as nondisplay applications of liquid crystals. Nondisplay applications are based on the liquid crystals molecular order sensitivity to the external incentive. This can be an external electric and magnetic field, temperature, chemical agents, mechanical stress, pressure, irradiation by different electromagnetic wave, or radioactive agents. Liquid crystals sensitivity for such wide spectrum of factors results in tremendous diversity of non-display applications. It starts from spatial light modulators for laser beam steering, adaptive optics through telecommunication area (light shutters, attenuators, and switches), cholesteric LC filters, LC thermometers, stress meters, dose meters to ends up with liquid crystals paints, and cosmetics. Another field of interest employing lyotropic liquid crystals is biomedicine where it plays an important role as a basic unit of the living organisms by means of plasma membranes of the living cells. More about existing nondisplay applications can be found in preferred reading materials.

Three types of liquid crystal displays (LCDs) have been developed: (1) transmissive, (2) reflective, and (3) transreflective. Each one has its own unique properties. In the following sections, we will introduce the basic operation principles of these display devices.

Transmissive TFT LCDs

A transmissive LCD uses a backlight to illuminate the LCD panel to achieve high brightness (300 to 500 nits) and high contrast ratio ($>2000:1$). Some transmissive LCDs, such as twisted-nematic (TN), do not use phase compensation films or multidomain structures so that their viewing angle is limited and they are more suitable for single viewer applications, for example, mobile displays and notebook computers. With phase compensation films and multidomain structures, the direct-view transmissive LCDs exhibit a wide viewing angle and high contrast ratio, and have been widely used for desktop computers and televisions. However, the cost of direct-view large screen LCDs is still relatively expensive. To obtain a screen diagonal larger than 2.5 m, projection displays, such as data projector, using transmissive microdisplays are still a favorable choice. There, a high power arc lamp or light emitting diode (LED) arrays are used as light source. Using a projection lens, the displayed image is magnified by more than 50 $\times$. To reduce the size of optics and cost, the LCD panel is usually made small (25 mm in diagonal) and each pixel size is ~ 20 to 40 μm . Thus, poly-silicon-based TFT LCD is the common choice.

Figure 28 shows the device structure of a transmissive thin-film-transistor (TFT) LCD using amorphous silicon (a-Si) transistors. LCD is a nonemissive display, that is, it does not emit light, instead, it functions as a two-dimensional spatial light modulator. Thus, a backlight is needed. A diffuser is used to homogenize the backlight in order to avoid hot spots. Some optical films are stacked to steer the Lambertian backlight to the central $\pm 40^\circ$ for improving display brightness. Since most LCDs require a linearly polarized light in order to obtain a high contrast ratio, two sheets of stretched dichroic polarizers are used for large size direct-view displays. The first glass substrate contains TFT arrays, which serve as independent light switches. Each display pixel is controlled by a TFT. Since TFT is sensitive and should be shielded from backlight illumination, the actual aperture ratio (the transparent ITO electrode area) is reduced to ~ 80 percent, depending on the pixel density. As the pixel density increases, the aperture ratio decreases. The LC layer is sandwiched between two ITO substrates whose inner surface is coated with a thin (80 to 100 nm) polyimide layer. Some LCDs (twisted-nematic, in-plane switching and fringe field switching) require rubbing but some (multidomain vertical alignment and patterned vertical alignment) do not. The cell gap is usually controlled at ~ 3.5 to 4.0 μm for a transmissive LCD. The performance of the display such as light throughput, response time, and viewing angle are all influenced by the LC configuration employed.

For direct-view LCDs, compact size, lightweight, and low power consumption are equally important as viewing angle, color, and contrast ratio. For direct-view LCDs, color filters are embedded in the inner side of the top (second) substrate. Three subpixels (red, green, and blue) form a color pixel. The size of

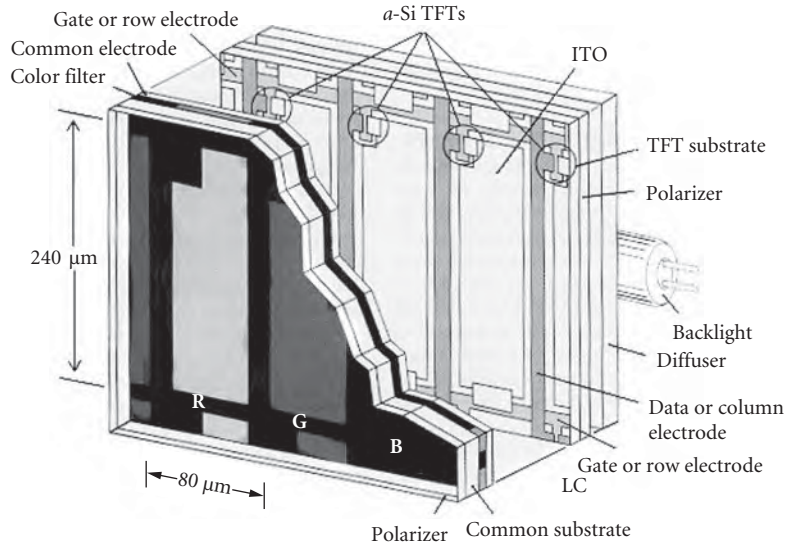


FIGURE 28 Device structure of a color pixel of a transmissive TFT LCD.

each subpixel is $\sim 80 \mu\text{m} \times 240 \mu\text{m}$. Each subpixel transmits only one color; the other colors are absorbed. Figure 29 depicts the emission spectra of a backlight (cold cathode fluorescent lamp, CCFL) and RGB light emitting diodes (LEDs), and the transmission spectra of RGB color filters.

From Fig. 29, the transmission spectra of RGB color filters are relatively broad. The advantage is to transmit more light however its color purity is degraded. The peak transmission of the RGB color filters is ~ 70 , 80 , and 90 percent, respectively. Roughly speaking, each color filter only transmits ~ 25 percent of the incident white light. The rest ~ 75 percent is absorbed by the color pigments. Moreover, CCFL emits two unwanted lines: blue-green ($\sim 480 \text{ nm}$) and orange ($\sim 580 \text{ nm}$). The blue-green light will transmit through the blue and green color filters simultaneously. Similarly, the orange light will transmit through the green and red color filters simultaneously. These leaked lights will downgrade the color purity (or color saturation) of the display. Therefore, the color gamut of a typical transmissive TFT LCD is ~ 75 percent of NTSC (National Television System Committee) standard. With improved CCFL spectra, the color gamut can reach ~ 92 percent. LEDs have narrower emission spectra that also match

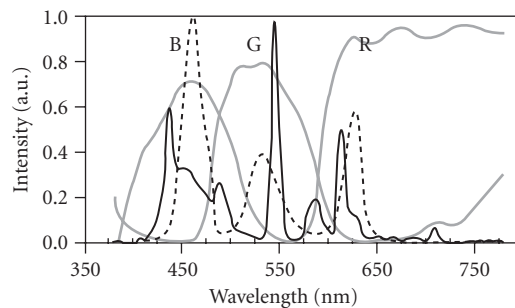


FIGURE 29 Transmission spectra of RGB color filters (thick gray lines), and emission spectra of CCFL backlight (thin black lines) and LEDs (dashed lines).

better with the transmission spectra of the color filters, thus the color gamut reaches ~ 120 percent.⁵⁹ For a display device, a wider color gamut is not necessarily better; natural colors are also important. After all, display is an art where perception plays an important role.

After taking into account the optical losses from polarizers, color filters, and TFT aperture ratio, the overall system optical efficiency is about 6 to 7 percent for a direct-view LCD. If wide view technology is included,⁶⁰ the total light efficiency is decreased to ~ 5 percent. Low optical efficiency implies high power consumption and more heat generation inside the display chassis. For a thin LCD, thermal dissipation is a critical issue. For portable displays, low power consumption is desirable because it lengthens the battery operating hours. Several approaches have been developed to reduce power consumption, for example, polarization conversion of backlight⁶¹ and two-dimensional LED backlight with local dimming capability.^{62–64} The use of LED backlight offers several additional advantages such as wide color gamut, high dynamic contrast ratio ($>50,000:1$), $\sim 2\times$ reduction in power consumption, and fast turn-on and off times (~ 10 ns) for reducing the motion picture image blurs.⁶⁵ Some technological concerns are color and power drifting as the junction temperature changes, and cost.

Reflective LCDs

Figure 30 shows a device structure of a TFT-based reflective LCD. The top linear polarizer and a broadband quarter-wave film forms an equivalent crossed polarizer for the incident and exit beams. This is because the LC modes work better under crossed-polarizer condition. The bumpy reflector not only reflects but also diffuses the ambient light to the observer in order to avoid specular reflection and widen the viewing angle. This is a critical part for reflective LCDs. The TFT is hidden beneath the bumpy reflector, thus the R-LCD can have a large aperture ratio ($\sim 90\%$). The light blocking layer (LBL) is used to absorb the scattered light from neighboring pixels. Two popular LCD modes have been widely used for R-LCDs: (1) VA cell, and (2) mixed-mode twisted nematic (MTN) cell. The VA cell utilizes the phase retardation effect while the MTN cell uses the combination of polarization rotation and birefringence effects.

In a reflective LCD, there is no built-in backlight unit; instead, it utilizes ambient light for reading out the displayed images. In comparison to transmissive LCDs, reflective LCDs have advantages in lower power consumption, lighter weight, and better sunlight readability. However, a reflective LCD is inapplicable under low or dark ambient conditions. Therefore, the TFT-based reflective LCD is gradually losing its ground.

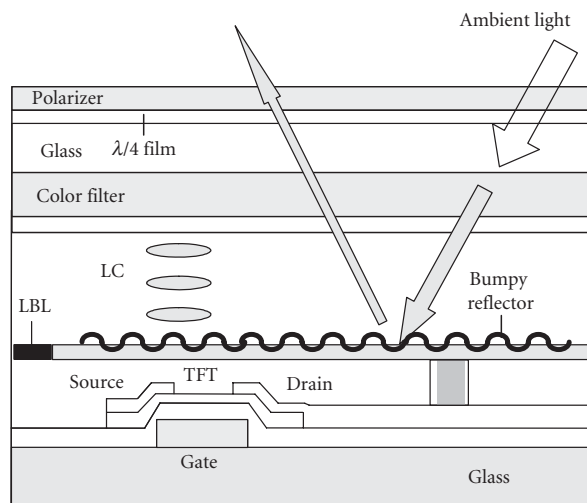


FIGURE 30 Device structure of a direct-view reflective LCD.

Flexible reflective LCDs using cholesteric liquid crystal display (Ch-LCD) and bistable nematic are gaining momentum because they can be used as electronic papers. These reflective LCDs use ambient light to readout the displayed images. Ch-LCD has helical structure which reflects color so that the display does not require color filters, neither polarizer. Thus, the reflectance for a given color band which depends on the pitch length and refractive index of the employed LC is relatively high (~30%). Moreover, it does not require a backlight so that its weight is light and the total device thickness can be thinner than 200 μm . Therefore, it is a strong contender for color flexible displays. Ch-LCD is a bistable device so that its power consumption is low, provided that the device is not refreshed too frequently. A major drawback of a reflective direct-view LCD is its poor readability under low ambient light.

Another reflective LCD developed for projection TVs is liquid-crystal-on-silicon (LCoS) microdisplay. Unlike a transmissive microdisplay, LCoS is a reflective device. Here the reflector employed is an aluminum metallic mirror. Crystalline silicon has high mobility so that the pixel size can be made small (<10 μm) and aperture ratio >90 percent. Therefore, the image not only has high resolution but also is seamless. By contrast, a transmissive microdisplay's aperture ratio is about 65 percent. The light blocked by the black matrices show up in the screen as dark patterns (also known as screen door effect). Viewing angle of a LCD is less critical in projection than direct-view displays because in a projection display the polarizing beam splitter has a narrower acceptance angle than the employed LCD.

Transflective LCDs

In a transflective liquid crystal display (TR-LCD), a pixel is divided into two parts: transmissive (T) and reflective (R). The T/R area ratio can vary from 80/20 to 20/80, depending on the applications. In dark ambient the backlight is on and the display works as a transmissive one, while at bright ambient the backlight is off and only the reflective mode is operational.

Dual-Cell-Gap Transflective LCDs In a TR-LCD, the backlight traverses the LC layer once, but the ambient light passes through twice. As a result, the optical path length is unequal. To balance the optical path difference between the T and R regions for a TR-LCD, dual-cell-gap device concept is introduced. The basic requirement for a TR-LCD is to find equal phase retardation between the T and R modes, which is

$$d_T(\Delta n)_T = 2d_R(\Delta n)_R \quad (22)$$

If T and R modes have the same effective birefringence, then the cell gap should be different. This is the so-called dual cell gap approach. On the other hand, if the cell gap is uniform (single cell gap approach), then we should find ways to make $(\Delta n)_T = 2(\Delta n)_R$. Let us discuss the dual cell gap approaches first.

Figure 31a shows the schematic device configuration of a dual-cell-gap TR-LCD. Each pixel is divided into a reflective region with cell gap d_R and a transmissive region with cell gap d_T . The LC employed could be homogeneous alignment (also known as ECB, electrically controlled birefringence) or vertical alignment, as long as it is a phase retardation type. To balance the phase retardation between the single and double pass of the T and R parts, we could set $d_T = 2d_R$. Moreover, to balance the color saturation due to single and double-pass discrepancy, we could use thinner or holed color filters in the R part. The top quarter wave plate is needed mainly for the reflective mode to obtain a high contrast ratio. Therefore, in the T region, the optic axis of the bottom quarter-wave plate should be aligned perpendicular to that of the top one so that their phase retardations are canceled.

A thin homogeneous cell is difficult to find a good common dark state for RGB wavelengths without a compensation film.⁵⁵ The compensation film can be designed into the top quarter-wave film shown in Fig. 31a to form a single film. Here, let us take a dual cell gap TR-LCD using VA (or MVA for wide-view) and MLC-6608 ($\Delta\epsilon = -4.2$, $\Delta n = 0.083$) as an example. We set $d_R = 2.25 \mu\text{m}$ in the R region and $d_T = 4.5 \mu\text{m}$ in the T region. Figure 31b depicts the voltage-dependent transmittance (VT) and reflectance (VR) curves at normal incidence. As expected, both VT and VR curves perfectly overlap with each other. Here $d_R\Delta n = 186.8 \text{ nm}$ and $d_T\Delta n = 373.5 \text{ nm}$ are intentionally designed to be larger than $\lambda/4$ (137.5 nm) and $\lambda/2$ (275 nm), respectively, in order to reduce the on-state voltage to $\sim 5 V_{\text{rms}}$.

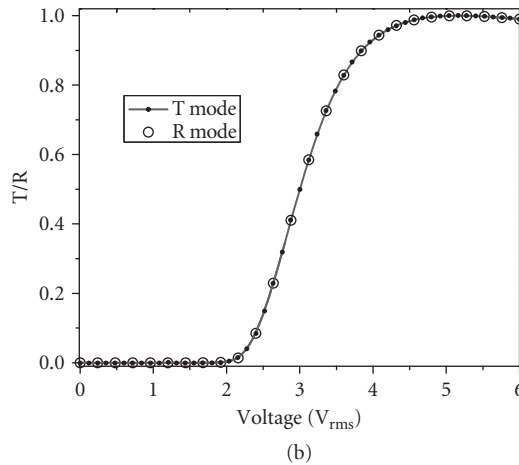
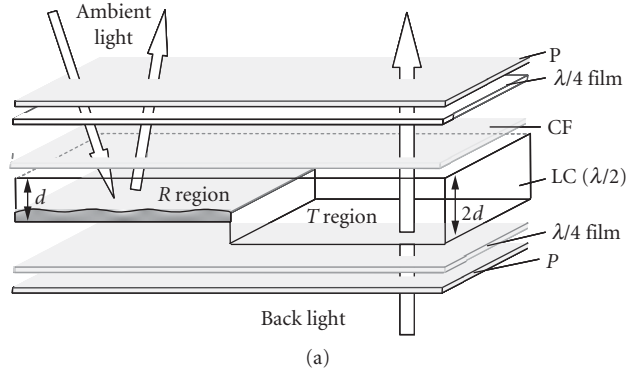


FIGURE 31 (a) Schematic device configuration of the dual-cell-gap TR-LCD. (b) Simulated VT and VR curves using VA (or MVA) cells. LC: MLC-6608, $dT = 4.5 \mu\text{m}$, $dR = 2.25 \mu\text{m}$, and $\lambda = 550 \text{ nm}$.

Three problems of dual-cell-gap TR-LCDs are encountered: (1) Due to the cell gap difference the LC alignment is distorted near the T and R boundaries. The distorted LCs will cause light scattering and degrade the device contrast ratio. Therefore, these regions should be covered by black matrices in order to retain a good contrast ratio. (2) The thicker cell gap in the T region results in a slower response time than the R region. Fortunately, the dynamic response requirement in mobile displays is not as strict as those for video applications. This response time difference, although not perfect, is still tolerable. (3) The view angle of the single-domain homogeneous cell mode is relatively narrow because the LC directors are tilted out of the plane by the longitudinal electric field. To improve view angle, a biaxial film⁶⁶ or a hybrid aligned nematic polymeric film⁶⁷ is needed. Because the manufacturing process is compatible with the LCD fabrication lines, the dual-cell-gap TR-LCD is widely used in commercial products, such as iPhones.

Single-Cell-Gap Transflective LCDs As its name implies, the single-cell-gap TR-LCD has a uniform cell gap in the T and R regions. Therefore, we need to find device concepts to achieve $(\Delta n)_T = 2(\Delta n)_R$. Several approaches have been proposed to solve this problem. In this section, we will discuss two

examples: (1) dual-TFT method in which one TFT is used to drive the T mode and another TFT to drive the R mode at a lower voltage,⁶⁸ and (2) divided-voltage method:⁶⁹ to have multiple R parts and the superimposed VR curve matches the VT curve.

Example 3: Dual-TFT Method Figure 32a shows the device structure of a TR-LCD using two TFTs to separately control the gamma curves of the T and R parts. Here, TFT-1 is connected to the bumpy reflector and TFT-2 is connected to the ITO of the transmissive part. Because of the double passes, the VR curve has a sharper slope than the VT curve and it reaches the peak reflectance at a lower voltage, as shown in Fig. 32b. Let us use a 4.5- μm vertically aligned LC layer with 88° pretilt angle as an example. The LC mixture employed is Merck MLC-6608 and wavelength is $\lambda = 550$ nm. From Fig. 32b, the peak reflectance occurs at $3 V_{\text{rms}}$ and transmittance at $5.5 V_{\text{rms}}$. Thus, the maximum voltage of TFT-1 should be set at 5.5 V and TFT-2 at 3 V. This driving scheme is also called double-gamma method.⁷⁰ The major advantage of this dual-TFT approach is its simplicity. However, each TFT takes up some real estate so that the aperture ratio for the T mode is reduced.

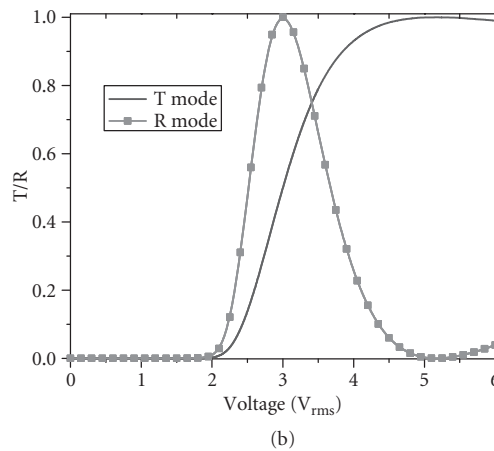
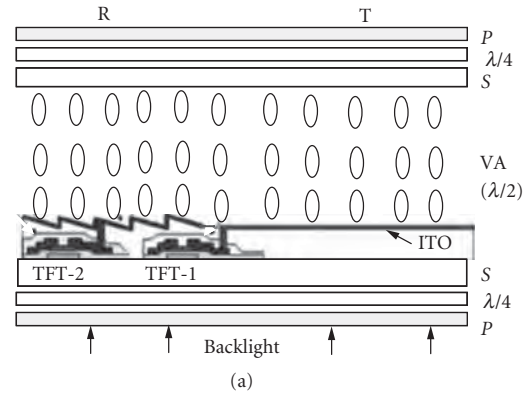


FIGURE 32 (a) Device structure of a dual-TFT TR-LCD and (b) simulated VT and VR curves using a 4.5- μm , MLC-6608 LC layer.

For a TR-LCD, the T mode should have priority over R mode. The major function of R mode is to preserve sunlight readability. In general, the viewing angle, color saturation, and contrast ratio of R mode are all inferior to T mode. In most lighting conditions except under direct sunlight, T mode is still the primary display.

Example 4: Divided-Voltage Method Figure 33a shows the device structure of a TR-LCD using divided voltage method.⁷¹ The R region consists of two sub-regions: R-I and R-II. Between R-II and bottom ITO, there is a passivation layer to weaken the electric field in the R-II region. As plotted in Fig. 33b, the VR-II curve in the R-II region has a higher threshold voltage than the VT curve due to this voltage shielding effect. To better match the VT curve, a small area in the T region is also used for R-I. The bumpy reflector in the R-I region is connected to the bottom ITO through a channeled electrode. Because of the double passes of ambient light, the VR-I curve is sharper than the VT curve. By properly choosing the R-I and R-II areas, we can match the VT and VR curves well, as shown in Fig. 33b.

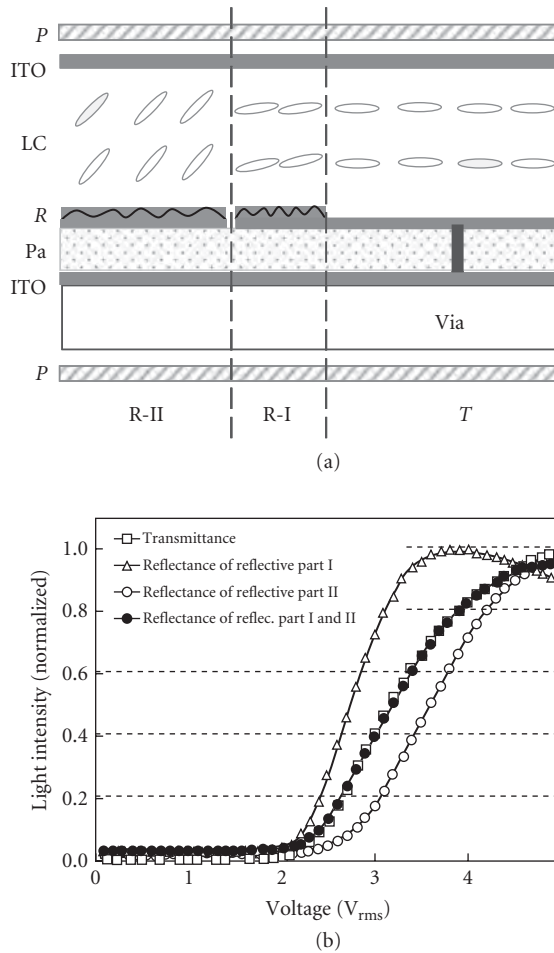


FIGURE 33 A transfective LCD using divided-voltage approach. (a) Device structure and (b) VT and VR curves at different regions. Here, P: polarizer, R: bumpy reflector, and Pa: passivation layer. (Redrawn from Ref. 71.)

8.8 POLYMER/LIQUID CRYSTAL COMPOSITES

Some types of liquid crystal display combine LC material and polymer material in a single device. Such LC/polymer composites are relatively new class of materials used for displays but also for light shutters or switchable windows. Typically LC/polymer components consist of calamitic low mass LCs and polymers and can be either polymer-dispersed liquid crystal (PDLC) or polymer-stabilized liquid crystals (PSLC).⁷²⁻⁷⁴ The basic difference between these two types comes from concentration ratio between LC and polymer. In case of PDLC there is typically around 1:1 percentage ratio of LC and polymer. In PSLC, the LC occupies 90 percent or more of the total composition. Such difference results in different phase separation process during composites polymerization. For equal concentration of LC and polymer, the LC droplets will form. But in case the LC is a majority, polymer will build up only walls or strings which divide LC into randomly aligned domains. Both types of composites operate between transparent state and scattering state. There are two requirements on the polymer for PDLC or PSLC device to work. First, refraction index of the polymer, n_p , must be equal to the refraction index for light polarized perpendicular to the director of the liquid crystal, (ordinary refractive index of the LC). Second, the polymer must induce the director of the LC in the droplets (PDLC) or domains (PSLC) to orient parallel to the surface of the polymer (Fig. 34). In the voltage OFF state the LC molecules in the droplets are partially aligned. In addition, the average director orientation $\mathbf{n}$ of the droplets exhibits a random distribution of orientation within the cell.

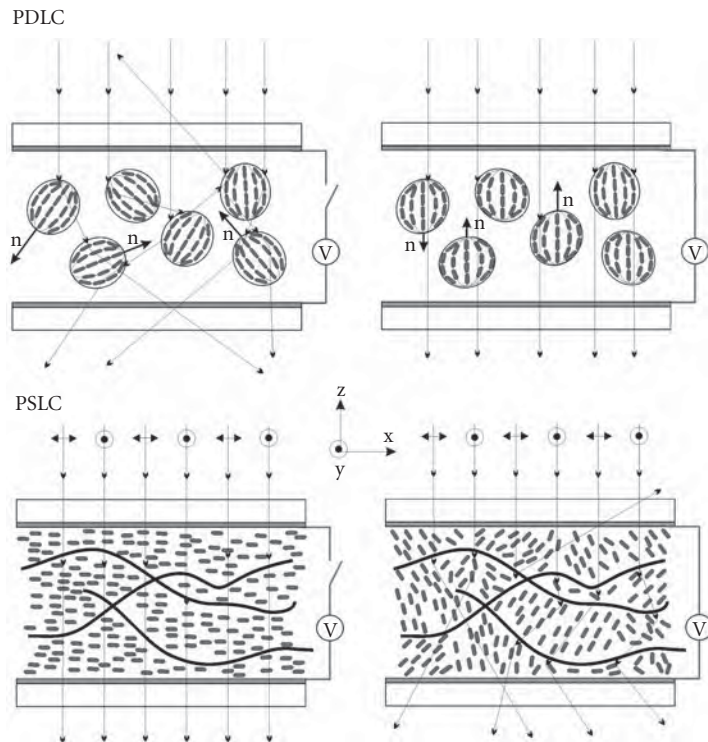


FIGURE 34 Schematic view and working principles of polymer/LC composites.

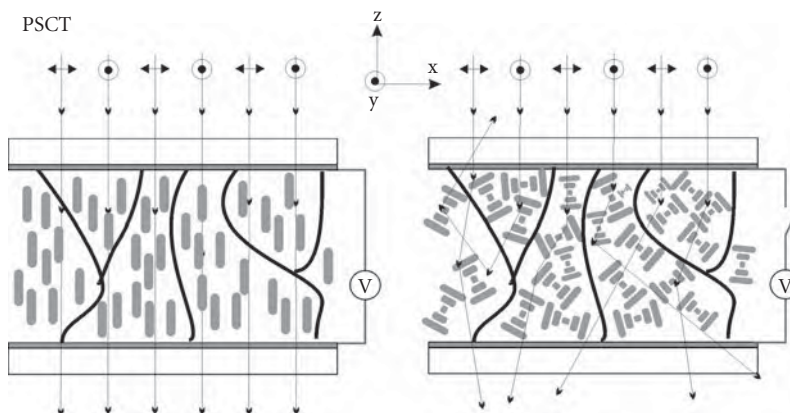


FIGURE 35 Schematic view of working principles of polymer-stabilized cholesteric LC light valve.

The incident unpolarized light is scattered if it goes through such a sample. When a sufficiently strong electric field (typically above $1 V_{\text{rms}}/\mu\text{m}$) is applied to the cell, all the LC molecules align parallel to the electric field. If the light is also propagating in the direction parallel to the field, then the beam of light is affected by ordinary refractive index of LC which is matched with refractive index of polymer, thus cell appears transparent. When the electric field is OFF, again the LC molecules go back to the previous random positions. A polymer mixed with a chiral liquid crystal is a special case of PS LC called as polymer stabilized cholesteric texture (PSCT). The ratio between polymer and liquid crystal remains similar to the one necessary for PS LC. When the voltage is not applied to the PSCT cell liquid crystals tends to have helical structure while the polymer network tends to keep LC director parallel to it (normal-mode PSCT). Therefore, the material has a poly-domain structure, as Fig. 35 shows. In this state the incident beam is scattered. When a sufficiently high electric field is applied across the cell, the liquid crystal is switched to the homeotropic alignment and, as a result, the cell becomes transparent.

8.9 SUMMARY

Liquid crystal was discovered more than 100 years ago and is finding widespread applications. This class of organic material exhibits some unique properties, such as good chemical and thermal stabilities, low operation voltage and low power consumption, and excellent compatibility with semiconductor fabrication processing. Therefore, it has dominated direct-view and projection display markets. The forecasted annual TFT LCD market is going to exceed \$100 billion by year 2011. However, there are still some technical challenges need to be overcome, for example, (1) faster response time for reducing motion picture blurs, (2) higher optical efficiency for reducing power consumption and lengthening battery life, (3) smaller color shift when viewed at oblique angles, (4) wider viewing angle with higher contrast ratio (ideally its viewing characteristics should be as good as an emissive display), and (5) lower manufacturing cost, especially for large screen TVs. In addition to displays, LC materials also share an important part of emerging photonics applications, such as spatial light modulators for laser beam steering and adaptive optics, adaptive-focus lens, variable optical attenuator for fiber-optic telecommunications, and LC-infiltrated photonic crystal fibers, just to name a few. Often neglected, lyotropic liquid crystals are important materials in biochemistry of the cell membranes.

8.10 REFERENCES

1. F. Reinitzer, *Monatsh. Chem.*, **9**:421 (1888); for English translation see, *Liq. Cryst.* **5**:7 (1989).
2. P. F. McManamon, T. A. Dorschner, D. L. Corkum, L. Friedman, D. S. Hobbs, M. Holz, S. Liberman, et al., *Proc. IEEE* **84**:268 (1996).
3. V. Vill, *Database of Liquid Crystalline Compounds for Personal Computers*, Ver. 4.6 (LCI Publishers, Hamburg 2005).
4. C. S. O'Hern and T. C. Lubensky, *Phys. Rev. Lett.* **80**:4345 (1998).
5. G. Friedel, *Ann. Physique* **18**:173 (1922).
6. P. G. De Gennes and J. Prost, *The Physics of Liquid Crystals*, 2nd ed. (Clarendon, Oxford, 1993).
7. P. R. Gerber, *Mol. Cryst. Liq. Cryst.* **116**:197 (1985).
8. N. A. Clark and S. T. Lagerwall, *Appl. Phys. Lett.* **36**:889 (1980).
9. A. D. L. Chandani, T. Hagiwara, Y. Suzuki, Y. Ouchi, H. Takezoe, and A. Fukuda, *Jpn. J. Appl. Phys.* **27**:L1265 (1988).
10. A. D. L. Chandani, E. Górecka, Y. Ouchi, H. Takezoe, and A. Fukuda, *Jpn. J. Appl. Phys.* **27**:L729 (1989).
11. J. W. Goodby, M. A. Waugh, S. M. Stein, E. Chin, R. Pindak, and J. S. Patel, *Nature* **337**:449 (1989).
12. T. C. Lubensky and S. R. Renn, *Mol. Cryst. Liq. Cryst.* **209**:349 (1991).
13. L. Schröder, *Z. Phys. Chem.* **11**:449 (1893).
14. J. J. Van Laar, *Z. Phys. Chem.* **63**:216 (1908).
15. W. Maier, G. Meier, and Z. Naturforsch. *Teil. A* **16**:262 (1961).
16. M. Schadt, *Displays* **13**:11 (1992).
17. G. Gray, K. J. Harrison, and J. A. Nash, *Electron. Lett.* **9**:130 (1973).
18. R. Dabrowski, *Mol. Cryst. Liq. Cryst.* **191**:17 (1990).
19. M. Schadt and W. Helfrich, *Appl. Phys. Lett.* **18**:127 (1971).
20. R. A. Soref, *Appl. Phys. Lett.* **22**:165 (1973).
21. M. Oh-e and K. Kondo, *Appl. Phys. Lett.* **67**:3895 (1995).
22. Y. Nakazono, H. Ichinose, A. Sawada, S. Naemura, and K. Tarumi, *Int'l Display Research Conference*, p. 65 (1997).
23. R. Tarao, H. Saito, S. Sawada, and Y. Goto, *SID Tech. Digest* **25**:233 (1994).
24. T. Geelhaar, K. Tarumi, and H. Hirschmann, *SID Tech. Digest* **27**:167 (1996).
25. Y. Goto, T. Ogawa, S. Sawada and S. Sugimori, *Mol. Cryst. Liq. Cryst.* **209**:1 (1991).
25. M. F. Schiekel and K. Fahrenschon, *Appl. Phys. Lett.* **19**:391 (1971).
26. Q. Hong, T. X. Wu, X. Zhu, R. Lu, and S. T. Wu, *Appl. Phys. Lett.* **86**:121107 (2005).
27. C. H. Wen, S. Gauza, and S. T. Wu, *Appl. Phys. Lett.* **87**:191909 (2005).
28. R. Lu, Q. Hong, and S. T. Wu, *J. Display Technology* **2**:217 (2006).
29. R. Eidenschink and L. Pohl, *US Patent* 4, 415, 470 (1983).
30. W. H. de Jeu, "The Dielectric Permittivity of Liquid Crystals" *Solid State Phys. Suppl.* **14**: "Liquid Crystals" Edited by L. Liebert. (Academic Press, New York, 1978); also, *Mol. Cryst. Liq. Cryst.* **63**:83 (1981).
31. M. Schadt, *Mol. Cryst. Liq. Cryst.* **89**:77 (1982).
32. H. K. Bucher, R. T. Klingbiel, and J. P. VanMeter, *Appl. Phys. Lett.* **25**:186 (1974).
33. H. Xianyu, Y. Zhao, S. Gauza, X. Liang, and S. T. Wu, *Liq. Cryst.* **35**:1129 (2008).
34. T. K. Bose, B. Campbell, and S. Yagihara, *Phys. Rev. A* **36**:5767 (1987).
35. C. H. Wen and S. T. Wu, *Appl. Phys. Lett.* **86**:231104 (2005).
36. I. C. Khoo and S. T. Wu, *Optics and Nonlinear Optics of Liquid Crystals* (World Scientific, Singapore, 1993).
37. S. T. Wu, E. Ramos, and U. Finkenzeller, *J. Appl. Phys.* **68**:78 (1990).
38. S. T. Wu, U. Efron, and L. D. Hess, *Appl. Opt.* **23**:3911 (1984).
39. S. T. Wu, *J. Appl. Phys.* **69**:2080 (1991).
40. S. T. Wu, C. S. Wu, M. Warenghem, and M. Ismaili, *Opt. Eng.* **32**:1775 (1993).

41. J. Li and S. T. Wu, *J. Appl. Phys.* **95**:896 (2004).
42. S. T. Wu, U. Efron and L. D. Hess, *Appl. Phys. Lett.* **44**:1033 (1984).
43. J. Li and S. T. Wu, *J. Appl. Phys.* **96**:170 (2004).
44. H. Mada and S. Kobayashi, *Mol. Cryst. Liq. Cryst.* **33**:47 (1976).
45. E. H. Stupp and M. S. Brennessoltz, *Projection Displays* (Wiley, New York, 1998).
46. J. Li, S. Gauza, and S. T. Wu, *Opt. Express* **12**:2002 (2004).
47. J. Li and S. T. Wu, *J. Appl. Phys.* **96**:19 (2004).
48. W. Maier and A. Saupe, *Z. Naturforsch. Teil A* **15**:287 (1960).
49. H. Gruler, and Z. Naturforsch. *Teil A* **30**:230 (1975).
50. W. M. Gelbart and A. Ben-Shaul, *J. Chem. Phys.* **77**:916 (1982).
51. S. T. Wu and C. S. Wu, *Liq. Cryst.* **8**:171 (1990). Seven commonly used models (see the references therein) have been compared in this paper.
52. M. A. Osipov and E. M. Terentjev, and Z. Naturforsch. *Teil A* **44**:785 (1989).
53. S. T. Wu and C. S. Wu, *Phys. Rev. A* **42**:2219 (1990).
54. S. T. Wu and C. S. Wu, *J. Appl. Phys.* **83**:4096 (1998).
55. M. Oh-e and K. Kondo, *Appl. Phys. Lett.* **67**:3895 (1995).
56. M.F. Schiekkel and K. Fahrenschoen, *Appl. Phys. Lett.* **19**:391 (1971).
57. K. Hanaoka, Y. Nakanishi, Y. Inoue, S. Tanuma, Y. Koike, and K. Okamoto, *SID Tech. Digest* **35**:1200 (2004).
58. S. G. Kim, S. M. Kim, Y. S. Kim, H. K. Lee, S. H. Lee, G. D. Lee, J. J. Lyu, and K. H. Kim, *Appl. Phys. Lett.* **90**:261910 (2007).
59. R. Lu, Q. Hong, Z. Ge, and S. T. Wu, *Opt. Express* **14**:6243 (2006).
60. D. K. Yang and S. T. Wu, *Fundamentals of Liquid Crystal Devices* (Wiley, New York, 2006).
61. J. M. Jonza, M. F. Weber, A. J. Ouderkirk, and C. A. Stover, *U.S. Patent* 5,962,114 (1999).
62. P. de Greef and H. G. Hulze, *SID Symp. Digest* **38**:1332 (2007).
63. H. Chen, J. Sung, T. Ha, and Y. Park, *SID Symp. Digest* **38**:1339 (2007).
64. F. C. Lin, C. Y. Liao, L. Y. Liao, Y. P. Huang, and H. P. Shieh, *SID Symp. Digest* **38**:1343 (2007).
65. M. Anandan, *J. SID* **16**:287 (2008).
66. M. Shibazaki, Y. Ukawa, S. Takahashi, Y. Iefuji, and T. Nakagawa, *SID Tech. Digest* **34**:90 (2003).
67. T. Uesaka, S. Ikeda, S. Nishimura, and H. Mazaki, *SID Tech. Digest* **28**:1555 (2007).
68. K. H. Liu, C. Y. Cheng, Y. R. Shen, C. M. Lai, C. R. Sheu, and Y. Y. Fan, *Proc. Int. Display Manuf. Conf.* p. 215 (2003).
69. C. Y. Tsai, M. J. Su, C. H. Lin, S. C. Hsu, C. Y. Chen, Y. R. Chen, Y. L. Tsai, C. M. Chen, C. M. Chang, and A. Lien, *Proc. Asia Display* **24** (2007).
70. C. R. Sheul, K. H. Liu, L. P. Hsin, Y. Y. Fan, I. J. Lin, C. C. Chen, B. C. Chang, C. Y. Chen, and Y. R. Shen, *SID Tech. Digest* **34**:653 (2003).
71. Y. C. Yang, J. Y. Choi, J. Kim, M. Han, J. Chang, J. Bae, D. J. Park, et al., *SID Tech. Digest* **37**:829 (2006).
72. J. L. Ferguson, *SID Symp. Digest* **16**:68 (1985).
73. J. W. Doane, N. A. Vaz, B. G. Wu, and S. Zumer, *Appl. Phys. Lett.* **48**:269 (1986).
74. R. L. Sutherland, V. P. Tondiglia, and L. V. Natarajan, *Appl. Phys. Lett.* **64**:1074 (1994).

8.11 BIBLIOGRAPHY

1. Chandrasekhar, S., *Liquid Crystals*, 2nd edn. (Cambridge University Press, Cambridge, England, 1992).
2. Collings, P. J., *Nature's Delicate Phase of Matter*, 2nd ed. (Princeton University Press, Princeton, N.J., 2001).
3. Collings, P. J., and M. Hird, *Introduction to Liquid Crystals Chemistry and Physics*, (Taylor & Francis, London 1997).
4. de Gennes, P. G., and J. Prost, *The Physics of Liquid Crystals*, 2nd ed. (Oxford University Press, Oxford, 1995).

5. de Jeu, W. H., *Physical Properties of Liquid Crystalline Materials* (Gorden and Breach, New York, 1980).
6. Demus, D., J. Goodby, G. W. Gray, H.-W. Spiess, and V. Vill, *Handbook of Liquid Crystals Vol. 1–4* (Wiley-VCH, Weinheim, New York, 1998).
7. Khoo, I. C., and S. T. Wu, *Optics and Nonlinear Optics of Liquid Crystals* (World Scientific, Singapore, 1993).
8. Kumar, S., *Liquid Crystals* (Cambridge University Press, Cambridge, England, 2001).
9. Oswald, P. and P. Pieranski, *Nematic and Cholesteric Liquid Crystals: Concepts and Physical Properties Illustrated by Experiments* (Taylor & Francis CRC Press, Boca Raton, FL, 2005).
10. Wu, S. T., and D.K. Yang, *Reflective Liquid Crystal Displays* (Wiley, New York, 2001).
11. Wu, S. T., and D. K. Yang, *Fundamentals of Liquid Crystal Devices* (Wiley, Chichester, England, 2006).
12. Yeh, P., and C. Gu, *Optics of Liquid Crystals* (Wiley, New York, 1999).

PART

4

FIBER OPTICS

This page intentionally left blank

OPTICAL FIBER COMMUNICATION TECHNOLOGY AND SYSTEM OVERVIEW

Ira Jacobs

*The Bradley Department of Electrical and Computer Engineering
Virginia Polytechnic Institute and State University
Blacksburg, Virginia*

9.1 INTRODUCTION

Basic elements of an optical fiber communication system include the transmitter [laser or light-emitting diode (LED)], fiber (multimode, single-mode, or dispersion-shifted), and the receiver [positive-intrinsic-negative (PIN) diode, avalanche photodiode (APD) detectors, coherent detectors, optical preamplifiers, receiver electronics]. Receiver sensitivities of digital systems are compared on the basis of the number of photons per bit required to achieve a given bit error probability, and eye degradation and error floor phenomena are described. Laser relative intensity noise and nonlinearities are shown to limit the performance of analog systems. Networking applications of optical amplifiers and wavelength-division multiplexing are considered, and future directions are discussed.

Although the light-guiding property of optical fibers has been known and used for many years, it is only relatively recently that optical fiber communications has become both a possibility and a reality.¹ Following the first prediction in 1966² that fibers might have sufficiently low attenuation for telecommunications, the first low-loss fiber (20 dB/km) was achieved in 1970.³ The first semiconductor laser diode to radiate continuously at room temperature was also achieved in 1970.⁴ The 1970s were a period of intense technology and system development, with the first systems coming into service at the end of the decade. The 1980s saw both the growth of applications (service on the first transatlantic cable in 1988) and continued advances in technology. This evolution continued in the 1990s with the advent of optical amplifiers and with the applications emphasis turning from point-to-point links to optical networks. The beginning of the 21st century has seen extensive fiber-to-the-home deployment as well as continued technology advances.

This chapter provides an overview of the basic technology, systems, and applications of optical fiber communication. It is an update and compression of material presented at a 1994 North Atlantic Treaty Organization (NATO) Summer School.⁵ Although there have been significant advances in technology and applications in subsequent years, the basics have remained essentially the same.

9.2 BASIC TECHNOLOGY

This section considers the basic technology components of an optical fiber communications link, namely the fiber, the transmitter, and the receiver, and discusses the principal parameters that determine communications performance.

Fiber

An optical fiber is a thin filament of glass with a central core having a slightly higher index of refraction than the surrounding cladding. From a physical optics standpoint, light is guided by total internal reflection at the core-cladding boundary. More precisely, the fiber is a dielectric waveguide in which there are a discrete number of propagating modes.⁶ If the core diameter and the index difference are sufficiently small, only a single mode will propagate. The condition for single-mode propagation is that the normalized frequency V be less than 2.405, where

$$V = \frac{2\pi a}{\lambda} \sqrt{n_1^2 - n_2^2} \quad (1)$$

and a is the core radius, λ is the free space wavelength, and n_1 and n_2 are the indexes of refraction of the core and cladding, respectively. Multimode fibers typically have a fractional index difference (Δ) between core and cladding of between 1 and 1.5 percent and a core diameter of between 50 and 100 μm . Single-mode fibers typically have $\Delta \approx 0.3$ percent and a core diameter of between 8 and 10 μm .

The fiber numerical aperture (NA), which is the sine of the half-angle of the cone of acceptance, is given by

$$\text{NA} = \sqrt{n_1^2 - n_2^2} = n_1 \sqrt{2\Delta} \quad (2)$$

Single-mode fibers typically have an NA of about 0.1, whereas the NA of multimode fibers is in the range of 0.2 to 0.3.

From a transmission system standpoint, the two most important fiber parameters are attenuation and bandwidth.

Attenuation There are three principal attenuation mechanisms in fiber: absorption, scattering, and radiative loss. Silicon dioxide has resonance absorption peaks in the ultraviolet (electronic transitions) and in the infrared beyond 1.6 μm (atomic vibrational transitions), but is highly transparent in the visible and near-infrared.

Radiative losses are generally kept small by using a sufficiently thick cladding (communication fibers have an outer diameter of 125 μm), a compressible coating to buffer the fiber from external forces, and a cable structure that prevents sharp bends.

In the absence of impurities and radiation losses, the fundamental attenuation mechanism is Rayleigh scattering from the irregular glass structure, which results in index of refraction fluctuations over distances that are small compared to the wavelength. This leads to a scattering loss

$$\alpha = \frac{B}{\lambda^4} \quad \text{with } B \approx 0.9 \frac{\text{dB}}{\text{km}} \mu\text{m}^4 \quad (3)$$

for “best” fibers. Attenuation as a function of wavelength is shown in Fig. 1. The attenuation peak at $\lambda = 1.4 \mu\text{m}$ is a resonance absorption due to small amounts of water in the fiber, although fibers are available in which this peak is absent. Initial systems operated at a wavelength around 0.85 μm owing to the availability of sources and detectors at this wavelength. Present systems (other than some short-distance data links) generally operate at wavelengths of 1.3 or 1.55 μm . The former, in addition to being low in attenuation (about 0.32 dB/km for best fibers), is the wavelength of minimum intramodal dispersion (see next section) for standard single-mode fiber. Operation at 1.55 μm allows even lower attenuation (minimum is about 0.16 dB/km) and the use of erbium-doped-fiber amplifiers (see Sec. 9.5), which operate at this wavelength.

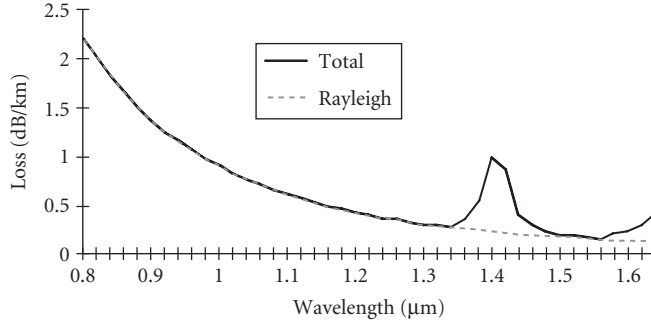


FIGURE 1 Fiber attenuation as a function of wavelength. Dashed curve shows Rayleigh scattering. Solid curve indicates total attenuation including resonance absorption at $1.38 \mu\text{m}$ from water and tail of infrared atomic resonances above $1.6 \mu\text{m}$.

Dispersion Pulse spreading (dispersion) limits the maximum modulation bandwidth (or maximum pulse rate) that may be used with fibers. There are two principal forms of dispersion: intermodal dispersion and intramodal dispersion. In multimode fiber, the different modes experience different propagation delays resulting in pulse spreading. For graded-index fiber, the lowest dispersion per unit length is given approximately by⁷

$$\frac{\delta\tau}{L} = \frac{n_1\Delta^2}{10c} \quad (\text{intermodal}) \quad (4)$$

[Grading of the index of refraction of the core in a nearly parabolic function results in an approximate equalization of the propagation delays. For a step-index fiber, the dispersion per unit length is $\delta\tau/L = n_1\Delta/c$, which for $\Delta = 0.01$ is 1000 times larger than that given by Eq. (4).]

Bandwidth is inversely proportional to dispersion, with the proportionality constant dependent on pulse shape and how bandwidth is defined. If the dispersed pulse is approximated by a Gaussian pulse with $\delta\tau$ being the full width at the half-power point, then the -3-dB bandwidth B is given by

$$B = 0.44/\delta\tau \quad (5)$$

Multimode fibers are generally specified by their bandwidth in a 1-km length. Typical specifications are in the range from 200 MHz to 1 GHz. Fiber bandwidth is a sensitive function of the index profile and is wavelength dependent, and the scaling with length depends on whether there is mode mixing.⁸ Also, for short-distance links, the bandwidth is dependent on the launch conditions. Multimode fibers are generally used only when the bit rates and distances are sufficiently small that accurate characterization of dispersion is not of concern, although this may be changing with the advent of graded-index plastic optical fiber for high-bit-rate short-distance data links.

Although there is no intermodal dispersion in single-mode fibers,* there is still dispersion within the single mode (intramodal dispersion) resulting from the finite spectral width of the source and the dependence of group velocity on wavelength. The intramodal dispersion per unit length is given by

$$\begin{aligned} \delta\tau/L &= D \delta\lambda & \text{for } D \neq 0 \\ &= 0.2S_o(\delta\lambda)^2 & \text{for } D = 0 \end{aligned} \quad (6)$$

*A single-mode fiber actually has two degenerate modes corresponding to the two principal polarizations. Any asymmetry in the transmission path removes this degeneracy and results in polarization dispersion. This is typically very small (in the range of 0.1 to 1 ps/km^{1/2}), but is of concern in long-distance systems using linear repeaters.⁹

where D is the dispersion coefficient of the fiber, $\delta\lambda$ is the spectral width of the source, and S_o is the dispersion slope

$$S_o = \frac{dD}{d\lambda} \text{ at } \lambda = \lambda_0 \quad \text{where} \quad D(\lambda_0) = 0 \quad (7)$$

If both intermodal and intramodal dispersion are present, the square of the total dispersion is the sum of the squares of the intermodal and intramodal dispersions. For typical digital systems, the total dispersion should be less than half the interpulse period T . From Eq. (5) this corresponds to an effective fiber bandwidth that is at least $0.88/T$.

There are two sources of intramodal dispersion: material dispersion, which is a consequence of the index of refraction being a function of wavelength, and waveguide dispersion, which is a consequence of the propagation constant of the fiber waveguide being a function of wavelength.

For a material with index of refraction $n(\lambda)$, the material dispersion coefficient is given by

$$D_{\text{mat}} = -\frac{\lambda}{c} \frac{d^2 n}{d\lambda^2} \quad (8)$$

For silica-based glasses, D_{mat} has the general characteristics shown in Fig. 2. It is about $-100 \text{ ps/km} \cdot \text{nm}$ at a wavelength of 820 nm, goes through zero at a wavelength near 1300 nm, and is about $20 \text{ ps/km} \cdot \text{nm}$ at 1550 nm.

For step-index single-mode fibers, waveguide dispersion is given approximately by¹⁰

$$D_{\text{wg}} \approx -\frac{0.025\lambda}{a^2 c n_2} \quad (9)$$

For conventional single-mode fiber, waveguide dispersion is small (about $-5 \text{ ps/km} \cdot \text{nm}$ at 1300 nm). The resultant $D(\lambda)$ is then slightly shifted (relative to the material dispersion curve) to longer wavelengths, but the zero-dispersion wavelength (λ_0) remains in the vicinity of 1300 nm. However, if the waveguide dispersion is made larger negative by decreasing a or equivalently by tapering the index of refraction in the core the zero-dispersion wavelength may be shifted to the vicinity of 1550 nm

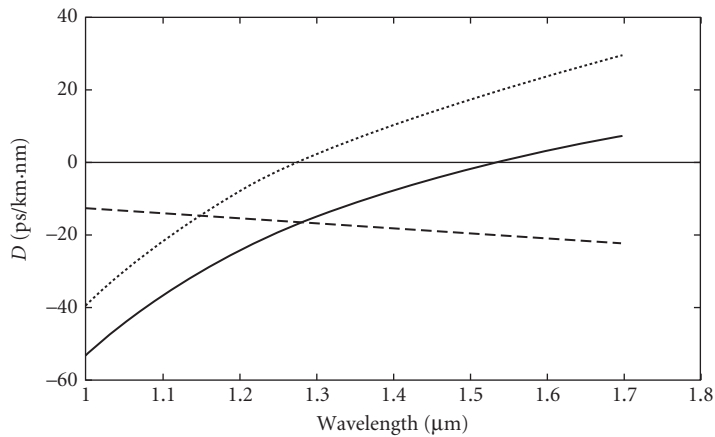


FIGURE 2 Intramodal dispersion coefficient as a function of wavelength. Dotted curve shows D_{mat} ; dashed curve shows D_{wg} to achieve D (solid curve) with zero dispersion at 1.55 μm .

(see Fig. 2). Such fibers are called *dispersion-shifted fibers* and are advantageous because of the lower fiber attenuation at this wavelength and the advent of erbium-doped-fiber amplifiers (see Sec. 9.5). Note that dispersion-shifted fibers have a smaller slope at the dispersion minimum ($S_0 \approx 0.06$ ps/km · nm² compared to $S_0 \approx 0.09$ ps/km · nm² for conventional single-mode fiber).

With more complicated index of refraction profiles, it is possible, at least theoretically, to control the shape of the waveguide dispersion such that the total dispersion is small in both the 1300- and 1550-nm bands, leading to dispersion-flattened fibers.¹¹

Transmitting Sources

Semiconductor light-emitting diodes (LEDs) or lasers are the primary light sources used in fiber-optic transmission systems. The principal parameters of concern are the power coupled into the fiber, the modulation bandwidth, and (because of intramodal dispersion) the spectral width.

Light-Emitting Diodes (LEDs) LEDs are forward-biased positive-negative (PN) junctions in which carrier recombination results in spontaneous emission at a wavelength corresponding to the energy gap. Although several milliwatts may be radiated from high-radiance LEDs, the radiation is over a wide angular range, and consequently there is a large coupling loss from an LED to a fiber. Coupling efficiency (η = ratio of power coupled to power radiated) from an LED to a fiber is given approximately by¹²

$$\begin{aligned} \eta &\approx (\text{NA})^2 & \text{for } r_s < a \\ \eta &\approx (a/r_s)^2 (\text{NA})^2 & \text{for } r_s > a \end{aligned} \quad (10)$$

where r_s is the radius of the LED. Use of large-diameter, high-NA multimode fiber improves the coupling from LEDs to fiber. Typical coupling losses are 10 to 20 dB for multimode fibers and more than 30 dB for single-mode fibers.

In addition to radiating over a large angle, LED radiation has a large spectral width (about 50 nm at $\lambda = 850$ nm and 100 nm at $\lambda = 1300$ nm) determined by thermal effects. Systems employing LEDs at 850 nm tend to be intramodal-dispersion-limited, whereas those at 1300 nm are intermodal-dispersion-limited.

Owing to the relatively long time constant for spontaneous emission (typically several nanoseconds), the modulation bandwidths of LEDs are generally limited to several hundred MHz. Thus, LEDs are generally limited to relatively short-distance, low-bit-rate applications.

Lasers In a laser, population inversion between the ground and excited states results in stimulated emission. In edge-emitting semiconductor lasers, this radiation is guided within the active region of the laser and is reflected at the end faces.* The combination of feedback and gain results in oscillation when the gain exceeds a threshold value. The spectral range over which the gain exceeds threshold (typically a few nanometers) is much narrower than the spectral width of an LED. Discrete wavelengths within this range, for which the optical length of the laser is an integer number of half-wavelengths, are radiated. Such a laser is termed a *multilongitudinal mode Fabry-Perot laser*. Radiation is confined to a much narrower angular range than for an LED, and consequently may be efficiently coupled into a small-NA fiber. Coupled power is typically about 1 mW.

The modulation bandwidth of lasers is determined by a resonance frequency caused by the interaction of the photon and electron concentrations.¹⁴ Although this resonance frequency was less than 1 GHz in early semiconductor lasers, improvements in materials have led to semiconductor lasers with resonance frequencies (and consequently modulation bandwidths) in excess of 10 GHz. This not only is important for very high-speed digital systems, but now also allows semiconductor lasers to be directly modulated with microwave signals. Such applications are considered in Sec. 9.7.

*In vertical cavity surface-emitting lasers (VCSELs), reflection is from internal “mirrors” grown within the semiconductor structure.¹³

Although multilongitudinal-mode Fabry-Perot lasers have a narrower spectral spread than LEDs, this spread still limits the high-speed and long-distance capability of such lasers. For such applications, single-longitudinal-mode (SLM) lasers are used. SLM lasers may be achieved by having a sufficiently short laser (less than 50 μm), by using coupled cavities (either external mirrors or cleaved coupled cavities¹⁵), or by incorporating a diffraction grating within the laser structure to select a specific wavelength. The latter has proven to be most practical for commercial application, and includes the distributed feedback (DFB) laser, in which the grating is within the laser active region, and the distributed Bragg reflector (DBR) laser, where the grating is external to the active region.¹⁶

There is still a finite line width for SLM lasers. For lasers without special stabilization, the line width is on the order of 0.1 nm. Expressed in terms of frequency, this corresponds to a frequency width of 12.5 GHz at a wavelength of 1550 nm. (Wavelength and frequency spread are related by $\delta f/f = -\delta\lambda/\lambda$, from which it follows that $\delta f = -c\delta\lambda/\lambda^2$.) Thus, unlike electrical communication systems, optical systems generally use sources with spectral widths that are large compared to the modulation bandwidth.

The finite line width (phase noise) of a laser is due to fluctuations of the phase of the optical field resulting from spontaneous emission. In addition to the phase noise contributed directly by the spontaneous emission, the interaction between the photon and electron concentrations in semiconductor lasers leads to a conversion of amplitude fluctuations to phase fluctuations, which increases the line width.¹⁷ If the intensity of a laser is changed, this same phenomenon gives rise to a change in the frequency of the laser (*chirp*). Uncontrolled, this causes a substantial increase in line width when the laser is modulated, which may cause difficulties in some system applications, possibly necessitating external modulation. However, the phenomenon can also be used to advantage. For appropriate lasers under small signal modulation, a change in frequency proportional to the input signal can be used to frequency-modulate and/or to tune the laser. Tunable lasers are of particular importance in networking applications employing wavelength-division multiplexing (WDM).¹⁸

Photodetectors

Fiber-optic systems generally use PIN or APD photodetectors. In a reverse-biased PIN diode, absorption of light in the intrinsic region generates carriers that are swept out by the reverse-bias field. This results in a photocurrent (I_p) that is proportional to the incident optical power (P_R), where the proportionality constant is the responsivity ($\mathfrak{R}$) of the photodetector; that is, $\mathfrak{R} = I_p/P_R$. Since the number of photons per second incident on the detector is power divided by the photon energy, and the number of electrons per second flowing in the external circuit is the photocurrent divided by the charge of the electron, it follows that the quantum efficiency (η = electrons/photons) is related to the responsivity by

$$\eta = \frac{hc}{q\lambda} \frac{I_p}{P_R} = \frac{1.24(\mu\text{m} \cdot \text{V})}{\lambda} \mathfrak{R} \quad (11)$$

For wavelengths shorter than 900 nm, silicon is an excellent photodetector, with quantum efficiencies of about 90 percent. For longer wavelengths, InGaAs is generally used, with quantum efficiencies typically around 70 percent. Very high bandwidths may be achieved with PIN photodetectors. Consequently, the photodetector does not generally limit the overall system bandwidth.

In an avalanche photodetector (APD), a larger reverse voltage accelerates carriers, causing additional carriers by impact ionization resulting in a current $I_{\text{APD}} = MI_p$, where M is the current gain of the APD. As we will note in Sec. 9.3, this can result in an improvement in receiver sensitivity.

9.3 RECEIVER SENSITIVITY

The receiver in a direct-detection fiber-optic communication system consists of a photodetector followed by electrical amplification and signal-processing circuits intended to recover the communications signal. Receiver sensitivity is defined as the average received optical power needed to

achieve a given communication rate and performance. For analog communications, the communication rate is measured by the bandwidth of the electrical signal to be transmitted (B), and performance is given by the signal-to-noise ratio (SNR) of the recovered signal. For digital systems, the communication rate is measured by the bit rate (R_b) and performance is measured by the bit error probability (P_e).

For a constant optical power transmitted, there are fluctuations of the received photocurrent about the average given by Eq. (11). The principal sources of these fluctuations are signal shot noise (quantum noise resulting from random arrival times of photons at the detector), receiver thermal noise, APD excess noise, and relative intensity noise (RIN) associated with fluctuations in intensity of the source and/or multiple reflections in the fiber medium.

Digital On-Off-Keying Receiver

It is instructive to define a normalized sensitivity as the average number of photons per bit ($\bar{N}_p$) to achieve a given error probability, which we take here to be $P_e = 10^{-9}$. Given $\bar{N}_p$, the received power when a 1 is transmitted is obtained from

$$P_R = 2\bar{N}_p R_b \frac{hc}{\lambda} \quad (12)$$

where the factor of 2 in Eq. (12) is because P_R is the peak power, and $\bar{N}_p$ is the average number of photons per bit.

Ideal Receiver In an ideal receiver individual photons may be counted, and the only source of noise is the fluctuation of the number of photons counted when a 1 is transmitted. This is a Poisson random variable with mean $2\bar{N}_p$. No photons are received when a 0 is transmitted. Consequently, an error is made only when a 1 is transmitted and no photons are received. This leads to the following expression for the error probability

$$P_e = \frac{1}{2} \exp(-2\bar{N}_p) \quad (13)$$

from which it follows that $\bar{N}_p = 10$ for $P_e = 10^{-9}$. This is termed the *quantum limit*.

PIN Receiver In a PIN receiver, the photodetector output is amplified, filtered, and sampled, and the sample is compared with a threshold to decide whether a 1 or 0 was transmitted. Let I be the sampled current at the input to the decision circuit scaled back to the corresponding value at the output of the photodetector. (It is convenient to refer all signal and noise levels to their equivalent values at the output of the photodetector.) I is then a random variable with means and variances given by

$$\mu_1 = I_p \quad \mu_0 = 0 \quad (14a)$$

$$\sigma_1^2 = 2qI_p B + \frac{4kTB}{R_e} \quad \sigma_0^2 = \frac{4kTB}{R_e} \quad (14b)$$

where the subscripts 1 and 0 refer to the bit transmitted, kT is the thermal noise energy, and R_e is the effective input noise resistance of the amplifier. Note that the noise values in the 1 and 0 states are different owing to the shot noise in the 1 state.

Calculation of error probability requires knowledge of the distribution of I under the two hypotheses. Under the assumption that these distributions may be approximated by gaussian distributions

with means and variances given by Eq. (14), the error probability may be shown to be given by (Chap. 4 in Ref. 19)

$$P_e = K \left(\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0} \right) \quad (15)$$

where

$$K(Q) = \frac{1}{\sqrt{2\pi}} \int_Q^\infty dx \exp(-x^2/2) = \frac{1}{2} \operatorname{erfc}(Q/\sqrt{2}) \quad (16)$$

It can be shown from Eqs. (11), (12), (14), and (15) that

$$\bar{N}_p = \frac{B}{\eta R_b} Q^2 \left[1 + \frac{1}{Q} \sqrt{\frac{8\pi kTC_e}{q^2}} \right] \quad (17)$$

where

$$C_e = \frac{1}{2\pi R_e B} \quad (18)$$

is the effective noise capacitance of the receiver, and from Eq. (16), $Q = 6$ for $P_e = 10^{-9}$. The minimum bandwidth of the receiver is half the bit rate, but in practice B/R_b is generally about 0.7.

The gaussian approximation is expected to be good when the thermal noise is large compared to the shot noise. It is interesting, however, to note that Eq. (17) gives $\bar{N}_p = 18$, when $C_e = 0$, $B/R_b = 0.5$, $\eta = 1$, and $Q = 6$. Thus, even in the shot noise limit, the gaussian approximation gives a surprisingly close result to the value calculated from the correct Poisson distribution. It must be pointed out, however, that the location of the threshold calculated by the gaussian approximation is far from correct in this case. In general, the gaussian approximation is much better in estimating receiver sensitivity than in establishing where to set receiver thresholds.

Low-input-impedance amplifiers are generally required to achieve the high bandwidths required for high-bit-rate systems. However, a low input impedance results in high thermal noise and poor sensitivity. High-input-impedance amplifiers may be used, but this narrows the bandwidth, which must be compensated for by equalization following the first-stage amplifier. Although this may result in a highly sensitive receiver, the receiver will have a poor dynamic range owing to the high gains required in the equalizer.²⁰ Receivers for digital systems are generally implemented with transimpedance amplifiers having a large feedback resistance. This reduces the effective input noise capacitance to below the capacitance of the photodiode, and practical receivers can be built with $C_e \approx 0.1 pF$. Using this value of capacitance and $B/R_b = 0.7$, $\eta = 0.7$, and $Q = 6$, Eq. (17) gives $\bar{N}_p \approx 2600$. Note that this is about 34 dB greater than the value given by the quantum limit.

APD Receiver In an APD receiver, there is additional shot noise owing to the excess noise factor F of the avalanche gain process. However, thermal noise is reduced because of the current multiplication gain M before thermal noise is introduced. This results in a receiver sensitivity given approximately by*

$$\bar{N}_p = \frac{B}{\eta R_b} Q^2 \left[F + \frac{1}{Q} \sqrt{\frac{8\pi kTC_e}{q^2 M^2}} \right] \quad (19)$$

*The gaussian approximation is not as good for an APD as for a PIN receiver owing to the nongaussian nature of the excess APD noise.

The excess noise factor is an increasing function of M , which results in an optimum M to minimize $\bar{N}_p$.²⁰ Good APD receivers at 1300 and 1550 nm typically have sensitivities of the order of 1000 photons per bit. Owing to the lower excess noise of silicon APDs, sensitivity of about 500 photons per bit can be achieved at 850 nm.

Impairments There are several sources of impairment that may degrade the sensitivity of receivers from the values given by Eqs. (17) and (19). These may be grouped into two general classes: eye degradations and signal-dependent noise.

An eye diagram is the superposition of all possible received sequences. At the sampling point, there is a spread of the values of a received 1 and a received 0. The difference between the minimum value of a received 1 and the maximum value of the received 0 is known as the *eye opening*. This is given by $(1 - \varepsilon)I_p$ where ε is the eye degradation. The two major sources of eye degradation are intersymbol interference and finite laser extinction ratio. Intersymbol interference results from dispersion, deviations from ideal shaping of the receiver filter, and low-frequency cutoff effects that result in direct current (DC) offsets.

Signal-dependent noises are phenomena that give a variance of the received photocurrent that is proportional to I_p^2 and consequently lead to a maximum signal-to-noise ratio at the output of the receiver. Principal sources of signal-dependent noise are laser relative intensity noise (RIN), reflection-induced noise, mode partition noise, and modal noise. RIN is a consequence of inherent fluctuations in laser intensity resulting from spontaneous emission. This is generally sufficiently small that it is not of concern in digital systems, but is an important limitation in analog systems requiring high signal-to-noise ratios (see Sec. 9.7). Reflection-induced noise is the conversion of laser phase noise to intensity noise by multiple reflections from discontinuities (such as at imperfect connectors.) This may result in a substantial RIN enhancement that can seriously affect digital as well as analog systems.²¹ Mode partition noise occurs when Fabry-Perot lasers are used with dispersive fiber. Fiber dispersion results in changing phase relation between the various laser modes, which results in intensity fluctuations. The effect of mode partition noise is more serious than that of dispersion alone.²² Modal noise is a similar phenomenon that occurs in multimode fiber when relatively few modes are excited and these interfere.

Eye degradations are accounted for by replacing Eq. (14a) by

$$\mu_1 - \mu_0 = (1 - \varepsilon)I_p \quad (20a)$$

and signal-dependent noise by replacing Eq. (14b) by

$$\sigma_1^2 = 2qI_pB + \frac{4kTB}{R_e} + \alpha^2 I_p^2 B \quad \sigma_0^2 = \frac{4kTB}{R_e} + \alpha^2 I_p^2 B \quad (20b)$$

where α^2 is the relative spectral density of the signal-dependent noise. (It is assumed that the signal-dependent noise has a large bandwidth compared to the signal bandwidth B .) With these modifications, the sensitivity of an APD receiver becomes

$$\bar{N}_p = \frac{\frac{B}{\eta R_b} \left(\frac{Q}{1 - \varepsilon} \right)^2 \left[F + \left(\frac{1 - \varepsilon}{Q} \right) \sqrt{\frac{8\pi kTC_e}{q^2 M^2}} \right]}{1 - \alpha^2 B \left(\frac{Q}{1 - \varepsilon} \right)^2} \quad (21)$$

The sensitivity of a PIN receiver is obtained by setting $F = 1$ and $M = 1$ in Eq. (21). It follows from Eq. (21) that there is a minimum error probability (*error floor*) given by

$$P_{e,\min} = K(Q_{\max}) \text{ where } Q_{\max} = \frac{1 - \varepsilon}{\alpha \sqrt{B}} \quad (22)$$

The existence of eye degradations and signal-dependent noise causes an increase in the receiver power (called *power penalty*) required to achieve a given error probability.

9.4 BIT RATE AND DISTANCE LIMITS

Bit rate and distance limitations of digital links are determined by loss and dispersion limitations. The following example is used to illustrate the calculation of the maximum distance for a given bit rate. Consider a 2.5 Gbit/s system at a wavelength of 1550 nm. Assume an average transmitter power of 0 dBm coupled into the fiber. Receiver sensitivity is taken to be 3000 photons per bit, which from Eq. (12) corresponds to an average receiver power of -30.2 dBm. Allowing a total of 8 dB for margin and for connector and cabling losses at the two ends gives a loss allowance of 22.2 dB. If the cabled fiber loss, including splices, is 0.25 dB/km, this leads to a loss-limited transmission distance of 89 km.

Assuming that the fiber dispersion is $D = 15$ ps/km · nm and source spectral width is 0.1 nm, this gives a dispersion per unit length of 1.5 ps/km. Taking the maximum allowed dispersion to be half the interpulse period, this gives a maximum dispersion of 200 ps, which then yields a maximum dispersion-limited distance of 133 km. Thus, the loss-limited distance is controlling.

Consider what happens if the bit rate is increased to 10 Gbit/s. For the same number of photons per bit at the receiver, the receiver power must be 6 dB greater than that in the preceding example. This reduces the loss allowance by 6 dB, corresponding to a reduction of 24 km in the loss-limited distance. The loss-limited distance is now 65 km (assuming all other parameters are unchanged). However, dispersion-limited distance scales inversely with bit rate, and is now 22 km. The system is now dispersion-limited. Dispersion-shifted fiber would be required to be able to operate at the loss limit.

Increasing Bit Rate

There are two general approaches for increasing the bit rate transmitted on a fiber: time-division multiplexing (TDM), in which the serial transmission rate is increased, and wavelength-division multiplexing (WDM), in which separate wavelengths are used to transmit independent serial bit streams in parallel. TDM has the advantage of minimizing the quantity of active devices but requires higher-speed electronics as the bit rate is increased. Also, as indicated by the preceding example, dispersion limitations will be more severe.

WDM allows use of existing lower-speed electronics, but requires multiple lasers and detectors as well as optical filters for combining and separating the wavelengths. Technology advances, including tunable lasers, transmitter and detector arrays, high-resolution optical filters, and optical amplifiers (Sec. 9.5) have made WDM more attractive, particularly for networking applications (Sec. 9.6).

Longer Repeater Spacing

In principal, there are three approaches for achieving longer repeater spacing than that calculated in the preceding text: lower fiber loss, higher transmitter powers, and improved receiver sensitivity (smaller $(\bar{N}_p)$). Silica-based fiber is already essentially at the theoretical Rayleigh scattering loss limit. There has been research on new fiber materials that would allow operation at wavelengths longer than 1.6 μm , with consequent lower theoretical loss values.²³ There are many reasons, however, why achieving such losses will be difficult, and progress in this area has been slow.

Higher transmitter powers are possible, but there are both nonlinearity and reliability issues that limit transmitter power. Since present receivers are more than 30 dB above the quantum limit, improved receiver sensitivity would appear to offer the greatest possibility. To improve the receiver sensitivity, it is necessary to increase the photocurrent at the output of the detector without introducing significant excess loss. There are two main approaches for doing so: optical amplification and optical mixing. Optical preamplifiers result in a theoretical sensitivity of 38 photons per bit²⁴ (6 dB above the quantum limit), and experimental systems have been constructed with sensitivities of about 100 photons per bit.²⁵ This will be discussed further in Sec. 9.5. Optical mixing (coherent receivers) will be discussed briefly in the following text.

Coherent Systems A photodetector provides an output current proportional to the magnitude square of the electric field that is incident on the detector. If a strong optical signal (*local oscillator*) coherent in phase with the incoming optical signal is added prior to the photodetector, then the photocurrent will contain a component at the difference frequency between the incoming and local oscillator signals. The magnitude of this photocurrent, relative to the direct detection case, is increased by the ratio of the local oscillator to the incoming field strengths. Such a coherent receiver offers considerable improvement in receiver sensitivity. With on-off keying, a heterodyne receiver (signal and local oscillator frequencies different) has a theoretical sensitivity of 36 photons per bit, and a homodyne receiver (signal and local oscillator frequencies the same) has a sensitivity of 18 photons per bit. Phase-shift keying (possible with coherent systems) provides a further 3-dB improvement. Coherent systems, however, require very stable signal and local oscillator sources (spectral linewidths need to be small compared to the modulation bandwidth) and matching of the polarization of the signal and local oscillator fields.²⁶ Differentially coherent systems (e.g., DPSK) in which the prior bit is used as a phase reference are simpler to implement and are beginning to find application.²⁷

An advantage of coherent systems, more so than improved receiver sensitivity, is that because the output of the photodetector is linear in the signal field, filtering for WDM demultiplexing may be done at the difference frequency (typically in the microwave range).^{*} This allows considerably greater selectivity than is obtainable with optical filtering techniques. The advent of optical amplifiers has slowed the interest in coherent systems, but there has been renewed interest in recent years.²⁸

9.5 OPTICAL AMPLIFIERS

There are two types of optical amplifiers: laser amplifiers based on stimulated emission and parametric amplifiers based on nonlinear effects (Chap. 10 in Ref. 32). The former are currently of most interest in fiber-optic communications. A laser without reflecting end faces is an amplifier, but it is more difficult to obtain sufficient gain for amplification than it is (with feedback) to obtain oscillation. Thus, laser oscillators were available much earlier than laser amplifiers.

Laser amplifiers are now available with gains in excess of 30 dB over a spectral range of more than 30 nm. Output saturation powers in excess of 10 dBm are achievable. The amplified spontaneous emission (ASE) noise power at the output of the amplifier, in each of two orthogonal polarizations, is given by

$$P_{\text{ASE}} = n_{\text{sp}} \frac{hc}{\lambda} B_o (G - 1) \quad (23)$$

where G is the amplifier gain, B_o is the bandwidth, and the spontaneous emission factor n_{sp} is equal to 1 for ideal amplifiers with complete population inversion.

Comparison of Semiconductor and Fiber Amplifiers

There are two principal types of laser amplifiers: semiconductor laser amplifiers (SLAs) and doped-fiber amplifiers. The erbium-doped-fiber amplifier (EDFA), which operates at a wavelength of 1.55 μm , is of most current interest.

The advantages of the SLA, similar to laser oscillators, are that it is pumped by a DC current, it may be designed for any wavelength of interest, and it can be integrated with electrooptic semiconductor components.

^{*}The difference frequency must be large compared to the modulation bandwidth. As modulation bandwidths have increased beyond 10 GHz this may necessitate difference frequencies greater than 100 GHz which may be difficult to implement.

The advantages of the EDFA are that there is no coupling loss to the transmission fiber, it is polarization-insensitive, it has lower noise than SLAs, it can be operated at saturation with no intermodulation owing to the long time constant of the gain dynamics, and it can be integrated with fiber devices. However, it does require optical pumping, with the principal pump wavelengths being either 980 or 1480 nm.

Communications Application of Optical Amplifiers

There are four principal applications of optical amplifiers in communication systems.^{29,30}

1. Transmitter power amplifiers
2. Compensation for splitting loss in distribution networks
3. Receiver preamplifiers
4. Linear repeaters in long-distance systems

The last application is of particular importance for long-distance networks (particularly under-sea systems), where a bit-rate-independent linear repeater allows subsequent upgrading of system capacity (either TDM or WDM) with changes only at the system terminals. Although amplifier noise accumulates in such long-distance linear systems, transoceanic lengths are achievable with amplifier spacings of about 60 km corresponding to about 15-dB fiber attenuation between amplifiers.

However, in addition to the accumulation of ASE, there are other factors limiting the distance of linearly amplified systems, namely dispersion and the interaction of dispersion and nonlinearity.³¹ There are two alternatives for achieving very long-distance, very high-bit-rate systems with linear repeaters: *solitons*, which are pulses that maintain their shape in a dispersive medium,³² and dispersion compensation.³³

9.6 FIBER-OPTIC NETWORKS

Networks are communication systems used to interconnect a number of terminals within a defined geographic area—for example, local area networks (LANs), metropolitan area networks (MANs), and wide area networks (WANs). In addition to the transmission function discussed throughout the earlier portions of this chapter, networks also deal with the routing and switching aspects of communications.

Passive optical networks utilize couplers to distribute signals to users. In an $N \times N$ ideal star coupler, the signal on each input port is uniformly distributed among all output ports. If an average power P_T is transmitted at a transmitting port, the power received at a receiving port (neglecting transmission losses) is

$$P_R = \frac{P_T}{N} (1 - \delta_N) \quad (24)$$

where δ_N is the excess loss of the coupler. If N is a power of 2, an $N \times N$ star may be implemented by $\log_2 N$ stages of 2×2 couplers. Thus, it may be conservatively assumed that

$$1 - \delta_N = (1 - \delta_2)^{\log_2 N} = N^{\log_2(1 - \delta_2)} \quad (25)$$

The maximum bit rate per user is given by the average received power divided by the product of the photon energy and the required number of photons per bit (N_p). The throughput Y is the product of the number of users and the bit rate per user, and from Eqs. (24) and (25) is therefore given by

$$Y = \frac{P_T}{N_p} \frac{\lambda}{hc} N^{\log_2(1 - \delta_2)} \quad (26)$$

Thus, the throughput (based on power considerations) is independent of N for ideal couplers ($\delta_2 = 0$) and decreases slowly with $N(N^{-0.17})$ for $10 \log(1 - \delta_2) = 0.5$ dB. It follows from Eq. (26) that for a power of 1 mW at $\lambda = 1.55 \mu\text{m}$ and with $N_p = 3000$, the maximum throughput is 2.6 Tbit/s.

This may be contrasted with a tapped bus, where it may be shown that optimum tap weight to maximize throughput is given by $1/N$, leading to a throughput given by³⁴

$$Y = \frac{P_T}{N_p} \frac{\lambda}{hc} \frac{1}{Ne^2} \exp(-2N\delta) \quad (27)$$

Thus, even for ideal ($\delta = 0$) couplers, the throughput decreases inversely with the number of users. If there is excess coupler loss, the throughput decreases exponentially with the number of users and is considerably less than that given by Eq. (26). Consequently, for a power-limited transmission medium, the star architecture is much more suitable than the tapped bus. The same conclusion does not apply to metallic media, where bandwidth rather than power limits the maximum throughput.

Although the preceding text indicates the large throughput that may be achieved in principle with a passive star network, it doesn't indicate how this can be realized. Most interest is in WDM networks.³⁵ The simplest protocols are those for which fixed-wavelength receivers and tunable transmitters are used. However, the technology is simpler when fixed-wavelength transmitters and tunable receivers are used, since a tunable receiver may be implemented with a tunable optical filter preceding a wideband photodetector. Fixed-wavelength transmitters and receivers involving multiple passes through the network are also possible, but this requires utilization of terminals as relay points. Protocol, technology, and application considerations for gigabit networks (networks having access at gigabit rates and throughputs at terabit rates) is an extensive area of research.³⁶

9.7 ANALOG TRANSMISSION ON FIBER

Most interest in fiber-optic communications is centered around digital transmission, since fiber is generally a power-limited rather than a bandwidth-limited medium. There are applications, however, where it is desirable to transmit analog signals directly on fiber without converting them to digital signals. Examples are cable television (CATV) distribution and microwave links such as entrance links to antennas and interconnection of base stations in mobile radio systems.

Carrier-to-Noise Ratio (CNR)

Optical intensity modulation is generally the only practical modulation technique for incoherent-detection fiber-optic systems. Let $f(t)$ be the carrier signal that intensity modulates the optical source. For convenience, assume that the average value of $f(t)$ is equal to 0, and that the magnitude of $f(t)$ is normalized to be less than or equal to 1. The received optical power may then be expressed as

$$P(t) = P_o[1 + mf(t)] \quad (28)$$

where m is the optical modulation index

$$m = \frac{P_{\max} - P_{\min}}{P_{\max} + P_{\min}} \quad (29)$$

The carrier-to-noise ratio is then given by

$$\text{CNR} = \frac{\frac{1}{2} m^2 \Re^2 P_o^2}{\text{RIN} \Re^2 P_o^2 B + 2q \Re P_o B + \langle i_{\text{th}}^2 \rangle B} \quad (30)$$

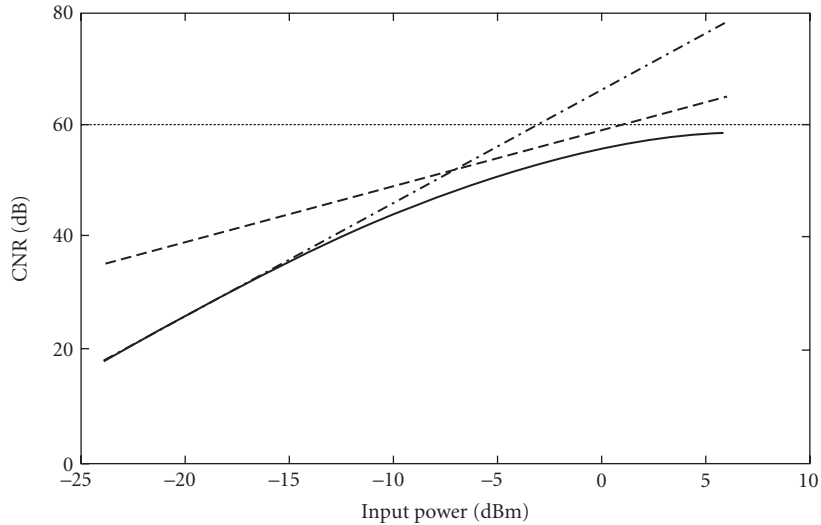


FIGURE 3 CNR as a function of input power. Straight lines indicate thermal noise (-----), shot noise (---), and RIN (.....) limits.

where $\mathfrak{R}$ is the photodetector responsivity, RIN is the relative intensity noise spectral density (denoted by α^2 in Sec. 9.3), and $\langle i_{th}^2 \rangle$ is the thermal noise spectral density (expressed as $4kT/R_e$ in Sec. 9.3). CNR is plotted in Fig. 3 as a function of received optical power for a bandwidth of $B = 4$ MHz (single video channel), optical modulation index $m = 0.05$, $\mathfrak{R} = 0.8$ A/W, $RIN = -155$ dB/Hz, and $\sqrt{\langle i_{th}^2 \rangle} = 7$ pA/ $\sqrt{\text{Hz}}$. At low received powers (typical of digital systems) the CNR is limited by thermal noise. However, to obtain the higher CNR generally needed by analog systems, shot noise and then ultimately laser RIN become limiting.

Analog Video Transmission on Fiber³⁷

It is helpful to distinguish between single-channel and multiple-channel applications. For the single-channel case, the video signal may directly modulate the laser intensity [amplitude-modulated (AM) system], or the video signal may be used to frequency-modulate an electrical subcarrier, with this subcarrier then intensity-modulating the optical source [frequency-modulated (FM) system]. Equation (30) gives the CNR of the recovered subcarrier. Subsequent demodulation of the FM signal gives an additional increase in signal-to-noise ratio. In addition to this FM improvement factor, larger optical modulation indexes may be used than in AM systems. Thus FM systems allow higher signal-to-noise ratios and longer transmission spans than AM systems.

Two approaches have been used to transmit multichannel video signals on fiber. In the first (AM systems), the video signals undergo electrical frequency-division multiplexing (FDM), and this combined FDM signal intensity modulates the optical source. This is conceptually the simplest system, since existing CATV multiplexing formats may be used.

In FM systems, the individual video channels frequency-modulate separate microwave carriers (as in satellite systems). These carriers are linearly combined and the combined signal intensity modulates a laser. Although FM systems are more tolerant than AM systems to intermodulation distortion and noise, the added electronics costs have made such systems less attractive than AM systems for CATV application.

Multichannel AM systems are of interest not only for CATV application but also for mobile radio applications to connect signals from a microcellular base station to a central processing station.

Relative to CATV applications, the mobile radio application has the additional complication of being required to accommodate signals over a wide dynamic power range.

Nonlinear Distortion

In addition to CNR requirements, multichannel analog communication systems are subject to intermodulation distortion. If the input to the system consists of a number of tones at frequencies ω_i , then nonlinearities result in intermodulation products at frequencies given by all sums and differences of the input frequencies. Second-order intermodulation gives intermodulation products at frequencies $\omega_i \pm \omega_j$, whereas third-order intermodulation gives frequencies $\omega_i \pm \omega_j \pm \omega_k$. If the signal frequency band is such that the maximum frequency is less than twice the minimum frequency, then all second-order intermodulation products fall outside the signal band, and third-order intermodulation is the dominant nonlinearity. This condition is satisfied for the transport of microwave signals (e.g., mobile radio signals) on fiber, but is not satisfied for wideband CATV systems, where there are requirements on composite second-order (CSO) and composite triple-beat (CTB) distortion.

The principal causes of intermodulation in multichannel fiber-optic systems are laser threshold nonlinearity,³⁸ inherent laser gain nonlinearity, and the interaction of chirp and dispersion.

9.8 TECHNOLOGY AND APPLICATIONS DIRECTIONS

Fiber-optic communication application in the United States began with metropolitan and short-distance intercity trunking at a bit rate of 45 Mbit/s, corresponding to the DS-3 rate of the North American digital hierarchy. Technological advances, primarily higher-capacity transmission and longer repeater spacings, extended the application to long-distance intercity transmission, both terrestrial and undersea. Also, transmission formats are now based on the synchronous digital hierarchy (SDH), termed synchronous optical network (SONET) in the U.S. OC-192 system* operating at 10 Gbit/s are widely deployed, with OC-768 40 Gbit/s systems also available. All of the signal processing in these systems (multiplexing, switching, performance monitoring) is done electrically, with optics serving solely to provide point-to-point links.

For long-distance applications, 10 Gbit/s dense wavelength-division multiplexing (DWDM), with channel spacings of 50 GHz and with upward of 100 wavelength channels, has extended the bit rate capability of fiber to greater than 1 Tbit/s in commercial systems and more than 3 Tbit/s in laboratory trials.³⁹ For local access, there is extensive interest in fiber directly to the premises⁴⁰ as well as hybrid combinations of optical and electronic technologies and transmission media.^{41,42}

The huge bandwidth capability of fiber optics (measured in tens of terahertz) is not likely to be utilized by time-division techniques alone, and DWDM technology and systems are receiving considerable emphasis, although work is also under way on optical time-division multiplexing (OTDM) and optical code-division multiplexing (OCDM).

Nonlinear phenomena, when uncontrolled, generally lead to system impairments. However, controlled nonlinearities are the basis of devices such as parametric amplifiers and switching and logic elements. Nonlinear optics will consequently continue to receive increased emphasis.

9.9 REFERENCES

1. J. Hecht, *City of Light: The Story of Fiber Optics*, Oxford University Press, New York, 1999.
2. C. K. Kao and G. A. Hockham, "Dielectric-Fiber Surface Waveguides for Optical Frequencies," *Proc. IEEE* **113**:1151-1158 (July 1966).
3. F. P. Kapron, et al., "Radiation Losses in Glass Optical Waveguides," *Appl. Phys. Lett.* **17**:423 (November 15, 1970).

*OC-*n* systems indicate optical channel at a bit rate of (51.84)*n* Mbit/s.

4. I. Hayashi, M. B. Panish, and P. W. Foy, "Junction Lasers which Operate Continuously at Room Temperature," *Appl. Phys. Lett.* **17**:109 (1970).
5. I. Jacobs, "Optical Fiber Communication Technology and System Overview," in O. D. D. Soares (ed.), *Trends in Optical Fibre Metrology and Standards*, NATO ASI Series, vol. 285, pp. 567–591, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.
6. D. Gloge, "Weakly Guiding Fibers," *Appl. Opt.* **10**:2252–2258 (October 1971).
7. R. Olshansky and D. Keck, "Pulse Broadening in Graded Index Fibers," *Appl. Opt.* **15**:483–491 (February 1976).
8. D. Gloge, E. A. J. Marcatili, D. Marcuse, and S. D. Personick, "Dispersion Properties of Fibers," in S. E. Miller and A. G. Chynoweth (eds.), *Optical Fiber Telecommunications*, chap. 4, Academic Press, New York, 1979.
9. Y. Namihiro and H. Wakabayashi, "Fiber Length Dependence of Polarization Mode Dispersion Measurements in Long-Length Optical Fibers and Installed Optical Submarine Cables," *J. Opt. Commun.* **2**:2 (1991).
10. W. B. Jones Jr., *Introduction to Optical Fiber Communication Systems*, pp. 90–92, Holt, Rinehart and Winston, New York, 1988.
11. L. G. Cohen, W. L. Mammel, and S. J. Jang, "Low-Loss Quadruple-Clad Single-Mode Lightguides with Dispersion Below 2 ps/km · nm over the 1.28 μm –1.65 μm Wavelength Range," *Electron. Lett.* **18**:1023–1024 (1982).
12. G. Keiser, *Optical Fiber Communications*, 3d ed., chap. 5, McGraw-Hill, New York, 2000.
13. N. M. Margalit, S. Z. Zhang, and J. E. Bowers, "Vertical Cavity Lasers for Telecom Applications," *IEEE Commun. Mag.* **35**:164–170 (May 1997).
14. J. E. Bowers and M. A. Pollack, "Semiconductor Lasers for Telecommunications," in S. E. Miller and I. P. Kaminow (eds.), *Optical Fiber Telecommunications II*, chap. 13, Academic Press, San Diego, CA, 1988.
15. W. T. Tsang, "The Cleaved-Coupled-Cavity (C^3) Laser," in *Semiconductors and Semimetals*, vol. 22, part B, chap. 5, pp. 257–373, 1985.
16. K. Kobayashi and I. Mito, "Single Frequency and Tunable Laser Diodes," *J. Lightwave Technol.* **6**:1623–1633 (November 1988).
17. T. Mukai and Y. Yamamoto, "AM Quantum Noise in 1.3 μm InGaAsP Lasers," *Electron. Lett.* **20**:29–30 (January 5, 1984).
18. J. Buus and E. J. Murphy, "Tunable Lasers in Optical Networks," *J. Lightwave Technol.* **24**:5–11 (January 2006).
19. G. P. Agrawal, *Fiber-Optic Communication Systems*, 3d ed., Wiley Interscience, New York, 2002.
20. S. D. Personick, "Receiver Design for Digital Fiber Optic Communication Systems I," *Bell Syst. Tech. J.* **52**: 843–874 (July–August 1973).
21. J. L. Gimlett and N. K. Cheung, "Effects of Phase-to-Intensity Noise Conversion by Multiple Reflections on Gigabit-per-Second DFB Laser Transmission Systems," *J. Lightwave Technol.* **LT-7**:888–895 (June 1989).
22. K. Ogawa, "Analysis of Mode Partition Noise in Laser Transmission Systems," *IEEE J. Quantum Electron.* **QE-18**:849–855 (May 1982).
23. D. C. Tran, G. H. Sigel, and B. Bendow, "Heavy Metal Fluoride Fibers: A Review," *J. Lightwave Technol.* **LT-2**:566–586 (October 1984).
24. P. S. Henry, "Error-Rate Performance of Optical Amplifiers," *Optical Fiber Communications Conference (OFC'89 Technical Digest)*, THK3, Houston, Texas, February 9, 1989.
25. O. Gautheron, G. Grandpierre, L. Pierre, J.-P. Thiery, and P. Kretzmeyer, "252 km Repeaterless 10 Gbits/s Transmission Demonstration," *Optical Fiber Communications Conference (OFC'93) Post-deadline Papers*, PD11, San Jose, California, February 21–26, 1993.
26. I. W. Stanley, "A Tutorial Review of Techniques for Coherent Optical Fiber Transmission Systems," *IEEE Commun. Mag.* **23**:37–53 (August 1985).
27. A. H. Gnauck, S. Chandrasekhar, J. Leutholdt, and L. Stulz, "Demonstration of 42.7 Gb/s DPSK Receiver with 45 Photons/Bit Sensitivity," *Photonics Technol. Letts.* **15**:99–101 (January 2003).
28. E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, "Coherent Detection Ion Optical Fiber Systems," *Optics Express* **16**:753–791 (January 9, 2008).
29. Bellcore, "Generic Requirements for Optical Fiber Amplifier Performance," Technical Advisory TA-NWT-001312, Issue 1, December 1992.
30. T. Li, "The Impact of Optical Amplifiers on Long-Distance Lightwave Telecommunications," *Proc. IEEE* **81**:1568–1579 (November 1993).

31. A. Naka and S. Saito, "In-Line Amplifier Transmission Distance Determined by Self-Phase Modulation and Group-Velocity Dispersion," *J. Lightwave Technol.* **12**:280–287 (February 1994).
32. G. P. Agrawal, *Nonlinear Fiber Optics*, 3d ed., chap. 5, Academic Press, San Diego, CA, 2001.
33. Bob Jopson and Alan Gnauck, "Dispersion Compensation for Optical Fiber Systems," *IEEE Commun. Mag.* **33**:96–102 (June 1995).
34. P. E. Green, Jr., *Fiber Optic Networks*, chap. 11, Prentice Hall, Englewood Cliffs, NJ, 1993.
35. M. Fujiwara, M. S. Goodman, M. J. O'Mahony, O. K. Tonguz, and A. E. Willner (eds.), Special Issue on Multiwavelength Optical Technology and Networks, *J. Lightwave Technology* **14**(6):932–1454 (June 1996).
36. P. J. Smith, D. W. Faulkner, and G. R. Hill, "Evolution Scenarios for Optical Telecommunication Networks Using Multiwavelength Transmission," *Proc. IEEE* **81**:1580–1587 (November 1993).
37. T. E. Darcie, K. Nawata, and J. B. Glabb, Special Issue on Broad-Band Lightwave Video Transmission, *J. Lightwave Technol.* **11**(1) (January 1993).
38. A. A. M. Saleh, "Fundamental Limit on Number of Channels in SCM Lightwave CATV System," *Electron. Lett.* **25**(12):776–777 (1989).
39. A. H. Gnauck, G. Charlet, P. Tran, et al., "25.6 Tb/s WDM Transmission of Polarization Multiplexed RZ-DQPSK Signals," *J. Lightwave Technol.* **26**:79–84 (January 1, 2008).
40. T. Koonen, "Fiber to the Home/Fiber to the Premises: What, Where, and When?" *Proc. IEEE* **94**:911–934 (May 2006).
41. C. Baack and G. Walf, "Photonics in Future Telecommunications," *Proc. IEEE* **81**:1624–1632 (November 1993).
42. G. C. Wilson, T. H. Wood, J. A. Stiles, et al., "FiberVista: An FTTH or FTTC System Delivering Broadband Data and CATV Services," *Bell Labs Tech. J.* **4**:300–322 (January–March 1999).

This page intentionally left blank

NONLINEAR EFFECTS IN OPTICAL FIBERS

John A. Buck

*Georgia Institute of Technology
School of Electrical and Computer Engineering
Atlanta, Georgia*

Fiber nonlinearities are important in optical communications, both as useful attributes and as characteristics to be avoided. They must be considered when designing long-range high-data-rate systems that involve high optical power levels and in which signals at multiple wavelengths are transmitted. The consequences of nonlinear transmission can include (1) the generation of additional signal bandwidth within a given channel, (2) modifications of the phase and shape of pulses, (3) the generation of light at other wavelengths at the expense of power in the original signal, and (4) crosstalk between signals at different wavelengths and polarizations. The first two, arising from self-phase modulation, can be used to advantage in the generation of *solitons*—pulses whose nonlinear phase modulation compensates for linear group dispersion in the fiber link¹ or in fiber gratings,² leading to pulses that propagate without changing shape or width (see Chap. 22). The third and fourth effects arise from stimulated Raman or Brillouin scattering or four-wave mixing. These can be used to advantage when it is desired to generate or amplify additional wavelengths, but they must usually be avoided in systems.

10.1 KEY ISSUES IN NONLINEAR OPTICS IN FIBERS

Optical fiber waveguides, being of glass compositions, do not possess large nonlinear coefficients. Nonlinear processes can nevertheless occur with high efficiencies since intensities are high and propagation distances are long. Even though power levels are usually modest (a few tens of milliwatts), intensities within the fiber are high due to the small cross-sectional areas involved. This is particularly true in single-mode fiber, where the LP_{01} mode typically presents an effective cross-sectional area of between 10^{-7} and 10^{-8} cm², thus leading to intensities on the order of MW/cm². Despite this, long interaction distances are usually necessary to achieve nonlinear mixing of any significance, so processes must be phase matched, or nearly so. Strategies to avoid unwanted nonlinear effects usually involve placing upper limits on optical power levels, and if possible, choosing other parameters such that phase mismatching occurs. Such choices may include wavelengths or wavelength spacing in wavelength-division multiplexed systems, or may be involved in special fiber waveguide designs.³

The generation of light through nonlinear mixing arises through polarization of the medium, which occurs through its interaction with intense light. The polarization consists of an array of phased dipoles in which the dipole moment is a nonlinear function of the applied field strength. In the classical picture, the dipoles, once formed, reradiate light to form the nonlinear output. The medium polarization is conveniently expressed through a power series expansion involving products of real electric fields:

$$\mathcal{P} = \epsilon_0 [\chi^{(1)} \cdot \mathcal{E} + \chi^{(2)} \cdot \mathcal{E}\mathcal{E} + \chi^{(3)} \cdot \mathcal{E}\mathcal{E}\mathcal{E} + \dots] = \mathcal{P}_L + \mathcal{P}_{NL} \quad (1)$$

in which the χ terms are the linear, second-, and third-order susceptibilities. Nonlinear processes are described through the product of two or more optical fields to form the nonlinear polarization, $\mathcal{P}_{NL}$, consisting of all terms of second order and higher in Eq. (1).

The second-order term in Eq. (1) [involving $\chi^{(2)}$] describes three-wave mixing phenomena, such as second-harmonic generation. The third-order term describes four-wave mixing (FWM) processes and stimulated scattering phenomena. In the case of optical fibers, second-order processes are generally not possible, since these effects require noncentrosymmetric media.⁴ In amorphous fiber waveguides, third-order effects [involving $\chi^{(3)}$] are usually seen exclusively, although second-harmonic generation can be observed in special instances.⁵

The interactions between fields and polarizations are described by the nonlinear wave equation:

$$\nabla^2 \mathcal{E} + n_0^2 \mu_0 \epsilon_0 \frac{\partial^2 \mathcal{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \mathcal{P}_{NL}}{\partial t^2} \quad (2)$$

where $\mathcal{E}$ and $\mathcal{P}$ are the sums of all electric fields and nonlinear polarizations that are present, and n_0 is the refractive index of the medium. The second-order differential equation is usually reduced to first order through the slowly varying envelope approximation (SVEA):

$$\left| \frac{\partial^2 E}{\partial z^2} \right| \ll \left| \frac{2\pi}{\lambda} \frac{\partial E}{\partial z} \right| \quad (3)$$

where E is the complex field amplitude. The interpretation of the SVEA is that the changes in field amplitude that occur over distances on the order of a wavelength are very large compared to variations in the rate of change over the same distance. The wave equation will separate according to frequencies or propagation directions, yielding sets of coupled differential equations that, under the SVEA, are first order. These describe the growth or decay of fields involved in the mixing process.

The requirement for phase matching is that the nonlinear polarization wave and the electric field associated with the generated wave propagate with the same phase constant; that is, their phase velocities are equal. Phase-matched processes in fiber include those that involve (1) interacting waves at the same wavelength and polarization, such as self- and cross-phase modulation, as well as other degenerate Kerr-type interactions, and (2) stimulated scattering processes (Raman and Brillouin), in addition to cross-phase modulation involving two wavelengths. Four-wave mixing processes involving light at different wavelengths can occur that are not precisely phase matched but that can nevertheless yield high efficiencies. Matters are further complicated by the fact that different nonlinear processes can occur simultaneously, with each affecting the performance of the other. Nonlinear effects are usually favored to occur under pulsed operation, since high peak powers can be achieved with comparatively modest average powers. Consequently, group velocity matching is desirable (although not always required) to achieve efficient mixing between pulses.

10.2 SELF- AND CROSS-PHASE MODULATION

Self-phase modulation (SPM) can occur whenever a signal having a time-varying amplitude is propagated in a nonlinear material. The origin of the effect is the refractive index of the medium, which will change with the instantaneous signal intensity. The complex nonlinear polarization for the process is:

$$P_{\text{NL}} = \frac{3}{4} \epsilon_0 \chi^{(3)} |E_0(z, t)|^2 E_0(z, t) \exp[i(\omega t - \beta z)] \quad (4)$$

where $E_0(t)$ is the time-varying electric field amplitude that describes the pulse or signal envelope, and where the frequency ω is the same as that of the input light. Incorporating this polarization and the field into the wave equation leads to a modified refractive index over the original zero-field value n_0 . The net index becomes:⁶

$$n = n_0 + n'_2 |E_0(z, t)|^2 \quad (5)$$

where the nonlinear refractive index is given by $n'_2 = \text{Re}\{3\chi^{(3)}/8n_0\}$. In fused silica it has the value $n'_2 = 6.1 \times 10^{-23} \text{ m}^2/\text{V}^2$.⁷ Equation (5) can also be expressed in terms of light intensity through $n(I) = n_0 + n_2 I(z, t)$, where $n_2 = 3.2 \times 10^{-20} \text{ m}^2/\text{W}$. In optical fibers the index is modified from the effective mode index of the single-mode fiber n_{eff} (which assumes the role of n_0).

The complex field as it propagates through the medium can be expressed as:

$$E = E_0(z, t) \exp(i\{\omega_0 t - [n_0 + n_2 I(z, t)] k_0 z\}) \quad (6)$$

which exhibits phase modulation that follows the shape of the intensity envelope. The instantaneous frequency is found through the time derivative of the phase:

$$\omega' = \omega_0 - n_2 k_0 z \frac{\partial I}{\partial t} \quad (7)$$

The effects of self-phase modulation on pulse propagation can be qualitatively observed from Eqs. (6) and (7). First, additional frequency components are placed on the pulse, thus increasing its spectral width. Second, a frequency sweep (chirp) is imposed on the pulse, the direction of which depends on the sign of $\partial I/\partial t$. The latter feature is particularly important in optical fibers, since the imposed frequency sweep from SPM will either add to or subtract from the chirp imposed by linear group dispersion. If the chirp directions for self-phase modulation and group dispersion are opposite, an effective cancellation may occur, leading to the formation of an optical soliton. In more conventional systems in which solitons are not employed, SPM must be considered as a possible benefit or detriment to performance, as some pulse shaping (which could include broadening or compression) can occur;^{8,9} however, such systems can in theory yield excellent performance.¹⁰ Furthermore, in systems employing fiber amplifiers, the change in refractive index associated with the signal-induced upper state population in erbium has been shown to be an important performance factor.¹¹ An additional effect can occur when pulse spectra lie within the anomalous group dispersion regime of the fiber; pulse breakup can occur as a result of *modulation instability*, in which the interplay between dispersive and nonlinear contributions to pulse shaping becomes unstable.¹²

Cross-phase modulation (XPM) is similar to SPM, except that two overlapping but distinguishable pulses (having, for example, different frequencies or polarizations) are involved. One pulse will modulate the index of the medium, which then leads to phase modulation of an overlapping pulse. XPM thus becomes a cross-talk mechanism between two channels if phase encoding is employed or if intensity modulation is used in dispersive systems.^{13,14} No transfer of energy occurs between channels, however, which distinguishes the process from other crosstalk mechanisms in which growth

of signal power in one channel occurs at the expense of power in another. The strength of the effect is enhanced by a factor of 2 over that which can be obtained by a single field acting on itself (the nonlinear refractive index n_2 is effectively doubled in XPM). The XPM process, while twice as strong as SPM, is effectively weakened by the fact that pulses of differing frequencies or polarizations are generally not group velocity matched, and so cannot maintain overlap indefinitely. The efficiency is further reduced if the interaction occurs between cross-polarized waves; in this case the nonlinear tensor element (and thus the effective nonlinear index) is a factor of $1/3$ less than the tensor element that describes copolarized waves.⁶

Self- and cross-phase modulation are analyzed by way of coupled equations of the *nonlinear Schrödinger* form,¹⁶ which describes the evolution over time and position of the electric field envelopes of two pulses, E_{0a} and E_{0b} , where SVEA is used and where pulse widths are on the order of 1 ps or greater:

$$\frac{\partial E_{0a}}{\partial z} + \beta_{1a} \frac{\partial E_{0a}}{\partial t} = -\frac{i}{2} \beta_{2a} \frac{\partial^2 E_{0a}}{\partial t^2} + i\gamma_a |E_{0a}|^2 E_{0a} + i\delta\gamma_a |E_{0b}|^2 E_{0a} - \frac{\alpha_a}{2} E_{0a} \quad (8)$$

$$\frac{\partial E_{0b}}{\partial z} + \beta_{1b} \frac{\partial E_{0b}}{\partial t} = -\frac{i}{2} \beta_{2b} \frac{\partial^2 E_{0b}}{\partial t^2} + i\gamma_b |E_{0b}|^2 E_{0b} + i\delta\gamma_b |E_{0a}|^2 E_{0b} - \frac{\alpha_b}{2} E_{0b} \quad (9)$$

In these equations, β_{1j} ($j=a,b$) are the group delays of the pulses at the two frequencies or polarizations over a unit distance; β_{2j} are the group dispersion parameters associated with the two pulses; and $\gamma_j = n_2 \omega_j / (cA_{\text{eff}})$, where A_{eff} is the effective cross-sectional area of the fiber mode. The coefficient δ is equal to 2 for copolarized pulses of different frequencies and is $2/3$ if the pulses are cross-polarized. Propagation loss characterized by coefficients α_j is assumed. The equation form that describes the propagation with SPM of a single pulse— E_{0a} , for example—is found from Eq. (8) by setting $E_{0b}=0$. The terms on the right sides of Eqs. (8) and (9) describe in order the effects of group dispersion, SPM, XPM, and loss. The equations can be solved using numerical techniques that are described in Refs. 15 and 16.

For subpicosecond pulses, the accuracy of Eqs. (8) and (9) begins to degrade as pulse band-widths increase with decreasing temporal width. Additional terms are usually incorporated in the equations as pulse widths are reduced to the vicinity of 100 fs. These embody (1) cubic dispersion, which becomes important as bandwidth increases, and (2) changes in group velocity with intensity. This latter effect can result in *self-steepening*, in which the pulse trailing edge shortens to the point of forming an optical shock front under appropriate conditions. An additional consequence of broad pulse spectra is that power conversion from high-frequency components within a pulse to those at lower frequencies can occur via stimulated Raman scattering, provided the interacting components are sufficiently separated in wavelength. The effect is an overall red shift of the spectrum. At sufficiently high intensities, cross-coupling between pulses having different center wavelengths can also occur through Raman scattering, regardless of pulse width.

10.3 STIMULATED RAMAN SCATTERING

In *stimulated Raman scattering* (SRS), coupling occurs between copropagating light waves whose frequency difference is in the vicinity of resonances of certain molecular oscillation modes. In silica-based fibers, stretch vibrational resonances occur between Si and O atoms in several possible modes within the glass matrix (see Ref. 17 for illustrations of the important modes in pure silica). In the Stokes process, light at frequency ω_2 (pump wave) is downshifted to light at ω_1 (Stokes wave), with the excess energy being absorbed by the lattice vibrational modes (manifested in the generation of optical phonons). The process is either spontaneous, in which the Stokes wave builds up from noise, or is stimulated, in which both waves are present in sufficient strength to generate a beat frequency that excites the oscillators and promotes coupling. A fiber Raman amplifier works on this principle, in which an input signal at ω_1 experiences gain in the presence of pump light at ω_2 . Figure 1 shows

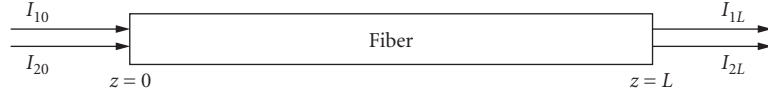


FIGURE 1 Beam geometry for stimulated Raman scattering in an optical fiber.

the beam geometry in which an input wave at ω_1 and intensity I_{10} can emerge at the far end with amplified value I_{1L} . This occurs in the presence of the pump wave at ω_2 that has initial intensity I_{20} and that emerges with depleted intensity I_{2L} .

Back-conversion from ω_1 to ω_2 (the inverse Raman effect) will also occur once the Stokes wave reaches sufficient intensity, but gain will only occur for the Stokes wave. Both processes are phase matched, and so occur with high efficiency in long fibers. The back-conversion process is to be distinguished from anti-Stokes scattering, in which pump light at ω_2 is upshifted to frequency ω_3 , with the additional energy being supplied by optical phonons associated with the previously excited medium. The anti-Stokes process is rarely seen in fiber transmission because (1) it is phase mismatched and (2) it requires a substantial population of excited oscillators, which is not the case at thermal equilibrium.

Figure 2 shows the measured Raman gain for the Stokes wave in fused silica. The gain is plotted as a function of difference frequency between the interacting waves measured in cm^{-1} (to convert this to wavelength shift, use the formula $\Delta\lambda = \lambda_p^2 \Delta f (\text{cm}^{-1})$, where λ_p is the pump wavelength). Other fiber constituents such as GeO_2 , P_2O_5 , and B_2O_3 exhibit their own Raman resonances, which occur at successively greater wavelength shifts;¹⁹ the effects of these will be weak, since their concentration in the fiber is generally small. Thus the dominant Raman shifts in optical fiber are associated with SiO_2 , and occur within the range of 440 to 490 cm^{-1} , as is evident in Fig. 2.

Nonlinear polarizations at frequencies ω_1 and ω_2 can be constructed that are proportional to products of the Stokes and pump fields, $E_1^{\omega_1}$ and $E_2^{\omega_2}$. These are of the form $P_{\text{NL}}^{\omega_1} \propto |E_2^{\omega_2}|^2 E_1^{\omega_1}$ (Stokes

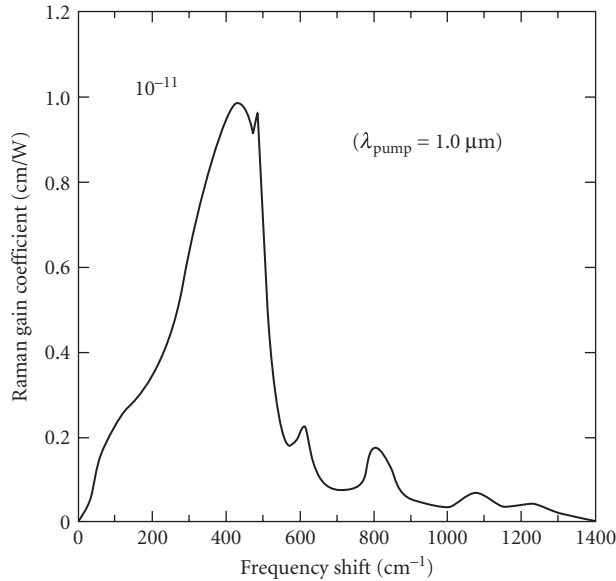


FIGURE 2 Raman gain spectrum in fused silica. (Adapted from Ref. 22. © 1980 IEEE.)

generation) and $P_{\text{NL}}^{\omega_2} \propto |E_1^{\omega_1}|^2 E_2^{\omega_2}$ (the inverse Raman effect). Substituting these polarizations and the two fields into the wave equation, using the SVEA, and assuming copolarized fields leads to the following coupled equations involving the Stokes and pump wave intensities I_1 and I_2 :¹⁸

$$\frac{dI_1}{dz} = g_r I_1 I_2 - \alpha I_1 \quad (10)$$

$$\frac{dI_2}{dz} = -\frac{\omega_2}{\omega_1} g_r I_1 I_2 - \alpha I_2 \quad (11)$$

where the loss terms involving α (the fiber loss per unit distance) are added phenomenologically. The Raman gain function g_r is expressed in a general way as

$$g_r = \frac{A}{\lambda_2} f(\lambda_1 - \lambda_2) \quad (12)$$

where A is a function of the material parameters and $f(\lambda_1 - \lambda_2)$ is a normalized line shape function, which is either derived from theory or experimentally measured (determined from Fig. 2, for example). With λ_2 expressed in μm , $A = 1.0 \times 10^{-11} \text{ cm} - \mu\text{m}/\text{W}$.²⁰ The solutions of Eqs. (10) and (11) are:

$$I_1(z) = \frac{\omega_1}{\omega_2} I_0 \exp(-\alpha z) \frac{\psi_r}{1 + \psi_r} \quad (13)$$

$$I_2(z) = I_0 \exp(-\alpha z) \frac{1}{1 + \psi_r} \quad (14)$$

In these equations, $I_0 = I_{20} + (\omega_1/\omega_2)I_{10}$, where I_{10} and I_{20} are the Stokes and pump intensities at the fiber input. The coupling parameter ψ_r assumes different forms, depending upon whether the input Stokes intensity I_{10} is present or not. If I_{10} is present, and if its magnitude is much greater than light from spontaneous Raman scattering, we have:

$$\psi_r = \frac{\omega_1}{\omega_2} \frac{I_{10}}{I_{20}} \exp(G_0) \quad (15)$$

When no Stokes input is present, the signal builds up from spontaneous Raman scattering, and the coupling parameter in this case becomes:

$$\psi_r = \frac{\hbar \omega_2 \Delta \omega_r}{4\sqrt{\pi}} \frac{1}{I_{20} A_{\text{eff}}} G_2^{-1/2} \exp(G_2) \quad (16)$$

with the gain parameters defined through $G_0 = g_r I_0 L_{\text{eff}}$ and $G_2 = g_r I_{20} L_{\text{eff}}$. The effective length of the fiber accounts for the reduction of Stokes and pump intensities as a result of loss, and is defined as

$$L_{\text{eff}} = \int_0^L \exp(-\alpha z) dz = \frac{1 - \exp(-\alpha L)}{\alpha} \quad (17)$$

The effective area of a single-mode fiber A_{eff} calculated through πr_0^2 , where r_0 is the mode field radius. For a multimode fiber, A_{eff} is usually taken as the core area, assuming that the power is uniformly distributed over the core. The power in the fiber is then $P_{1,2} = I_{1,2} A_{\text{eff}}$.

Two basic issues concerning SRS are of interest in fiber communication systems. First, pump-to-Stokes coupling provides a mechanism for crosstalk from short- to long-wavelength channels. This will occur most efficiently if the channel frequency spacing is in the vicinity of that associated with the maximum Raman gain. The Raman gain peak at approximately 500 cm^{-1} corresponds to a frequency spacing of 15 THz , meaning that operation at $1.55 \mu\text{m}$ produces a Stokes wave of about $1.67 \mu\text{m}$

wavelength. Two-channel operation at these wavelengths would lead to a maximum allowable signal level of about 50 mW.²¹ In WDM systems, within the 1.53- to 1.56- μm erbium-doped fiber amplifier window, channel spacings of 50 or 100 GHz are used. Raman gain is thus considerably reduced, but is still sufficient to cause appreciable crosstalk, which can lead to system penalties of between 1 and 3 dB depending on the number of channels.²² Second, and of more importance to single-wavelength systems, is the conversion to Stokes power from the original signal—a mechanism by which signal power can be depleted. A related problem is walkoff²³ occurring between the signal and Stokes pulses, since these will have different group delays. Walkoff is a means for aliasing to occur in digital transmission, unless the signal is filtered at the output. If pulses are of subpicosecond widths, additional complications arise due to the increased importance of SPM and XPM.²⁴ In any event, an upper limit must be placed on the signal power if significant conversion to Stokes power is to be avoided. In single-wavelength systems, where crosstalk is not an issue, pulse peak powers must be kept below about 500 mW to avoid significant SRS conversion.²⁵

A useful criterion is the so-called critical condition (or Raman threshold), defined as the condition under which the output Stokes and signal powers are equal. This occurs when $\psi_r = 1$, which, from Eq. (16), leads to $G_2 \approx 16$. SRS can also be weakened by taking advantage of the gain reduction that occurs as signal (pump) wavelengths increase, as shown in Eq. (12). For example, operation at 1.55 μm yields less SRS for a given signal power than operation at 1.3 μm .

Apart from the need to reduce SRS, the effect can be used to advantage in wavelength conversion and in amplification. Fiber Raman lasers have proven to be good sources of tunable radiation and operate at multiple Stokes wavelengths.²⁶ Specifically, a Stokes wave can serve as a pump to generate an additional (higher-order) Stokes wave at a longer wavelength.²⁷ Fiber Raman amplifiers are used routinely in long-haul systems.²⁸

10.4 STIMULATED BRILLOUIN SCATTERING

The stimulated Brillouin scattering (SBS) process involves the input of a single intense optical wave at frequency ω_2 , which initiates a copropagating acoustic wave at frequency ω_p . The acoustic wave is manifested as a traveling index grating in the fiber, which back-diffracts a portion of the original input. The backward (Stokes) wave is Doppler-shifted to a lower frequency ω_1 and is proportional to the phase conjugate of the input.²⁹ The backward wave is amplified as it propagates, with the gain increasing with increasing input (pump) power.

The beam interaction geometry is shown in Fig. 3. Usually, the Stokes wave builds up spontaneously, but can be inputted at the far end. The effect can be understood by considering a case in which counter-propagating Stokes and pump waves exist that together form a moving interference pattern whose velocity is proportional to the difference frequency $\omega_2 - \omega_1$. Coupling between the waves will occur via SBS when the interference pattern velocity is in the vicinity of the acoustic wave velocity v_p . It is the interference pattern that forms and reinforces the acoustic wave through electrostriction. With a single input, spontaneous scattering from numerous shock waves occurs, with preferential feedback from the acoustic wave that matches the condition just described. With the Stokes wave generated (although it is initially weak), the acoustic wave is reinforced, and so backscattering increases.

In terms of the wave vector magnitudes, the condition for phase matching is given by $k_p = k_1 + k_2$. Since the sound frequency is much less than those of the two optical waves, we can write $k_p \approx 2k_2$. Then, since $k_p = \omega_p/v_p$, it follows that $\omega_p \approx 2n\omega_2 v_p/c$, where n is the refractive index (assumed to

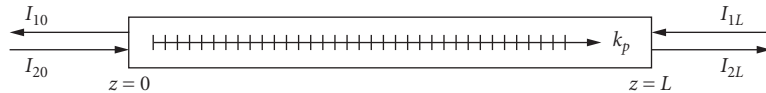


FIGURE 3 Beam geometry for stimulated Brillouin scattering in an optical fiber.

be the same value at both optical frequencies). The Brillouin frequency shift under phase-matched conditions thus becomes

$$\omega_2 - \omega_1 \approx 2n\omega_2 \frac{v_p}{c} \quad (18)$$

This yields a value of about 11 GHz, with $v_p \approx 6$ km/s in fused silica and $\lambda_2 = 1.55 \mu\text{m}$.

The process can be described by the nonlinear polarization produced by the product of complex fields, E_1 , E_2^* , and E_2 ; this yields a polarization at ω_1 that propagates with wavevector k_1 in the direction of the Stokes wave. Another polarization, describing back-coupling from Stokes to pump, involves the product $E_1 E_1^* E_2$. Substituting fields and polarizations into the wave equation yields the following coupled equations that describe the evolution of the optical intensities with distance (pp. 287–290 of Ref. 18):

$$\frac{dI_1}{dz} = -g_b I_1 I_2 + \alpha I_1 \quad (19)$$

$$\frac{dI_2}{dz} = -g_b I_1 I_2 - \alpha I_2 \quad (20)$$

where α is the linear loss coefficient. The Brillouin gain is given by

$$g_b = g_{b0} \left(1 + \frac{4(\omega_1 - \omega_{10})^2}{v_p^2 \alpha_p^2} \right)^{-1} \quad (21)$$

where ω_{10} is the Stokes frequency at precise phase matching, α_p is the loss coefficient for the acoustic wave, and the peak gain g_{b0} is a function of the material parameters. The Brillouin line width, defined as the full width at half-maximum of g_b , is $\Delta\omega_b = v_p \alpha_p$. In optical fibers, $\Delta f_b = \Delta\omega_b/2\pi$ is typically between 10 and 30 MHz and $g_{b0} = 4.5 \times 10^{-9}$ cm/W.²⁰ Signal bandwidths in high-data-rate communication systems greatly exceed the Brillouin line width, and so SBS is typically too weak to be considered a source of noise or signal depletion. This is to be compared to stimulated Raman scattering, which supports considerable gain over approximately 5 THz. Consequently, SRS is a much more serious problem in high-data-rate systems.

Using analysis methods similar to those employed in SRS, a critical condition (or threshold) can be defined for SBS, at which the backscattered power is equal to the input power:³⁰

$$\frac{\omega_1 k_B T \Delta\omega_b}{4\sqrt{\pi} \omega_p I_{20} A_{\text{eff}}} G_b^{-3/2} \exp(G_b) = 1 \quad (22)$$

where k_B is the Boltzmann's constant and T is the temperature in Kelvin. The gain parameter is:

$$G_b = g_b I_{20} L_{\text{eff}} \quad (23)$$

with L_{eff} as defined in Eq. (17). Equation (22) is approximately satisfied when $G_b \approx 21$.³⁰ In practice, the backscattered power will always be less than the input power, since pump depletion will occur. Nevertheless, this condition is used as a benchmark to determine the point at which SBS becomes excessive in a given system.³¹ In one study, it was found that $G_b \approx 21$ yields the pump power required to produce an SBS output that is at the level of Rayleigh back-scattering.³² Pump powers required to achieve threshold can be on the order of a few milliwatts for CW or narrowband signals, but these increase substantially for broadband signals.³³ Reduction of SBS is accomplished in practice by

lowering the input signal power (I_{20}) or by taking advantage of the reduction in g_b that occurs when signal bandwidths ($\Delta\omega$) exceed the Brillouin line width. Specifically, if $\Delta\omega \gg \Delta\omega_b$,

$$g_b(\Delta\omega) \approx g_b \frac{\Delta\omega_b}{\Delta\omega} \quad (24)$$

10.5 FOUR-WAVE MIXING

The term *four-wave mixing* in fibers is generally applied to wave coupling through the electronic nonlinearity in which at least two frequencies are involved and in which frequency conversion is occurring. The fact that the electronic nonlinearity is involved distinguishes four-wave mixing interactions from stimulated scattering processes because in the latter the medium was found to play an active role through the generation or absorption of optical phonons (in SRS) or acoustic phonons (in SBS). If the nonlinearity is electronic, bound electron distributions are modified according to the instantaneous optical field configurations. For example, with light at two frequencies present, electron positions can be modulated at the difference frequency, thus modulating the refractive index. Additional light will encounter the modulated index and can be up- or downshifted in frequency. In such cases, the medium plays a passive role in the interaction, as it does not absorb applied energy or release energy previously stored. The self- and cross-phase modulation processes also involve the electronic nonlinearity, but in those cases, power conversion between waves is not occurring—only phase modulation.

As an illustration of the process, consider the interaction of two strong waves at frequencies ω_1 and ω_2 , which mix to produce a downshifted (Stokes) wave at ω_3 and an upshifted (anti-Stokes) wave at ω_4 . The frequencies have equal spacing, that is, $\omega_1 - \omega_3 = \omega_2 - \omega_1 = \omega_4 - \omega_2$ (Fig. 4). All fields assume the real form:

$$\mathcal{E}_j = \frac{1}{2} E_{oj} \exp[i(\omega_j t - \beta_j z) + c.c.] \quad j = 1 - 4 \quad (25)$$

The nonlinear polarization will be proportional to $\mathcal{E}^3$, where $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 + \mathcal{E}_4$. With all fields copolarized, complex nonlinear polarizations at ω_3 and ω_4 appear that have the form:

$$P_{NL}^{\omega_3} - \frac{3}{4} \epsilon_0 \chi^{(3)} E_{01}^2 E_{02}^* \exp[i(2\omega_1 - \omega_2)t] \exp[-i(2\beta^{\omega_1} - \beta^{\omega_2})z] \quad (26)$$

$$P_{NL}^{\omega_4} - \frac{3}{4} \epsilon_0 \chi^{(3)} E_{02}^2 E_{01}^* \exp[i(2\omega_2 - \omega_1)t] \exp[-i(2\beta^{\omega_2} - \beta^{\omega_1})z] \quad (27)$$

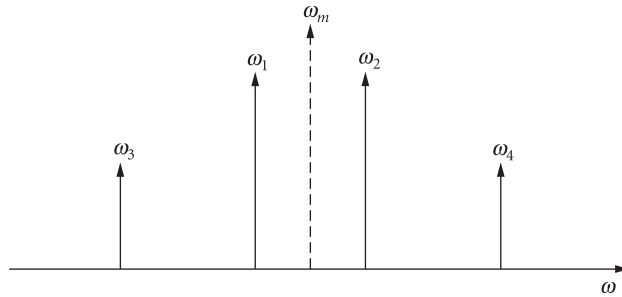


FIGURE 4 Frequency diagram for four-wave mixing, showing pump frequencies (ω_1 and ω_2) and sideband frequencies (ω_3 and ω_4).

where $\omega_3 = 2\omega_1 - \omega_2$, $\omega_4 = 2\omega_1 - \omega_3$, and $\chi^{(3)}$ is proportional to the nonlinear refractive index n_2' . The significance of these polarizations lies not only in the fact that waves at the sideband frequencies ω_3 and ω_4 can be generated, but that preexisting waves at those frequencies can experience gain in the presence of the two pump fields at ω_1 and ω_2 , thus forming a parametric amplifier. The sideband waves will contain the amplitude and phase information on the pumps, thus making this process an important crosstalk mechanism in multiwavelength communication systems. Under phase-matched conditions, the gain associated with FWM is more than twice the peak gain in SRS.³⁴

The wave equation, when solved in steady state, yields the output intensity at either one of the sideband frequencies.³⁵ For a medium of length L , having loss coefficient α , the sideband intensities are related to the pump intensities through

$$I^{\omega_3} \propto \left(\frac{n_2 L_{\text{eff}}}{\lambda_m} \right)^2 I^{\omega_2} (I^{\omega_1})^2 \eta \exp(-\alpha L) \quad (28)$$

$$I^{\omega_4} \propto \left(\frac{n_2 L_{\text{eff}}}{\lambda_m} \right)^2 I^{\omega_1} (I^{\omega_2})^2 \eta \exp(-\alpha L) \quad (29)$$

where L_{eff} is defined in Eq. (17), and where

$$\eta = \frac{\alpha^2}{\alpha^2 + \Delta\beta^2} \left(1 + \frac{4 \exp(-\alpha L) \sin^2(\Delta\beta L/2)}{(1 - \exp(-\alpha L))^2} \right) \quad (30)$$

Other FWM interactions can occur, involving products of intensities at three different frequencies rather than two as demonstrated here. In such cases, the output wave intensities are increased by a factor of 4 over those indicated in Eqs. (28) and (29).

One method of suppressing four-wave mixing in WDM systems includes the use of unequal channel spacing.³⁶ This ensures, for example, that $\omega_3 \neq 2\omega_1 + \omega_2$, where ω_1 , ω_2 , and ω_3 are assigned channel frequencies. More common methods involve phase-mismatching the process in some way. This is accomplished by increasing $\Delta\beta$, which has the effect of decreasing η in Eqs. (28) and (29). Note that in the low-loss limit, where $\alpha \rightarrow 0$, Eq. (30) reduces to $\eta = (\sin^2(\Delta\beta L/2)/(\Delta\beta L/2)^2)$. The $\Delta\beta$ expressions associated with wave generation at ω_3 and ω_4 are given by

$$\Delta\beta(\omega_3) = 2\beta^{\omega_1} - \beta^{\omega_2} - \beta^{\omega_3} \quad (31)$$

and

$$\Delta\beta(\omega_4) = 2\beta^{\omega_2} - \beta^{\omega_1} - \beta^{\omega_4} \quad (32)$$

It is possible to express Eqs. (31) and (32) in terms of known fiber parameters by using a Taylor series for the propagation constant, where the expansion is about frequency ω_m as indicated in Fig. 4, where $\omega_m = (\omega_2 + \omega_1)/2$

$$\beta \approx \beta_0 + (\omega - \omega_m)\beta_1 + \frac{1}{2}(\omega - \omega_m)^2\beta_2 + \frac{1}{6}(\omega - \omega_m)^3\beta_3 \quad (33)$$

In Eq. (33), β_1 , β_2 , and β_3 are, respectively, the first, second, and third derivatives of β with respect to ω , evaluated at ω_m . These in turn relate to the fiber dispersion parameter D (ps/nm · km) and its first derivative with respect to wavelength through $\beta_2 = -(\lambda_m^2/2\pi c)D(\lambda_m)$ and $\beta_3 = (\lambda_m^3/2\pi^2 c^2)[D(\lambda_m) + (\lambda_m/2)(dD/d\lambda)]_{\lambda_m}$ where $\lambda_m = 2\pi c/\omega_m$. Using these relations along with Eq. (33) in Eqs. (31) and (32) results in:

$$\Delta\beta(\omega_3, \omega_4) \approx 2\pi c \frac{\Delta\lambda^2}{\lambda_m^2} \left[D(\lambda_m) \pm \frac{\Delta\lambda}{2} \frac{dD}{d\lambda} \Big|_{\lambda_m} \right] \quad (34)$$

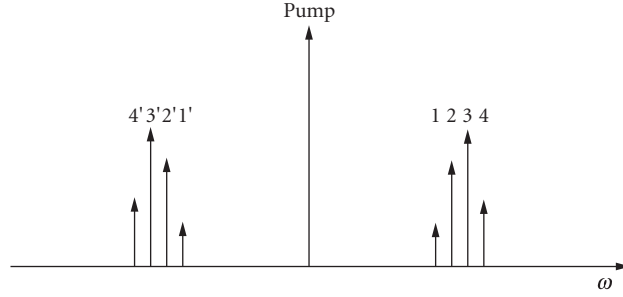


FIGURE 5 Frequency diagram for spectral inversion using four-wave mixing with a single pump frequency.

where the plus sign is used for $\Delta\beta(\omega_3)$, the minus sign is used for $\Delta\beta(\omega_1)$, and $\Delta\lambda = \lambda_1 - \lambda_2$. Phase matching is not completely described by Eq. (34), since cross-phase modulation plays a subtle role, as discussed in Ref. 16. Nevertheless, Eq. (34) does show that the retention of moderate values of dispersion D is a way to reduce FWM interactions that would occur, for example, in WDM systems. As such, modern commercial fiber intended for use in WDM applications will have values of D that are typically in the vicinity of 4 ps/nm · km.³⁷ With WDM operation in conventional dispersion-shifted fiber (with the dispersion zero near 1.55 μm), having a single channel at the zero dispersion wavelength can result in significant four-wave mixing.³⁸ Methods that were found to reduce four-wave mixing in such cases include the use of cross-polarized signals in dispersion-managed links³⁹ and operation within a longer-wavelength band near 1.6 μm⁴⁰ at which dispersion is appreciable and where gain-shifted fiber amplifiers are used.⁴¹

Examples of other cases involving four-wave mixing include single-wavelength systems, in which the effect has been successfully used in a demultiplexing technique for TDM signals.⁴² In another case, coupling through FWM can occur between a signal and broadband amplified spontaneous emission (ASE) in links containing erbium-doped fiber amplifiers.⁴³ As a result, the signal becomes spectrally broadened and exhibits phase noise from the ASE. The phase noise becomes manifested as amplitude noise under the action of dispersion, producing a form of modulation instability.

An interesting application of four-wave mixing is spectral inversion. Consider a case that involves the input of a strong single-frequency pump wave along with a relatively weak wave having a spectrum of finite width positioned on one side of the pump frequency. Four-wave mixing leads to the generation of a wave whose spectrum is the “mirror image” of that of the weak wave, in which the mirroring occurs about the pump frequency. Figure 5 depicts a representation of this, where four frequency components comprising a spectrum are shown along with their imaged counterparts. An important application of this is pulses that have experienced broadening with chirping after propagating through a length of fiber exhibiting linear group dispersion.⁴⁴ Inverting the spectrum of such a pulse using four-wave mixing has the effect of reversing the direction of the chirp (although the pulse center wavelength is displaced to a different value). When the spectrally inverted pulse is propagated through an additional length of fiber having the same dispersive characteristics, the pulse will compress to nearly its original input width. Compensation for nonlinear distortion has also been demonstrated using this method.⁴⁵

10.6 CONCLUSION

An overview of fiber nonlinear effects has been presented here in which emphasis is placed on the basic concepts, principles, and perspectives on communication systems. Space is not available to cover the more subtle details of each effect or the interrelations between effects that often occur. The

text by Agrawal¹⁶ is recommended for further in-depth study, which should be supplemented by the current literature. Nonlinear optics in fibers and in fiber communication systems comprises an area whose principles and implications are still not fully understood. It thus remains an important area of current research.

10.7 REFERENCES

1. L. F. Mollenauer and P. V. Mamyshev, "Massive Wavelength-Division Multiplexing with Solitons," *IEEE J. Quantum Electron.* **34**:2089–2102 (1998).
2. C. M. de Sterke, B. J. Eggleton, and P. A. Krug, "High-Intensity Pulse Propagation in Uniform Gratings and Grating Superstructures," *IEEE J. Lightwave Technol.* **15**:1494–1502 (1997).
3. L. Clark, A. A. Klein, and D. W. Peckham, "Impact of Fiber Selection and Nonlinear Behavior on Network Upgrade Strategies for Optically Amplified Long Interoffice Routes," *Proceedings of the 10th Annual National Fiber Optic Engineers Conference*, vol. 4, 1994.
4. Y. R. Shen, *The Principles of Nonlinear Optics*, Wiley-Interscience, New York, 1984, p. 28.
5. R. H. Stolen and H. W. K. Tom, "Self-Organized Phase-Matched Harmonic Generation in Optical Fibers," *Opt. Lett.* **12**:585–587 (1987).
6. R. W. Boyd, *Nonlinear Optics*, Academic Press, San Diego, 2008.
7. R. H. Stolen and C. Lin, "Self-Phase Modulation in Silica Optical Fibers," *Phys. Rev. A* **17**:1448–1453 (1978).
8. G. Bellotti, A. Bertaina, and S. Bigo, "Dependence of Self-Phase Modulation Impairments on Residual Dispersion in 10-Gb/s-Based Terrestrial Transmission Using Standard Fiber," *IEEE Photon. Technol. Lett.* **11**:824–826 (1999).
9. M. Stern, J. P. Heritage, R. N. Thurston, and S. Tu, "Self-Phase Modulation and Dispersion in High Data Rate Fiber Optic Transmission Systems," *IEEE J. Lightwave Technol.* **8**:1009–1015 (1990).
10. D. Marcuse and C. R. Menyuk, "Simulation of Single-Channel Optical Systems at 100 Gb/s," *IEEE J. Lightwave Technol.* **17**:564–569 (1999).
11. S. Reichel and R. Zengerle, "Effects of Nonlinear Dispersion in EDFA's on Optical Communication Systems," *IEEE J. Lightwave Technol.* **17**:1152–1157 (1999).
12. M. Karlsson, "Modulational Instability in Lossy Optical Fibers," *J. Opt. Soc. Am. B* **12**:2071–2077 (1995).
13. R. Hui, K. R. Demarest, and C. T. Allen, "Cross-Phase Modulation in Multispan WDM Optical Fiber Systems," *IEEE J. Lightwave Technol.* **17**:1018–1026 (1999).
14. S. Bigo, G. Billotti, and M. W. Chbat, "Investigation of Cross-Phase Modulation Limitation over Various Types of Fiber Infrastructures," *IEEE Photon. Technol. Lett.* **11**:605–607 (1999).
15. L. F. Mollenauer and J. P. Gordon, *Solitons in Optical Fibers*, Academic Press, Boston, 2006.
16. G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed., Academic Press, San Diego, 2006.
17. G. Herzberg, *Infra-Red and Raman Spectroscopy of Polyatomic Molecules*, Van Nostrand, New York, 1945, pp. 99–101.
18. J. A. Buck, *Fundamentals of Optical Fibers*, 2nd ed., Wiley-Interscience, Hoboken, 2004.
19. F. L. Galeener, J. C. Mikkelsen Jr., R. H. Geils, and W. J. Mosby, "The Relative Raman Cross Sections of Vitreous SiO₂, GeO₂, B₂O₃, and P₂O₅," *Appl. Phys. Lett.* **32**:34–36 (1978).
20. R. H. Stolen, "Nonlinear Properties of Optical Fibers," in S. E. Miller and A. G. Chynoweth (eds.), *Optical Fiber Telecommunications*, Academic Press, New York, 1979.
21. A. R. Chraplyvy, "Optical Power Limits in Multi-Channel Wavelength Division Multiplexed Systems due to Stimulated Raman Scattering," *Electron. Lett.* **20**:58–59 (1984).
22. F. Forghieri, R. W. Tkach, and A. R. Chraplyvy, "Effect of Modulation Statistics on Raman Crosstalk in WDM Systems," *IEEE Photon. Technol. Lett.* **7**:101–103 (1995).
23. R. H. Stolen and A. M. Johnson, "The Effect of Pulse Walkoff on Stimulated Raman Scattering in Fibers," *IEEE J. Quantum Electron.* **22**:2154–2160 (1986).
24. C. H. Headley III and G. P. Agrawal, "Unified Description of Ultrafast Stimulated Raman Scattering in Optical Fibers," *J. Opt. Soc. Am. B* **13**:2170–2177 (1996).

25. R. H. Stolen, J. P. Gordon, W. J. Tomlinson, and H. A. Haus, "Raman Response Function of Silica Core Fibers," *J. Opt. Soc. Am. B* **6**:1159–1166 (1988).
26. L. G. Cohen and C. Lin, "A Universal Fiber-Optic (UFO) Measurement System Based on a Near-IR Fiber Raman Laser," *IEEE J. Quantum Electron.* **14**:855–859 (1978).
27. K. X. Liu and E. Garmire, "Understanding the Formation of the SRS Stokes Spectrum in Fused Silica Fibers," *IEEE J. Quantum Electron.* **27**:1022–1030 (1991).
28. J. Bromage, "Raman Amplification for Fiber Communication Systems," *IEEE J. Lightwave Technol.* **22**:79–93 (2004).
29. A. Yariv, *Quantum Electronics*, 3d ed., Wiley, New York, 1989, pp. 513–516.
30. R. G. Smith, "Optical Power Handling Capacity of Low Loss Optical Fibers as Determined by Stimulated Raman and Brillouin Scattering," *Appl. Opt.* **11**:2489–2494 (1972).
31. A. R. Chraplyvy, "Limitations on Lightwave Communications Imposed by Optical Fiber Nonlinearities," *IEEE J. Lightwave Technol.* **8**:1548–1557 (1990).
32. X. P. Mao, R. W. Tkach, A. R. Chraplyvy, R. M. Jopson, and R. M. Derosier, "Stimulated Brillouin Threshold Dependence on Fiber Type and Uniformity," *IEEE Photon. Technol. Lett.* **4**:66–68 (1992).
33. C. Edge, M. J. Goodwin, and I. Bennion, "Investigation of Nonlinear Power Transmission Limits in Optical Fiber Devices," *Proc. IEEE* **134**:180–182 (1987).
34. R. H. Stolen, "Phase-Matched Stimulated Four-Photon Mixing in Silica-Fiber Waveguides," *IEEE J. Quantum Electron.* **11**:100–103 (1975).
35. R. W. Tkach, A. R. Chraplyvy, F. Forghieri, A. H. Gnauck, and R. M. Derosier, "Four-Photon Mixing and High-Speed WDM Systems," *IEEE J. Lightwave Technol.* **13**:841–849 (1995).
36. F. Forghieri, R. W. Tkach, and A. R. Chraplyvy, and D. Marcuse, "Reduction of Four-Wave Mixing Crosstalk in WDM Systems Using Unequally-Spaced Channels," *IEEE Photon. Technol. Lett.* **6**:754–756 (1994).
37. AT&T Network Systems data sheet 4694FS-Issue 2 LLC, "TrueWave Single Mode Optical Fiber Improved Transmission Capacity," December 1995.
38. D. Marcuse, A. R. Chraplyvy, and R. W. Tkach, "Effect of Fiber Nonlinearity on Long-Distance Transmission," *IEEE J. Lightwave Technol.* **9**:121–128 (1991).
39. E. A. Golovchenko, N. S. Bergano, and C. R. Davidson, "Four-Wave Mixing in Multispan Dispersion Managed Transmission Links," *IEEE Photon. Technol. Lett.* **10**:1481–1483 (1998).
40. M. Jinno et al, "1580 nm Band, Equally-Spaced 8×10 Gb/s WDM Channel Transmission Over 360 km (3×120 km) of Dispersion-Shifted Fiber Avoiding FWM Impairment," *IEEE Photon. Technol. Lett.* **10**:454–456 (1998).
41. H. Ono, M. Yamada, and Y. Ohishi, "Gain-Flattened Er^{3+} -Fiber Amplifier for A WDM Signal in the 1.57–1.60 μm Wavelength Region," *IEEE Photon. Technol. Lett.* **9**:596–598 (1997).
42. P. O. Hedekvist, M. Karlsson, and P. A. Andrekson, "Fiber Four-Wave Mixing Demultiplexing with Inherent Parametric Amplification," *IEEE J. Lightwave Technol.* **15**:2051–2058 (1997).
43. R. Hui, M. O'Sullivan, A. Robinson, and M. Taylor, "Modulation Instability and Its Impact on Multispan Optical Amplified IMDD Systems: Theory and Experiments," *IEEE J. Lightwave Technol.* **15**:1071–1082 (1997).
44. A. H. Gnauck, R. M. Jopson, and R. M. Derosier, "10 Gb/s 360 km Transmission over Dispersive Fiber Using Midsystem Spectral Inversion," *IEEE Photon. Technol. Lett.* **5**:663–666 (1993).
45. A. H. Gnauck, R. M. Jopson, and R. M. Derosier, "Compensating the Compensator: A Demonstration of Nonlinearity Cancellation in a WDM System," *IEEE Photon. Technol. Lett.* **7**:582–584 (1995).

This page intentionally left blank

Philip St. J. Russell and Greg J. Pearce

Max-Planck Institute for the Science of Light
Erlangen, Germany

11.1 GLOSSARY

A_i	nonlinear effective area of subregion i
c	velocity of light in vacuum
d	hole diameter
D	$\frac{\partial}{\partial \lambda} \frac{1}{v_g} = \frac{\partial^2 \beta_m}{\partial \lambda \partial \omega}$ the group velocity dispersion of mode m in engineering units (ps/nm · km)
k	vacuum wavevector $2\pi/\lambda = \omega/c$
n_2^i	nonlinear refractive index of subregion i
$n_{\max}$	maximum index supported by the PCF cladding (fundamental space-filling mode)
n_z	z -component of refractive index
n_q^∞	$\sqrt{\sum_i n_i^2 A_i / \sum_i A_i}$ the area-averaged refractive index of an arbitrary region q of a microstructured fiber, where n_i and A_i are respectively the refractive indices and total area of subregion i
V	$k\rho\sqrt{n_{co}^2 - n_{cl}^2}$ the normalized frequency for a step-index fiber
V_{gen}	$k\Lambda\sqrt{n_1^2 - n_2^2}$ a generalized form of V for a structure made from two materials of index n_1 and n_2
β	component of wavevector along the fiber axis
β_m	axial wavevector of guided mode
$\beta_{\max}$	maximum possible axial component of wavevector in the PCF cladding
Δ	$(n_{co} - n_{cl})/n_{cl}$, where n_{co} and n_{cl} are, respectively, the core and cladding refractive indices of a conventional step-index fiber
Δ_∞	$(n_{co}^\infty - n_{cl}^\infty)/n_{cl}^\infty$, where n_{co}^∞ and n_{cl}^∞ are the refractive indices of core and cladding in the long wavelength limit
γ	$W^{-1}\text{km}^{-1}$ nonlinear coefficient of an optical fiber
$\epsilon(\mathbf{r}_T)$	relative dielectric constant of a PCF as a function of transverse position $\mathbf{r}_T = (x, y)$
λ	vacuum wavelength
λ_{eff}^i	$2\pi/\sqrt{k^2 n_i^2 - \beta^2}$ the effective transverse wavelength in subregion i

- Λ interhole spacing, period, or pitch
- ρ core radius
- ω angular frequency of light

11.2 INTRODUCTION

Photonic crystal fibers (PCFs)—fibers with a periodic transverse microstructure—have been in practical existence as low loss waveguides since early 1996.^{1–4} The initial demonstration took 4 years of technological development, and since then the fabrication techniques have become more and more sophisticated. It is now possible to manufacture the microstructure in air-glass PCF to accuracies of 10 nm on the scale of 1 μm , which allows remarkable control of key optical properties such as dispersion, birefringence, nonlinearity, and the position and width of the photonic bandgaps (PBGs) in the periodic “photonic crystal” cladding. PCF has in this way extended the range of possibilities in optical fibers, both by improving well-established properties and introducing new features such as low loss guidance in a hollow core.

Standard single-mode telecommunications fiber (SMF), with a normalized core-cladding refractive index difference Δ approximately 0.4 percent, a core radius of ρ approximately 4.5 μm , and of course a very high optical clarity (better than 5 km/dB at 1550 nm), is actually quite limiting for many applications. Two major factors contribute to this. The first is the smallness of Δ , which causes bend loss (0.5 dB at 1550 nm in Corning SMF-28 for one turn around a mandrel 32 mm in diameter¹⁰) and limits the degree to which group velocity dispersion and birefringence can be manipulated. Although much higher values of Δ can be attained (modified chemical vapor deposition yields an index difference of 0.00146 per mol % GeO_2 , up to a maximum $\Delta \sim 10\%$ for 100 mol %¹¹), the single-mode core radius becomes very small and the attenuation rises through increased absorption and Rayleigh scattering. The second factor is the reliance on total internal reflection (TIR), so that guidance in a hollow core is impossible, however useful it would be in fields requiring elimination of glass-related nonlinearities or enhancement of laser interactions with dilute or gaseous media. PCF has made it possible to overcome these limitations, and as a result many new applications of optical fibers are emerging.

Outline of Chapter

In the next section a brief history of PCF is given, and in Sec. 11.4 fabrication techniques are reviewed. Numerical modeling and analysis are covered in Sec. 11.5 and the optical properties of the periodic photonic crystal cladding in Sec. 11.6. The characteristics of guidance are discussed in Sec. 11.7. In Sec. 11.8 the nonlinear characteristics of guidance are reviewed and intrafiber devices (including cleaving and splicing) are discussed in Sec. 11.9. Brief conclusions, including a list of applications, are drawn in Sec. 11.10.

11.3 BRIEF HISTORY

The original motivation for developing PCF was the creation of a new kind of dielectric waveguide—one that guides light by means of a two-dimensional PBG. In 1991, the idea that the well-known “stop-bands” in multiply periodic structures (for a review see Ref. 12) could be extended to eliminate all photonic states¹³ was leading to attempts worldwide to fabricate three-dimensional PBG materials. At that time the received wisdom was that the refractive index difference needed to create a PBG

in two dimensions was large—of order 2.2:1. It was not widely recognized that the refractive index difference requirements for PBG formation in two dimensions are greatly relaxed if, as in a fiber, propagation is predominantly along the third axis—the direction of invariance.

Photonic Crystal Fibers

The original 1991 idea, then, was to trap light in a hollow core by means of a two-dimensional “photonic crystal” of microscopic air capillaries running along the entire length of a glass fiber.¹⁴ Appropriately designed, this array would support a PBG for incidence from air, preventing the escape of light from a hollow core into the photonic crystal cladding and avoiding the need for TIR.

The first 4 years of work on understanding and fabricating PCF were a journey of exploration. The task of solving Maxwell’s equations numerically made good progress, culminating in a 1995 paper that showed that photonic bandgaps did indeed exist in two-dimensional silica-air structures for “conical” incidence from vacuum—this being an essential prerequisite for hollow-core guidance.¹⁵ Developing a suitable fabrication technique took rather longer. After 4 years of trying different approaches, the first successful silica-air PCF structure was made in late 1995 by stacking 217 silica capillaries (8 layers outside the central capillary), specially machined with hexagonal outer and a circular inner cross sections. The diameter-to-pitch ratio d/Λ of the holes in the final stack was approximately 0.2, which theory showed was too small for PBG guidance in a hollow core, so it was decided to make a PCF with a solid central core surrounded by 216 air channels (Fig. 1a).^{2,5,6}

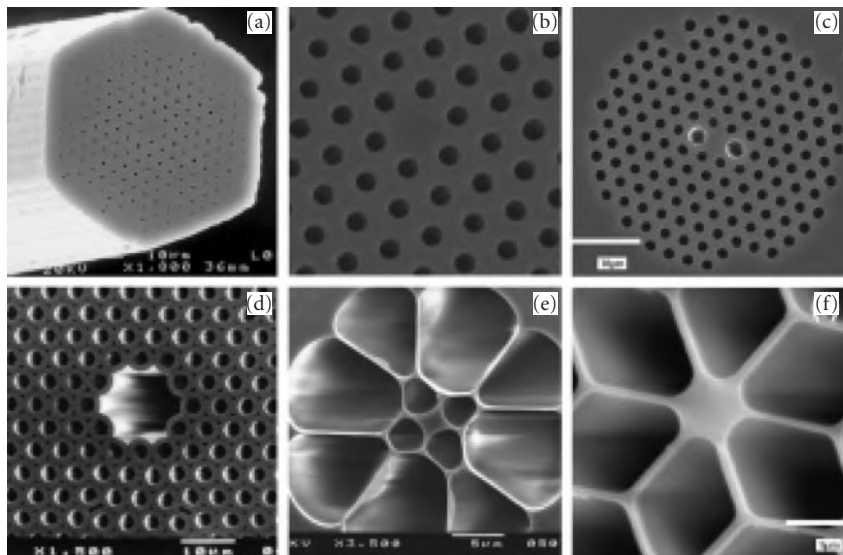


FIGURE 1 Selection of scanning electron micrographs of PCF structures. (a) The first working PCF—the solid glass core is surrounded by a triangular array of 300 nm diameter air channels, spaced 2.3 μm apart;^{5,6} (b) detail of a low loss solid-core PCF (interhole spacing $\sim 2 \mu\text{m}$); (c) birefringent PCF (interhole spacing 2.5 μm); (d) the first hollow-core PCF (core diameter $\sim 10 \mu\text{m}$);⁷ (e) a PCF extruded from Schott SF6 glass with a core approximately 2 μm in diameter;⁸ and (f) PCF with very small core (diameter 800 nm) and zero GVD wavelength 560 nm.⁹

This led to the discovery of endlessly single-mode PCF, which, if it guides at all, only supports the fundamental guided mode.¹⁶ The success of these initial experiments led rapidly to a whole series of new types of PCF—large mode area,¹⁷ dispersion-controlled,^{9,18} hollow core,⁷ birefringent,¹⁹ and multicore.²⁰

These initial breakthroughs led quickly to applications, perhaps the most celebrated being the report in 2000 of supercontinuum generation from unamplified Ti:sapphire fs laser pulses in a PCF with a core small enough to give zero dispersion at 800 nm wavelength (subsection “Supercontinuum Generation” in Sec. 11.8).²¹

Bragg Fibers

In the late 1960s and early 1970s, theoretical proposals were made for another kind of fiber with a periodically structured cross section.^{22,23} This was a cylindrical “Bragg” fiber that confines light within an annular array of rings of high and low refractive index arranged concentrically around a central core. A group in France has made a solid-core version of this structure using modified chemical vapor deposition.²⁴ Employing a combination of polymer and chalcogenide glass, researchers in the United States have realized a hollow-core version of a similar structure,²⁵ reporting 1 dB/m loss at 10 μm wavelength (the losses at telecom wavelengths are as yet unspecified). This structure guides light in the TE_{01} mode, used in microwave telecommunications because of its ultralow loss; the field moves away from the attenuating waveguide walls as the frequency increases, resulting in very low losses, although the guide must be kept very straight to avoid the fields entering the cladding and experiencing high absorption.

11.4 FABRICATION TECHNIQUES

Photonic crystal fiber (PCF) structures are currently produced in many laboratories worldwide using a variety of different techniques (see Fig. 2 for schematic drawings of some example structures). The first stage is to produce a “preform”—a macroscopic version of the planned microstructure in the drawn PCF. There are many ways to do this, including stacking of capillaries and rods,⁵ extrusion,^{8,26–28} sol-gel casting,²⁹ injection molding and drilling. The materials used range from silica to compound glasses, chalcogenide glasses, and polymers.³⁰

The most widely used technique is stacking of circular capillaries (Fig. 3). Typically, meter-length capillaries with an outer diameter of approximately 1 mm are drawn from a starting tube of high-purity synthetic silica with a diameter of approximately 20 mm. The inner/outer diameter of the starting tube, which typically lies in the range from 0.3 up to beyond 0.9, largely determines the d/Λ value in the drawn fiber. The uniformity in diameter and circularity of the capillaries must be controlled to at least 1 percent of the diameter. They are stacked horizontally, in a suitably shaped jig, to form the desired crystalline arrangement. The stack is bound with wire before being inserted into a jacketing tube, and the whole assembly is then mounted in the preform feed unit for drawing down to fiber. Judicious use of pressure and vacuum during the draw allows some limited control over the final structural parameters, for example the d/Λ value.

Extrusion offers an alternative route to making PCF, or the starting tubes, from bulk glass; it permits formation of structures that are not readily made by stacking. While not suitable for silica (no die material has been identified that can withstand the $\sim 2000^\circ\text{C}$ processing temperatures without contaminating the glass), extrusion is useful for making PCF from compound silica glasses, tellurites, chalcogenides, and polymers—materials that melt at lower temperatures. Figure 1e shows the cross section of a fiber extruded, through a metal die, from a commercially available glass (Schott SF6).⁸ PCF has also been extruded from tellurite glass, which has excellent IR transparency out to beyond 4 μm , although the reported fiber losses (a few dB/m) are as yet rather high.^{27,31–33} Polymer PCFs, first developed in Sydney, have been successfully made using many different approaches, for example extrusion, casting, molding, and drilling.³⁰

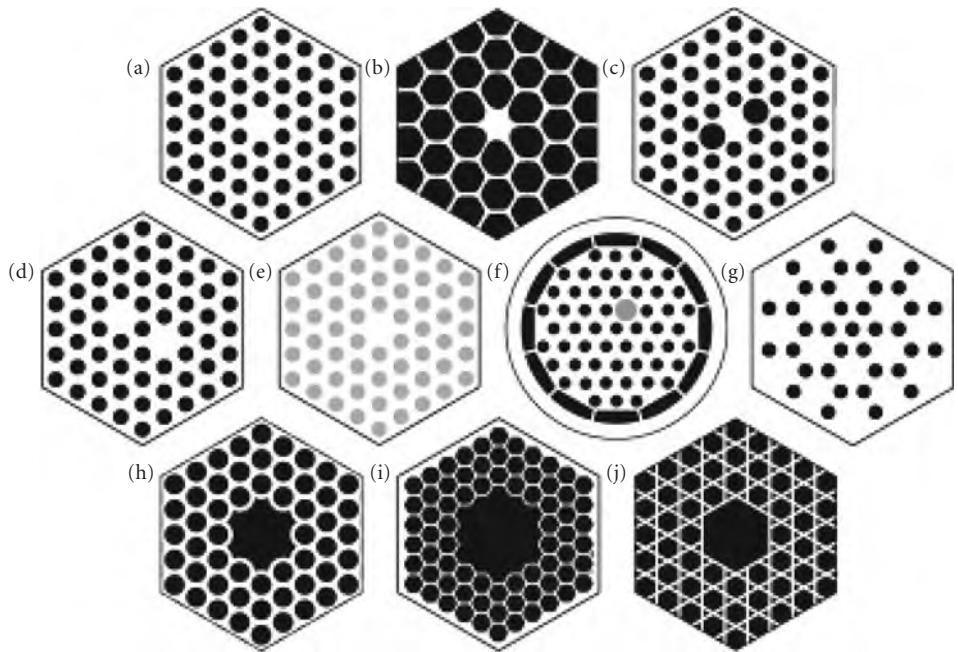


FIGURE 2 Representative sketches of different types of PCF. The black regions are hollow, the white regions are pure glass, and the gray regions doped glass. (a) Endlessly single-mode solid core; (b) highly nonlinear (high air-filling fraction, small core, characteristically distorted holes next to the core); (c) birefringent; (d) dual-core; (e) all-solid glass with raised-index doped glass strands (colored gray) in the cladding; (f) double-clad PCF with off-set doped lasing core and high numerical aperture inner cladding for pumping (the photonic crystal cladding is held in place by thin webs of glass); (g) “carbon-ring” array of holes for PBG guidance in core with extra hole; (h) seven-cell hollow core; (i) 19-cell hollow core with high air-filling fraction (in a real fiber, surface tension smoothes out the bumps on the core surround); and (j) hollow-core with Kagomé lattice in the cladding.

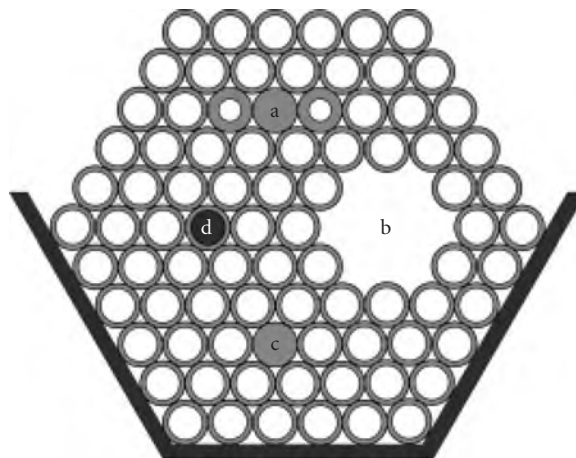


FIGURE 3 Preform stack containing (a) birefringent solid core; (b) seven-cell hollow core; (c) solid isotropic core; and (d) doped core. The capillary diameters are approximately 1 mm—large enough to ensure that they remain stiff for stacking.

Design Approach

The successful design of a PCF for a particular application is not simply a matter of using numerical modeling (see next section) to calculate the parameters of a structure that yields the required performance. This is because the fiber drawing process is not lithographic, but introduces its own highly reproducible types of distortion through the effects of viscous flow, surface tension, and pressure. As a result, even if the initial preform stack precisely mimics the theoretically required structure, several modeling and fabrication iterations are usually needed before a successful design can be reached.

11.5 MODELING AND ANALYSIS

The complex structure of PCF—in particular the large refractive index difference between glass and air—makes its electromagnetic analysis challenging. Maxwell's equations must usually be solved numerically, using one of a number of specially developed techniques.^{15,34–38} Although standard optical fiber analyses and number of approximate models are occasionally helpful, these are only useful as rough guidelines to the exact behavior unless checked against accurate numerical solutions.

Maxwell's Equations

In most practical cases, a set of equal frequency modes is more useful than a set of modes of different frequency sharing the same value of axial wavevector component β . It is therefore convenient to arrange Maxwell's equations with β^2 as eigenvalue

$$(\nabla^2 + k^2 \epsilon(\mathbf{r}_T) + [\nabla \ln \epsilon(\mathbf{r}_T)] \wedge \nabla \wedge) \mathbf{H}_T = \beta^2 \mathbf{H}_T \quad (1)$$

where all the field vectors are taken in the form $\mathbf{Q} = \mathbf{Q}_T(\mathbf{r}_T) e^{-j\beta z}$, $\epsilon_T(\mathbf{r}_T)$ is the dielectric constant, $\mathbf{r}_T = (x, y)$ is position in the transverse plane, and $k = \omega/c$ is the vacuum wavevector. This form allows material dispersion to be easily included, something which is not possible if the equations are set up with k^2 as eigenvalue. Written out explicitly in cartesian coordinates Eq. (1) yields two equations relating h_x and h_y

$$\begin{aligned} \frac{\partial^2 h_x}{\partial y^2} + \frac{\partial^2 h_x}{\partial x^2} - \frac{\partial \ln \epsilon}{\partial y} \left(\frac{\partial h_x}{\partial y} - \frac{\partial h_y}{\partial x} \right) + (\epsilon k^2 - \beta^2) h_x &= 0 \\ \frac{\partial^2 h_y}{\partial x^2} + \frac{\partial^2 h_y}{\partial y^2} + \frac{\partial \ln \epsilon}{\partial x} \left(\frac{\partial h_x}{\partial y} - \frac{\partial h_y}{\partial x} \right) + (\epsilon k^2 - \beta^2) h_y &= 0 \end{aligned} \quad (2)$$

and a third differential equation relating h_x , h_y , and h_z , which is however not required to solve Eq. (2).

Scalar Approximation

In the paraxial scalar approximation the second term inside the operator in Eq. (1), which gives rise to the middle terms in Eq. (2) that couple between the vector components of the field, can be neglected, yielding a scalar wave equation

$$\nabla^2 \mathbf{H}_T + [k^2 \epsilon(\mathbf{r}_T) - \beta^2] \mathbf{H}_T = 0 \quad (3)$$

This leads to a scaling law, similar to the one used in standard analyses of step-index fiber,³⁹ that can be used to parameterize the fields.⁴⁰ Defining Λ as the interhole spacing and n_1 and n_2 the refractive indices of the two materials used to construct a particular geometrical shape of photonic crystal, the mathematical forms of the fields and the dispersion relations are identical provided the generalized V -parameter

$$V_{\text{gen}} = k\Lambda\sqrt{n_1^2 - n_2^2} \quad (4)$$

is held constant. This has the interesting (though in the limit not exactly practical) consequence that bandgaps can exist for vanishingly small index differences, provided the structure is made sufficiently large (see subsection “All-Solid Structures” in Sec. 11.7).

Numerical Techniques

A common technique for solving Eq. (1) employs a Fourier expansion to create a basis set of plane waves for the fields, which reduces the problem to the inversion of a matrix equation, suitable for numerical computation.³⁷ Such an implicitly periodic approach is especially useful for the study of the intrinsic properties of PCF claddings. However, in contrast to versions of Maxwell’s equations with k^2 as eigenvalue,⁴¹ Eq. (1) is non-Hermitian, which means that standard matrix inversion methods for Hermitian problems cannot straightforwardly be applied. An efficient iterative scheme can, however, be used to calculate the inverse of the operator by means of fast Fourier transform steps. This method is useful for accurately finding the modes guided in a solid-core PCF, which are located at the upper edge of the eigenvalue spectrum of the inverted operator. In hollow-core PCF, however (or other fibers relying on a cladding bandgap to confine the light), the modes of interest lie in the interior of the eigenvalue spectrum. A simple transformation can, however, be used to move the desired interior eigenvalues to the edge of the spectrum, greatly speeding up the calculations and allowing as many as a million basis waves to be incorporated.^{42,43} To treat PCFs with a central guiding core in an otherwise periodic lattice, a supercell is constructed, its dimensions being large enough so that, once tiled, the guided modes in adjacent cores do not significantly interact.

The choice of a suitable numerical method often depends on fiber geometry, as some methods can exploit symmetries or regularity of structure to increase efficiency. Other considerations are whether material dispersion is significant (more easily included in fixed-frequency methods), and whether leakage losses or a treatment of leaky modes (requiring suitable boundaries on the computational domain) are desired. If the PCF structure consists purely of circular holes, for example, the multipole or Rayleigh method is a particularly fast and efficient method.^{34,35} It uses Mie theory to evaluate the scattering of the field incident on each hole. Other numerical techniques include expanding the field in terms of Hermite-Gaussian functions,^{38,44} the use of finite-difference time-domain (FDTD) analysis (a simple and versatile tool for exploring waveguide geometries⁴⁵) or finite-difference method in the frequency domain,⁴⁶ and the finite-element approach.⁴⁷ Yet another approach is a source-model technique which uses two sets of fictitious elementary sources to approximate the fields inside and outside circular cylinders.⁴⁸

11.6 CHARACTERISTICS OF PHOTONIC CRYSTAL CLADDING

The simplest photonic crystal cladding is a biaxially periodic, defect-free, composite material with its own well-defined dispersion and band structure. These properties determine the behavior of the guided modes that form at cores (or “structural defects” in the jargon of photonic crystals).

A convenient graphical tool is the propagation diagram—a map of the ranges of frequency and axial wavevector component β where light is evanescent in all transverse directions regardless of its polarization state (Fig. 4).¹⁵ The vertical axis is the normalized frequency $k\Lambda$, and the horizontal axis is the normalized axial wavevector component $\beta\Lambda$. Light is unconditionally cutoff from propagating (due to either TIR or a PBG) in the black regions.

In any subregion of isotropic material (glass or air) at fixed optical frequency, the maximum possible value of $\beta\Lambda$ is given by $k\Lambda n$, where n is the refractive index (at that frequency) of the region under consideration. For $\beta < kn$ light is free to propagate, for $\beta > kn$ it is evanescent and at $\beta = kn$ the critical angle is reached—denoting the onset of TIR for light incident from a medium of index larger than n .

The slanted guidelines (Fig. 4) denote the transitions from propagation to evanescence for air, the photonic crystal, and glass. At fixed optical frequency for $\beta < k$, light propagates freely in every subregion of the structure. For $k < \beta < kn_g$ (n_g is the index of the glass), light propagates in the glass substrands and is evanescent in the hollow regions. Under these conditions the “tight binding” approximation holds, and the structure may be viewed as an array of coupled glass waveguides.

The photonic bandgap “fingers” in Fig. 4 are most conveniently investigated by plotting the photonic density of states (DOS),⁴³ which shows graphically the density of allowed modes in the PCF cladding relative to vacuum (Fig. 5). Regions of zero DOS are photonic bandgaps, and by plotting a quantity such as $(\beta - k)\Lambda$ it is possible to see clearly how far a photonic bandgap extends below the light line.

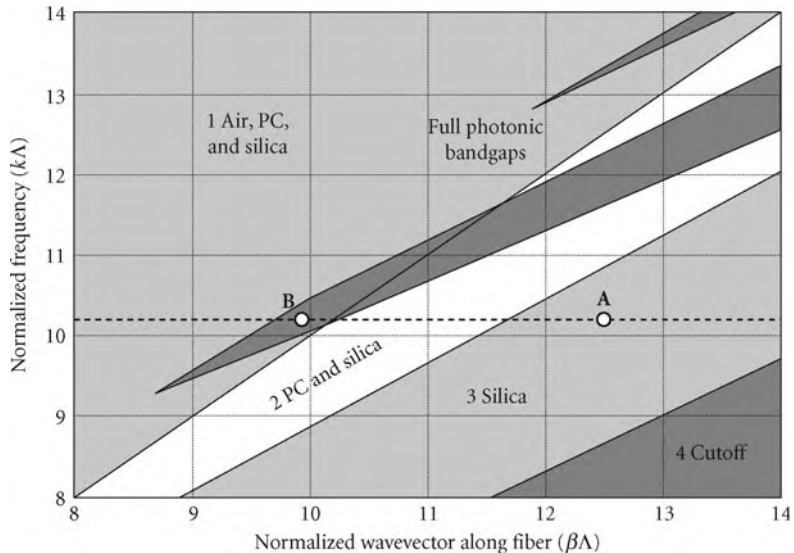


FIGURE 4 Propagation diagram for a triangular array of circular air holes (radius $\rho = 0.47\Lambda$) in silica glass, giving an air-filling fraction of 80 percent. Note the different regions where light is (4) cutoff completely (dark), (3) able to propagate only in silica glass (light gray), (2) able to propagate also in the photonic crystal cladding (white), and (1) able to propagate in all regions (light gray). Guidance by total internal reflection in a silica core is possible at point A. The “fingers” indicate the positions of full two-dimensional photonic bandgaps, which can be used to guide light in air at positions such as B where a photonic bandgap crosses the light line $k = \beta$.

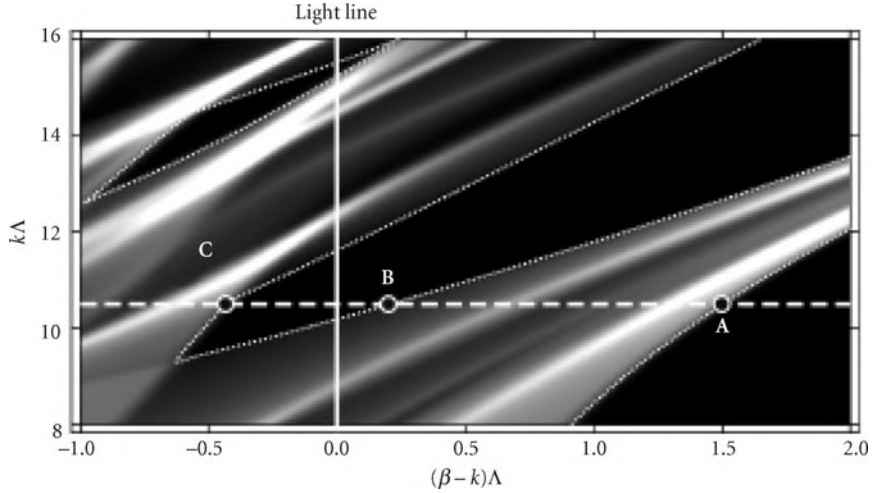


FIGURE 5 Photonic density of states (DOS) for the fiber structure described in Fig. 4, where black regions show zero DOS and lighter regions show higher DOS. The edges of full two-dimensional photonic bandgaps and the band edge of the fundamental space-filling mode are highlighted with thin dotted white lines. The vertical white line is the light line, and the labeled points mark band edges at the frequency of the thick dashed white line, discussed in the section “Maximum Refractive Index and Band Edges.”

Maximum Refractive Index and Band Edges

The maximum axial refractive index $n_{\max} = \beta_{\max}/k$ in the photonic crystal cladding lies in the range $k < \beta < k n_g$ as expected of a composite glass-air material. This value coincides with the z -pointing “peaks” of the dispersion surfaces in reciprocal space, where multiple values of transverse wavevector are allowed, one in each tiled Brillouin zone. For a constant value of β slightly smaller than $\beta_{\max}$, these wavevectors lie on small approximately circular loci, with a transverse component of group velocity that points normal to the circles in the direction of increasing frequency. Thus, light can travel in *all* directions in the transverse plane, even though its wavevectors are restricted to values lying on the circular loci. The real-space distribution of this field is shown in Fig. 6, together with the fields at two other band edges.

The maximum axial refractive index $n_{\max}$ depends strongly on frequency, even though neither the air nor the glass are assumed dispersive in the analysis; microstructuring itself creates dispersion, through a balance between transverse energy storage and energy flow that is highly dependent upon frequency.

By averaging the square of the refractive index in the photonic crystal cladding it is simple to show that

$$n_{\max} \rightarrow \sqrt{(1-F)n_g^2 - F n_a^2} \quad k\Lambda \rightarrow 0 \quad (5)$$

in the long-wavelength limit for a scalar approximation, where F is the air-filling fraction and n_a is the index in the holes (which we take to be 1 for the rest of this subsection).

As the wavelength of the light falls, the optical fields are better able to distinguish between the glass regions and the air. The light piles up more and more in the glass, causing the effective $n_{\max}$ “seen” by

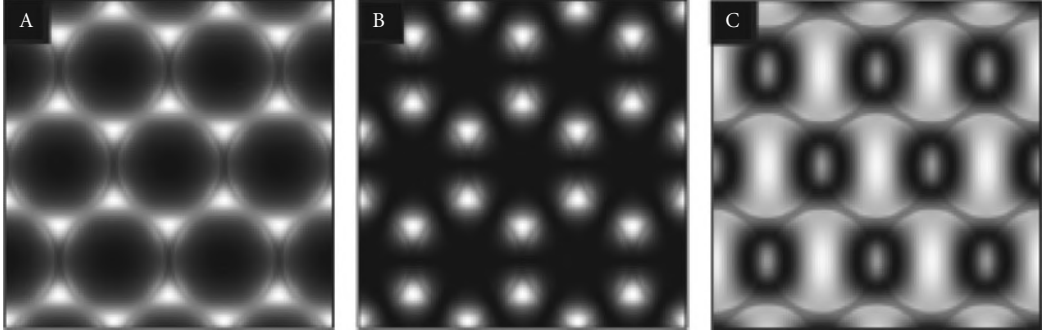


FIGURE 6 Plots showing the magnitude of the axial Poynting vector at band edges A, B, and C as shown in Fig. 5. White regions have large Poynting vector magnitude. A is the fundamental space-filling mode, for which the field amplitudes are in phase between adjacent unit cells. In B (the “dielectric” edge) the field amplitudes change sign between antinodes, and in C (the “air” edge) the central lobe has the opposite sign from the six surrounding lobes.

it to change. In the limit of small wavelength $k\Lambda \rightarrow \infty$ light is strongly excluded from the air holes by TIR, and the field profile “freezes” into a shape that is independent of wavelength. The variation of $n_{\max}$ with frequency may be estimated by expanding fields centered on the air holes in terms of Bessel functions and applying symmetry.¹⁶ Defining the normalized parameters

$$u = \Lambda \sqrt{k^2 n_g^2 - \beta^2} \quad v = k\Lambda \sqrt{n_g^2 - 1} \quad (6)$$

the analysis yields the polynomial fit (see Sec. 11.11):

$$u(v) = (0.00151 + 2.62v^{-1} + 0.0155v - 0.000402v^2 + 3.63 \times 10^{-6}v^3)^{-1} \quad (7)$$

for $d/\Lambda = 0.4$ and $n_g = 1.444$. This polynomial is accurate to better than 1 percent in the range $0 < v < 50$. The resulting expression for $n_{\max}$ is plotted in Fig. 7 against the parameter v .

Transverse Effective Wavelength

The transverse effective wavelength in the i th material is defined as follows:

$$\lambda_{\text{eff}}^i = \frac{2\pi}{\sqrt{k^2 n_i^2 - \beta^2}} \quad (8)$$

where n_i is its refractive index. This wavelength can be many times the vacuum value, tending to infinity at the critical angle $\beta \rightarrow kn_i$, and being imaginary when $\beta > kn_i$. It is a measure of whether or not the light is likely to be resonant within a particular feature of the structure, for example, a hole or a strand of glass, and defines PCF as a wavelength-scale structure.

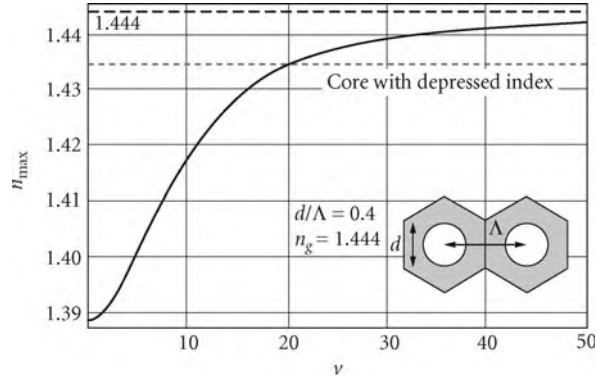


FIGURE 7 Maximum axial refractive index in the photonic crystal cladding as a function of the normalized frequency parameter ν for $d/\lambda = 0.4$ and $n_g = 1.444$. For this filling fraction of air (14.5%), the value at long wavelength ($\nu \rightarrow 0$) is $n_{\max} = 1.388$, in agreement with Eq. (5). The horizontal dashed gray line represents the case when the core is replaced with a glass of refractive index $n_{\text{co}} = 1.435$ (below that of silica), when guidance ceases for $\nu > 20$.

Photonic Bandgaps

Full two-dimensional PBGs exist in the black finger-shaped regions on Fig. 4. Some of these extend into the region $\beta < k$ where light is free to propagate in vacuum, confirming the feasibility of trapping light within a hollow core.

The bandgap edges coincide with points where resonances in the cladding unit cells switch on and off, that is, the eigenvalues of the unitary inter-unit-cell field transfer matrices change from $\exp(\pm j\phi)$ (propagation) to $\exp(\pm \gamma)$ (evanescence). At these transitions, depending on the band edge, the light is to a greater or lesser degree preferentially redistributed into the low or high index subregions. For example, at fixed optical frequency and small values of β , leaky modes peaking in the low index channels form a pass-band that terminates when the standing wave pattern has 100 percent visibility (Fig. 6c). For the high index strands (Fig. 6a and b) on the other hand, the band of real states is bounded by a lower value of β where the field amplitude changes sign between selected pairs of adjacent strands (depending on the lattice geometry), and an upper bound where the field amplitude does not change sign between the strands (this field distribution yields $n_{\max}$).

11.7 LINEAR CHARACTERISTICS OF GUIDANCE

In SMF, guided modes form within the range of axial refractive indices $n_{\text{cl}} < n_z < n_{\text{co}}$, when light is evanescent in the cladding ($n_z = \beta/k$; core and cladding indices are n_{co} and n_{cl}). In PCF, three distinct guidance mechanisms exist: a modified form of TIR,^{16,49} photonic bandgap guidance^{7,50} and a low leakage mechanism based on a Kagomé cladding structure.^{51,52} In the following subsections we explore the role of resonance and antiresonance, and discuss chromatic dispersion, attenuation mechanisms, and guidance in cores with refractive indices raised and lowered relative to the “mean” cladding value.

Resonance and Antiresonance

It is helpful to view the guided modes as being confined (or not confined) by resonance and antiresonance in the unit cells of the cladding crystal. If the core mode finds no states in the cladding with which it is phase-matched, light cannot leak out. This is a familiar picture in many areas of photonics. What is perhaps not so familiar is the use of the concept in two dimensions, where a repeating unit is tiled to form a photonic crystal cladding. This allows the construction of an intuitive picture of “cages,” “bars,” and “windows” for light and actually leads to a blurring of the distinction between guidance by modified TIR and photonic bandgap effects.

Positive Core-Cladding Index Difference

This type of PCF may be defined as one where the mean cladding refractive index in the long wavelength limit, $k \rightarrow 0$, [Eq. (5)] is lower than the core index (in the same limit). Under the correct conditions (high air-filling fraction), PBG guidance may also occur in this case, although experimentally the TIR-guided modes will dominate.

Controlling Number of Modes A striking feature of this type of PCF is that it is “endlessly single-mode” (ESM), that is, the core does not become multimode in the experiments, no matter how short the wavelength of the light.¹⁶ Although the guidance in some respects resembles conventional TIR, it turns out to have some interesting and unique features that distinguish it markedly from step-index fiber. These are due to the piecewise discontinuous nature of the core boundary—sections where (for $n_z > 1$) air holes strongly block the escape of light are interspersed with regions of barrier-free glass. In fact, the cladding operates in a regime where the transverse effective wavelength, Eq. (8), in silica is comparable with the glass substructures in the cladding. The zone of operation in Fig. 4 is $n_{\max} < n_z < n_g$ (point A).

In a solid-core PCF, taking the core radius $\rho = \Lambda$ and using the analysis in Ref. 16, the effective V -parameter can be calculated. This yields the plot in Fig. 8, where the full behavior from very low to very high frequency is predicted (the glass index was kept constant at 1.444). As expected, the number of guided modes approximately $V_{\text{PCF}}^2/2$ is almost independent of wavelength at high frequencies;

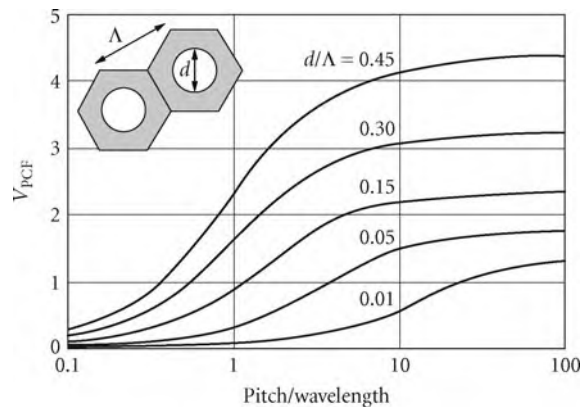


FIGURE 8 V -parameter for solid-core PCF (triangular lattice) plotted against the ratio of hole spacing to vacuum wavelength for different values of d/Λ . Numerical modeling shows that ESM behavior is maintained for $V_{\text{PCF}} \leq 4$ or $d/\Lambda \leq 0.43$.

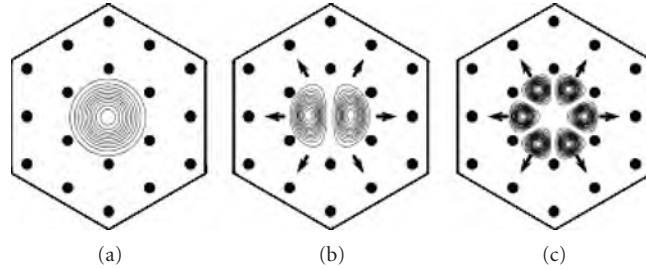


FIGURE 9 Schematic of modal filtering in a solid-core PCF: (a) The fundamental mode is trapped whereas (b) and (c) higher-order modes leak away through the gaps between the air holes.

the single-mode behavior is determined solely by the geometry. Numerical modeling shows that if $d/\Lambda < 0.43$ the fiber never supports any higher-order guided modes, that is, it is ESM.

This behavior can be understood by viewing the array of holes as a modal filter or “sieve” (Fig. 9). The fundamental mode in the glass core has a transverse effective wavelength $\lambda_{\text{eff}}^s \approx 4\Lambda$. It is thus unable to “squeeze through” the glass channels between the holes, which are $\Lambda - d$ wide and thus below the Rayleigh resolution limit $\approx \lambda_{\text{eff}}^s / 2 = 2\Lambda$. Provided the relative hole size d/Λ is small enough, higher-order modes are able to escape their transverse effective wavelength is shorter so they have higher resolving power. As the holes are made larger, successive higher-order modes become trapped.

ESM behavior may also be viewed as being caused by strong wavelength dispersion in the photonic crystal cladding, which forces the core-cladding index step to fall as the wavelength gets shorter (Fig. 7).^{16,49} This counteracts the usual trend toward increasingly multimode behavior at short wavelengths. In the limit of very short wavelength the light strikes the glass-air interfaces at glancing incidence, and is strongly rejected from the air holes. In this regime the transverse single-mode profile does not change with wavelength. As a consequence the angular divergence (roughly twice the numerical aperture) of the emerging light is proportional to wavelength; in SMFs it is approximately constant owing to the appearance of more and more higher-order guided modes as the frequency increases.

Thus, the refractive index of the photonic crystal cladding increases with optical frequency, tending toward the index of silica glass in the short wavelength limit. If the core is made from a glass of refractive index lower than that of silica (e.g., fluorine-doped silica), guidance is lost at wavelengths shorter than a certain threshold value (see Fig. 7).⁵³ Such fibers have the unique ability to prevent transmission of short wavelength light—in contrast to conventional fibers which guide more and more modes as the wavelength falls.

Ultra-Large Area Single-Mode The modal filtering in ESM-PCF is controlled only by the geometry (d/Λ for a triangular lattice). A corollary is that the behavior is quite independent of the absolute size of the structure, permitting single-mode fiber cores with arbitrarily large areas. A single-mode PCF with a core diameter of 22 μm at 458 nm was reported in 1998.¹⁷ In conventional step-index fibers, where $V < 2.405$ for single-mode operation, this would require uniformity of core refractive index to approximately part in 10^5 —very difficult to achieve if MCVD is used to form the doped core. Larger mode areas allow higher power to be carried before the onset of intensity-related nonlinearities or damage, and have obvious benefits for delivery of high laser power, fiber amplifiers, and fiber lasers. The bend-loss performance of such large-core PCFs is discussed in subsection “Bend Loss” in Sec. 11.7.

Fibers with Multiple Cores The stacking procedure makes it straightforward to produce multicore fiber. A preform stack is built up with a desired number of solid (or hollow) cores, and drawn down to fiber in the usual manner.²⁰ The coupling strength between the cores depends on

the sites chosen, because the evanescent decay rate of the fields changes with azimuthal direction. Applications include curvature sensing.⁵⁴ More elaborate structures can be built up, such as fibers with a central single-mode core surrounded by a highly multimode cladding waveguide are useful in applications such as high power cladding-pumped fiber lasers^{55,56} and two-photon fluorescence sensors⁵⁷ (see Fig. 2f).

Negative Core-Cladding Index Difference

Since TIR cannot operate under these circumstances, low loss waveguiding is only possible if a PBG exists in the range $\beta < kn_{co}$.

Hollow-Core Silica/Air In silica-air PCF, larger air-filling fractions and small interhole spacings are necessary to achieve photonic bandgaps in the region $\beta < k$. The relevant operating region on Fig. 4 is to the left of the vacuum line, inside one of the bandgap fingers (point B). These conditions ensure that light is free to propagate, and form guided modes, within the hollow core while being unable to escape into the cladding. The number N of such modes is controlled by the depth and width of the refractive index “potential well” and is approximately given by

$$N \approx k^2 \rho^2 (n_{\text{high}}^2 - n_{\text{low}}^2) / 2 \quad (9)$$

where n_{high} and n_{low} are the refractive indices at the edges of the PBG at fixed frequency and ρ is the core radius. Since the bandgaps are quite narrow ($n_{\text{high}}^2 - n_{\text{low}}^2$ is typically a few percent) the hollow core must be sufficiently large if a guided mode is to exist at all. In the first hollow-core PCF, reported in 1999,⁷ the core was formed by omitting seven capillaries from the preform stack (Fig. 1d). An electron micrograph of a more recent structure, with a hollow core made by removing 19 missing capillaries from the stack, is shown in Fig. 10.⁵⁸

In hollow-core PCF, guidance can only occur when a photonic bandgap coincides with a core resonance. This means that only restricted bands of wavelength are guided. This feature can be very useful for suppressing parasitic transitions by filtering away the unwanted wavelengths, for example, in fiber lasers and in stimulated Raman scattering in gases.⁵⁹

Higher Refractive Index Glass Achieving a bandgap in higher refractive index glasses for $\beta < k$ presents at first glance a dilemma. Whereas a larger refractive index contrast generally yields wider bandgaps, the higher “mean” refractive index seems likely to make it more difficult to achieve bandgaps

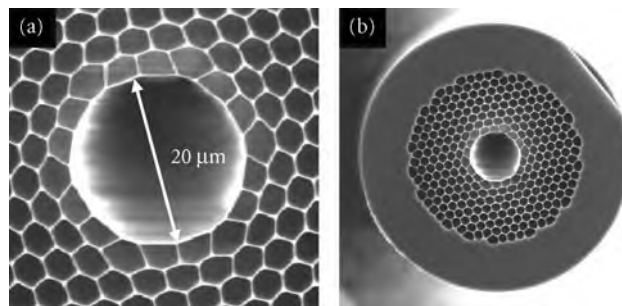


FIGURE 10 Scanning electron micrographs of a low loss hollow PCF (manufactured by BlazePhotonics Ltd.) with attenuation approximately 1 dB/km at 1550 nm wavelength: (a) detail of the core (diameter 20.4 μm) and (b) the complete fiber cross section.

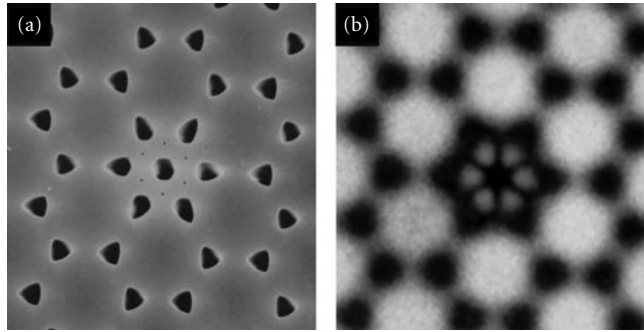


FIGURE 11 (a) A PCF with a “carbon-ring” lattice of air holes and an extra central hole to form a low index core. (b) When white light is launched, only certain bands of wavelength are transmitted in the core—here a six-lobed mode (in the center of the image, blue in the original near-field image) emerges from the end-face.⁵⁰

for incidence from vacuum. Although this argument holds in the scalar approximation, the result of calculations show that vector effects become important at higher levels of refractive index contrast (e.g., 2:1 or higher) and a new species of bandgap appears for smaller filling fractions of air than in silica-based structures. The appearance of this new type of gap means that it is actually easier to obtain wide bandgaps with higher index glasses such as tellurites or chalcogenides.⁴³

Surface States on Core-Cladding Boundary The first PCF that guided by photonic bandgap effects consisted of a lattice of air holes arranged in the same way as the carbon rings in graphite. The core was formed by introducing an extra hole at the center of one of the rings, its low index precluding the possibility of TIR guidance.⁵⁰ When white light was launched into the core region, a colored mode was transmitted—the colors being dependent on the absolute size to which the fiber was drawn. The modal patterns had six equally strong lobes, disposed in a flower-like pattern around the central hole. Closer examination revealed that the light was guided not in the air holes but in the six narrow regions of glass surrounding the core (Fig. 11). The light remained in these regions, despite the close proximity of large “rods” of silica, full of modes. This is because, for particular wavelengths, the phase velocity of the light in the core is not coincident with any of the phase velocities available in the transmission bands created by nearest-neighbor coupling between rod modes. Light is thus unable to couple over to them and so remains trapped in the core.

Similar guided modes are commonly seen in hollow-core PCF, where they form surface states (analogous with electronic surface states in semiconductor crystals) on the rim of the core, confined on the cladding side by photonic bandgap effects. These surface states become phase-matched to the air-guided mode at certain wavelengths, and if the two modes share the same symmetry they couple to form an anticrossing on the frequency-wavevector diagram (Figs. 12 and 13). Within the anticrossing region, the modes share the characteristics of both an air-guided mode and a surface mode, and this consequently perturbs the group velocity dispersion and contributes additional attenuation (see subsection “Absorption and Scattering” in Sec. 11.7).^{60–62}

All-Solid Structures In all-solid bandgap guiding fibers the core is made from low index glass and is surrounded by an array of high index glass strands.^{63–65} Since the mean core-cladding index contrast is negative, TIR cannot operate, and photonic bandgap effects are the only possible guidance mechanism. These structures have some similarities with one-dimensional “ARROW” structures, where antiresonance plays an important role.⁶⁶

When the cladding strands are antiresonant, light is confined to the central low index core by a mechanism not dissimilar to the modal filtering picture in subsection “Controlling Number of

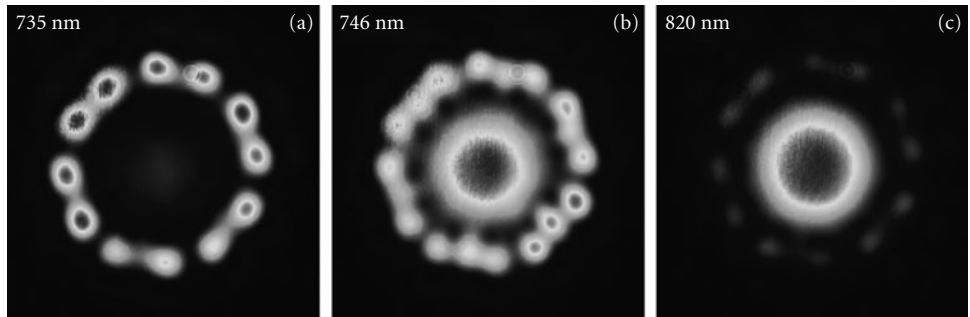


FIGURE 12 Near-field end-face images of the light transmitted in hollow-core PCF designed for 800 nm transmission. For light launched in the core mode, at 735 nm an almost pure surface mode is transmitted, at 746 nm a coupled surface-core mode, and at 820 nm an almost pure core mode.⁶⁰ (The ring-shaped features are an artifact caused by converting from false color to gray scale; the intensity *increases* toward the dark centers of the rings.) (Images courtesy G. Humbert, University of Bath).⁶⁰

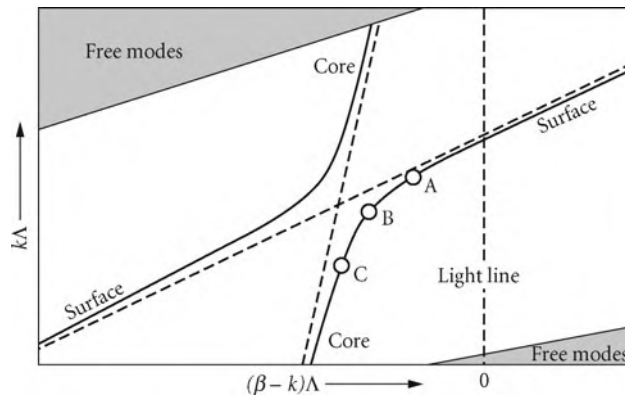


FIGURE 13 Example mode trajectories showing the anticrossing of a core-guided mode with a surface mode in a hollow-core PCF. The dotted lines show the approximate trajectories of the two modes in the absence of coupling (for instance if the modes are of different symmetries), and the vertical dashed line is the light line. The gray regions, within which the mode trajectories are not shown, are the band edges; the white region is the photonic bandgap. Points A, B, and C are the approximate positions of the modes shown in Fig. 12.

Modes” in Sec. 11.7;⁵² the high index cores act as the “bars of a cage,” so that no features in the cladding are resonant with the core mode, resulting in a low loss guided mode. Guidance is achieved over wavelength ranges that are punctuated with high loss windows where the cladding “bars” become resonant (Fig. 14). Remarkably, it is possible to achieve photonic bandgap guidance by this mechanism even at index contrasts of 1 percent,^{63,67} with losses as low as 20 dB/km at 1550 nm.⁶⁸

Low Leakage Guidance The transmission bands are greatly widened in hollow-core PCFs with a kagomé lattice in the cladding⁵² (Fig. 2j). The typical attenuation spectrum of such a fiber has a loss of order 1 dB/m over a bandwidth of 1000 nm or more. Numerical simulations show that, while the cladding structure supports no bandgaps, the density of states is greatly reduced near the vacuum

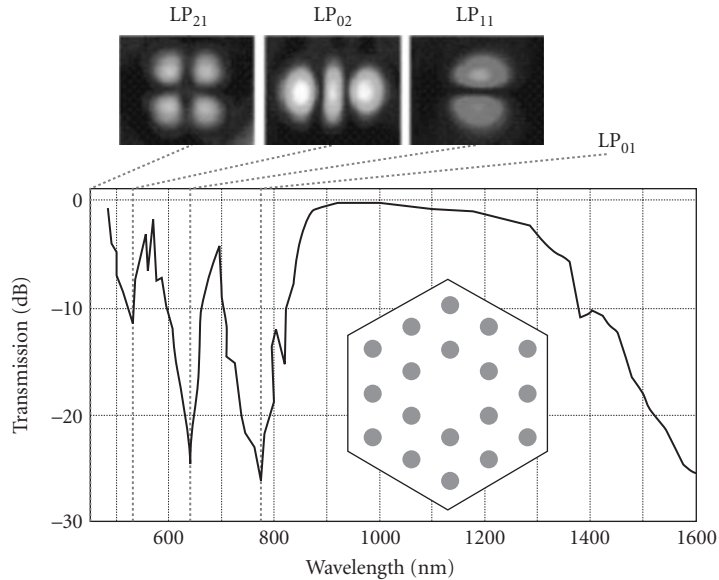


FIGURE 14 Lower: Measured transmission spectrum (using a white-light supercontinuum source) for a PCF with a pure silica core and a cladding formed by an array of Ge-doped strands ($d/\Lambda = 0.34$ hole spacing $\sim 7 \mu\text{m}$, index contrast 1.05:1). The transmission is strongly attenuated when the core mode becomes phase-matched to different LP_{nm} “resonances” in the cladding strands. Upper: Experimental images [left to right, taken with blue (500 nm), green (550 nm), and red (650 nm) filters] of the near-field profiles in the cladding strands at three such wavelengths. The fundamental LP_{01} resonance occurs at approximately 820 nm and the four-lobed blue resonance lies off the edge of the graph.

line. The consequential poor overlap between the core states, together with the greatly reduced number of cladding states, appears to slow down the leakage of light—though the precise mechanism is still a matter of debate.⁵²

Birefringence

The modes of a perfect sixfold symmetric core and cladding structure are not birefringent.⁶⁹ In practice, however, the large glass-air index difference means that even slight accidental distortions in the structure yield a degree of birefringence. Therefore, if the core is deliberately distorted so as to become twofold symmetric, extremely high values of birefringence can be achieved. For example, by introducing capillaries with different wall thicknesses above and below a solid glass core (Figs. 1c and 2g), values of birefringence some 10 times larger than in conventional fibers can be obtained.⁷⁰ It is even possible to design and fabricate strictly single-polarization PCFs in which only one polarization state is guided.⁷¹ By filling selected holes with a polymer, the birefringence can be thermally tuned.⁷² Hollow-core PCF with moderate levels of birefringence ($\sim 10^{-4}$) can be realized either by forming an elliptical core or by adjusting the structural design of the core surround.^{73,74}

Experiments show that the birefringence in PCF is some 100 times less sensitive to temperature variations than in conventional fibers, which is important in many applications.^{75–77} This is because traditional “polarization maintaining” fibers (bow-tie, elliptical core, or Panda) contain at least two different glasses, each with a different thermal expansion coefficient. In such structures, the resulting temperature-dependent stresses make the birefringence a strong function of temperature.

Group Velocity Dispersion

Group velocity dispersion (GVD)—which causes different frequencies of light to travel at different group velocities—is a factor crucial in the design of telecommunications systems and in all kinds of nonlinear optical experiments. PCF offers greatly enhanced control of the magnitude and sign of the GVD as a function of wavelength. In many ways this represents an even greater opportunity than a mere enhancement of the effective nonlinear coefficient.

Solid Core As the optical frequency increases, the GVD in SMF changes sign from anomalous ($D > 0$) to normal ($D < 0$) at approximately $1.3 \mu\text{m}$. In solid-core PCF as the holes get larger, the core becomes more and more isolated, until it resembles an isolated strand of silica glass (Fig. 15). If the whole structure is made very small (core diameters $< 1 \mu\text{m}$ have been made) the zero dispersion point of the fundamental guided mode can be shifted to wavelengths in the visible.^{9,18} For example, the PCF in Fig. 1f has a dispersion zero at 560 nm .

By careful design, the wavelength dependence of the GVD can also be reduced in PCFs at much lower air-filling fractions. Figure 16 shows the flattened GVD profiles of three PCFs with cores several μm in diameter.^{78,79} These fibers operate in the regime where SMF is multimoded. Although the fundamental modes in both SMF and PCF have similar dispersion profiles, the presence of higher-order modes (not guided in the PCF, which is endlessly single mode) makes the use of SMF impractical.

A further degree of freedom in GVD design may be gained by working with multicomponent glasses, such as Schott SF6, where the intrinsic zero dispersion point occurs at approximately $1.8 \mu\text{m}$.⁸ In highly nonlinear small-core PCF, this shifts the whole dispersion landscape to longer wavelengths than in a silica-based PCF with the same core size and geometry.

Hollow Core Hollow-core fiber behaves in many respects rather like a circular-cylindrical hollow metal waveguide, which has anomalous dispersion (the group velocity increases as the frequency rises). The main difference, however, is that the dispersion changes sign at the high frequency edge, owing to the approach of the photonic band edge and the weakening of the confinement (Fig. 17).⁸⁰

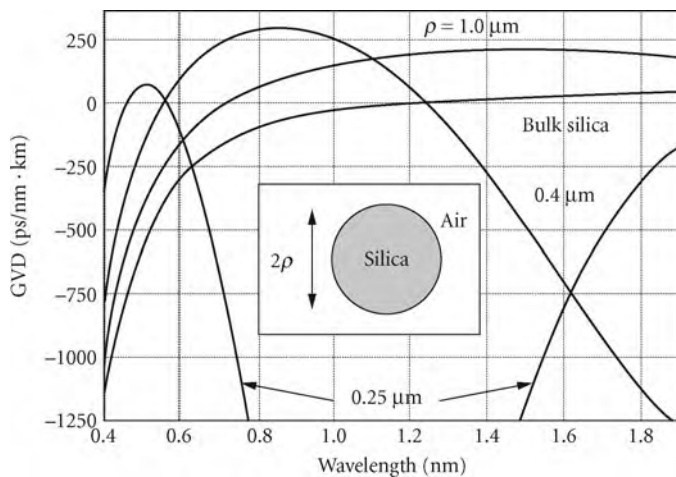


FIGURE 15 The calculated group velocity dispersion of three circular strands of silica glass, radii 0.25 , 0.4 , and $1.0 \mu\text{m}$, compared with the dispersion of bulk glass. The narrower strands have two dispersion zeros within the transparency window of silica.

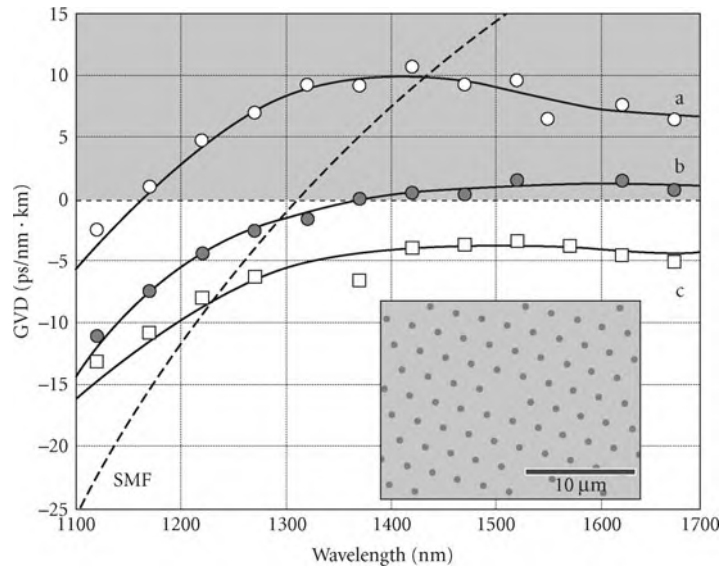


FIGURE 16 Group velocity dispersion profiles, against wavelength, for three different PCFs designed to have lowlevel ultraflattened GVD.^{78,79} The curve for Corning SMF-28 is included for comparison.

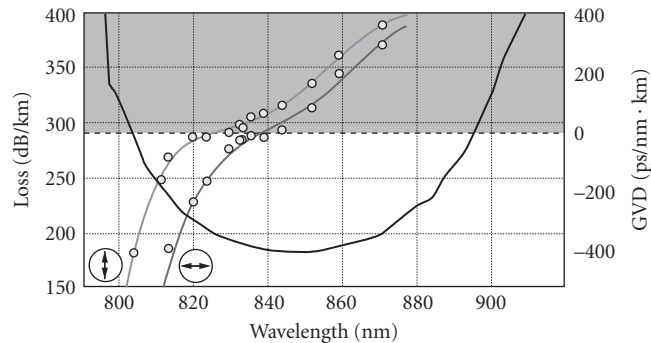


FIGURE 17 Measured attenuation and GVD spectra for a hollow-core PCF designed for 850 nm transmission.⁸⁰ The core is slightly elliptical, so the dispersion in each eigenstate of polarization is different.

Attenuation Mechanisms

An advantage common to all fibers is the very large extension ratio from preform to fiber, which has the effect of smoothing out imperfections, resulting in a transverse structure that is extremely invariant with distance along the fiber. This is the chief reason for the ultralow attenuation displayed by fibers, compared to other waveguide structures. In PCF the losses are governed by two main parameters: the fraction of light in glass and the roughness at the glass-air interfaces. The light-in-glass fraction can be controlled by judicious design, and ranges from close to 100 percent in solid core fibers to less than 1 percent in the best hollow-core fibers.

Absorption and Scattering The best reported loss in solid-core PCF, from a group in Japan, stands at 0.28 dB/km at 1550 nm, with a Rayleigh scattering coefficient of $0.85 \text{ dB} \cdot \text{km}^{-1} \cdot \mu\text{m}^{-4}$. A 100 km length of this fiber was used in the first PCF-based penalty-free dispersion-managed soliton transmission system at 10 Gb/s.^{81,82} The slightly higher attenuation compared to SMF is due to roughness at the glass-air interfaces.⁶¹

It is hollow-core PCF, however, that has the greatest potential for extremely low loss, since the light is traveling predominantly in empty (or gas-filled) space. Although the best reported attenuation in hollow-core PCF stands at 1.2 dB/km,⁵⁸ values below 0.2 dB/km or even lower seem feasible with further development of the technology. The prospect of improving on conventional fiber, at the same time greatly reducing the nonlinearities associated with a solid glass core, is intriguing. By using infrared glasses, transmission can be extended into the infrared⁸³ and recent work shows that silica hollow-core PCF can even be used with acceptable levels of loss in the mid-IR,⁸⁴ owing to the very low overlap between the optical field and the glass.

In the latest hollow-core silica PCF, with loss levels approaching 1 dB/km at 1550 nm, very small effects can contribute significantly to the attenuation floor. The ultimate loss limit in such fibers is determined by surface roughness caused by thermally driven capillary waves, which are present at all length scales. These interface ripples freeze in when the fiber cools, introducing high scattering losses for modes that are concentrated at the interfaces, such as surface modes guided on the edge of the core. The pure core mode does not itself “feel” the ripples very strongly, except at anticrossing wavelengths where it becomes phase-matched to surface modes, causing light to move to the surface and experience enhanced scattering.

The result is a transmission spectrum consisting of windows of high transparency punctuated with bands of high attenuation (Fig. 18). This picture has been confirmed by measurements of the surface roughness in hollow-core PCFs, the angular distribution of the power scattered out of the core, and the wavelength dependence of the minimum loss of fibers drawn to different scales.⁵⁸ The thin glass shell surrounding the hollow core can be designed to be antiresonant with the core mode, permitting further exclusion of light from the glass.⁸⁵

Ignoring material dispersion, the whole transmission landscape shifts linearly in wavelength in proportion to the overall size of the structure (a consequence of Maxwell’s equations). This means

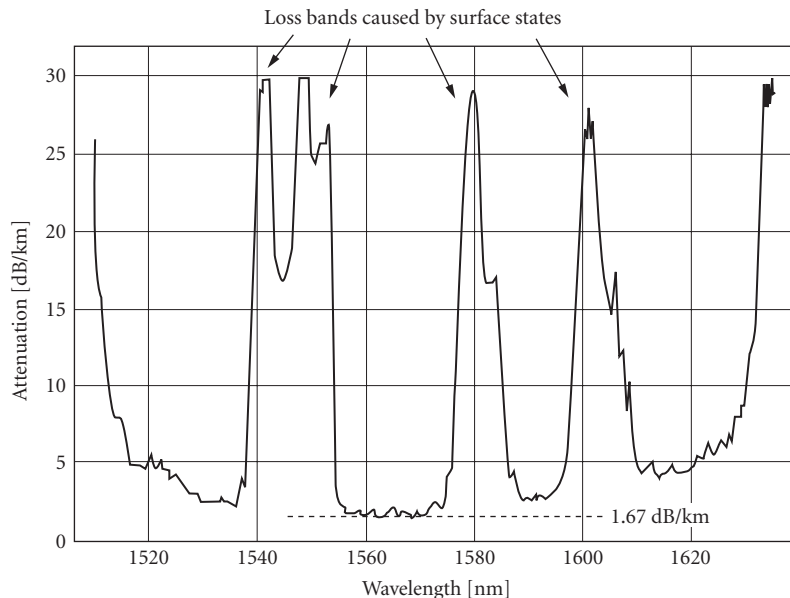


FIGURE 18 Attenuation spectrum of a typical ultralow loss hollow-core PCF designed for operation in the 1550 nm telecommunications band (see micrographs in Fig. 10).

that the smallest loss at a given wavelength will be obtained by drawing a fiber to a particular diameter. The optical overlap with the surface roughness scales inversely with the size with the fiber, and the scattering itself may be regarded as being governed by the density of states into which scattering can occur, which in three-dimensions scales as λ^{-2} . Thus the wavelength of minimum loss scales as λ^{-3} , in contrast to the λ^{-4} dependence of Rayleigh scattering in bulk glass.

Bend Loss Conventional fibers suffer additional loss if bent beyond a certain critical radius R_c , which depends on wavelength, core-cladding refractive index step and—most notably—the third power of core radius ρ .³⁹ For wavelengths longer than a certain value (the “long wavelength bend edge”) all guidance is effectively lost.

A starting point for understanding bend loss in solid core ESM-PCF (perhaps the most interesting case) is the long wavelength limit. ESM behavior occurs when $d/\Lambda < 0.43$, which sets the highest air-filling fraction at 16.8 percent and yields an area-averaged cladding refractive index of 1.388 (silica index of 1.444)—valid in the limit $k \rightarrow 0$. This index step is some 10 times higher than in Corning SMF-28, making ESM-PCF relatively much less susceptible to bend loss at long wavelengths. For a step-index fiber with a Ge-doped core, 40 mol % of GeO_2 would be needed to reach the same index step (assuming 0.0014 index change per mol % GeO_2).¹¹ The result is that the long-wavelength bend edge in ESM-PCF is in the infrared beyond the transparency window of silica glass, even when the core radius is large.⁸⁶

ESM-PCF also exhibits a *short* wavelength bend edge, caused by bend-induced coupling from the fundamental to higher-order modes, which of course leak out of the core.¹⁶ The critical bend radius for this loss varies as

$$R_c \sim \Lambda^3/\lambda^2 \quad (10)$$

compared to R_c approximately λ for SMF. The reciprocal dependence on λ^2 makes it inevitable that a short-wavelength bend edge will appear in ESM-PCF. Following⁸⁶ in taking the pre-factor in Eq. (10) as unity, R_c can be plotted against wavelength for different values of core radius ($\approx$ interhole spacing). This is illustrated in Fig. 19. A step-index fiber, with the same core-cladding index step as

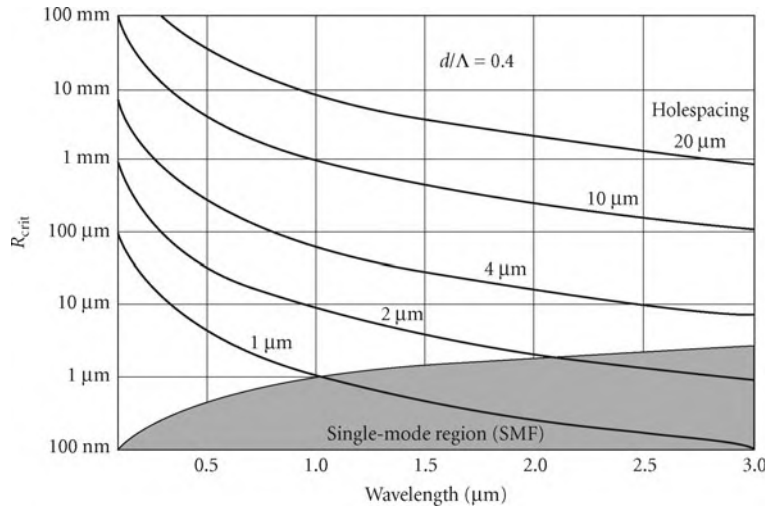


FIGURE 19 Short-wavelength critical bend radii for ESM-PCF with $d/\Lambda = 0.4$ plotted against vacuum wavelength for different values of hole-spacing (approximately equal to the core radius). As the wavelength increases at constant core size, a step-index fiber with the same core/cladding index as the PCF in the long-wavelength limit [1.444:1.388 using Eq. (5)] becomes single-mode when the curves enter the shaded region. The step-index fiber is multi-mode over wide parameter ranges where ESM-PCF has negligible short-wavelength bend loss.

ESM-PCF in the long wavelength limit, is multimode over wide parameter ranges where ESM-PCF has negligible bend loss.

In contrast, hollow-core PCF is experimentally very insensitive to bend loss—in many cases no appreciable drop in transmission is observed until the fiber breaks. This is because the effective depth of “potential well” for the guided light (see section “Negative Core-Cladding Index Difference”), given by the distance $\Delta\beta$ between the edges of the photonic bandgap, is substantially larger than in SMF.

Confinement Loss The photonic crystal cladding in a realistic PCF is of course finite in extent. For a guided mode, the Bloch waves in the cladding are evanescent, just like the evanescent plane waves in the cladding of a conventional fiber. If the cladding is not thick enough, the evanescent field amplitudes at the cladding/coating boundary can be substantial, causing attenuation. In the solid core case for small values of d/Λ the resulting loss can be large unless a sufficiently large number of periods is used.⁷⁸

Very similar losses are observed in hollow-core fibers, where the “strength” of the photonic bandgap (closely related to its width in β) determines how many periods are needed to reduce confinement loss to acceptable levels. Numerical modeling is useful for giving an indication of how many periods are needed to reach a required loss level. The cladding field intensity in the ultralow loss PCF reported in⁵⁸ falls by approximately 9 dB per period, reaching –63 dB at the edge of the photonic crystal region.

11.8 NONLINEAR CHARACTERISTICS OF GUIDANCE

The ability to enhance or reduce the effective nonlinear coefficients, and at the same time control the magnitude and wavelength dependence of the GVD, makes PCF a versatile vehicle for studies of nonlinear effects.

Kerr Nonlinearities

The optical Kerr effect drives effects such as four-wave mixing, self-phase modulation, modulation instability, and soliton formation. To take account of the differing proportions of light in glass and air,⁸⁷ it is necessary to derive an expression for the effective nonlinear γ coefficient, which in the scalar approximation takes the form

$$\gamma = \frac{k \iint n(x, y) n_2(x, y) \psi^4(x, y) dx dy}{n_m \left(\iint \psi^2(x, y) dx dy \right)^2} W^{-1} m^{-1} \quad (11)$$

where ψ is the transverse field amplitude profile, $n(x, y)$ and $n_2(x, y)$ are the linear and nonlinear refractive index profiles, and n_m is the modal phase index. The integrals must be evaluated over the entire transverse plane, which in practice means over the area where the field amplitudes are nonnegligible. Equation (11) can be reexpressed for a composite PCF made from two or more homogeneous materials

$$\gamma = k \sum_i \frac{n_2^i}{A_i} \quad (12)$$

where n_2^i is the nonlinear refractive index of material i and A_i its nonlinear effective area

$$A_i = \frac{n_m \left(\iint \psi^2(x, y) dx dy \right)^2}{\iint u_i(x, y) \psi^4(x, y) dx dy} \quad (13)$$

where $u_i(x, y)$ equals n_i in regions made from material i and zero elsewhere.

The nonlinear refractive indices of silica glass and air are, respectively, $2.5 \times 10^{-20} \text{ m}^2\text{W}^{-1}$ and $2.9 \times 10^{-23} \text{ m}^2\text{W}^{-1}$. Multicomponent glasses typically have values an order of magnitude or more higher, and the hollow channels can of course be filled with another gas with a different nonlinear coefficient.

The highest nonlinear coefficient available in conventional step-index fibers is $\gamma: 20 \text{ W}^{-1}\text{km}^{-1}$ at 1550 nm.⁸⁸ By comparison, a solid-core PCF similar to the one in Fig. 1f but with a core diameter 1 μm has a nonlinearity of $\gamma: 240 \text{ W}^{-1}\text{km}^{-1}$, at 850 nm, and values as high as $\gamma: 550 \text{ W}^{-1}\text{km}^{-1}$ at 1550 nm have been measured for PCFs made from multicomponent glasses.⁸⁹ In complete contrast, hollow-core PCF has extremely low levels of nonlinearity, owing to the small overlap between the glass and the light. In a recent example, a fiber was reported with a nonlinear coefficient $\gamma = 0.023 \text{ W}^{-1}\text{km}^{-1}$ (some $10^4 \times$ smaller than in a typical highly nonlinear solid-core PCF).^{90,91}

Although the level of nonlinearity is clearly important, the actual nonlinear effects that appear in a particular case are also strongly dependent on the magnitude, sign and wavelength dependence of the GVD as well as on the characteristics of the pump laser.⁹² They are determined by the relative values of the nonlinear length $L_{\text{nl}} = \gamma^{-1}P_0^{-1}$, where P_0 is the peak power, the dispersion length $L_D = \tau^2/|\beta_2|$, where τ is the pulse duration and the effective fiber length $L_{\text{eff}} = [1 - \exp(-\alpha L)]/\alpha$ where $\alpha \text{ m}^{-1}$ is the power attenuation coefficient.⁹³ For a solid-core PCF with $\gamma: 240 \text{ W}^{-1}\text{km}^{-1}$, a peak power of 10 kW yields $L_{\text{nl}} < 0.5 \text{ mm}$. For typical values of loss (usually between 1 and 100 dB/km) $L_{\text{eff}} \gg L_{\text{nl}}$ and the nonlinearity dominates. For dispersion values in the range $-300 < \beta_2 < 300 \text{ ps}^2/\text{km}$ and pulse durations $\tau = 200 \text{ fs}$, $L_D > 0.1 \text{ m}$. Since both these lengths are much longer than the nonlinear length, it is easy to observe strong nonlinear effects.

Supercontinuum Generation One of the most successful applications of nonlinear PCF is to supercontinuum (SC) generation from ps and fs laser pulses. When high power pulses travel through a material, their frequency spectrum can be broadened by a range of interconnected nonlinear effects.⁹⁴ In bulk materials, the preferred pump laser is a regeneratively amplified Ti:sapphire system producing high (mJ) energy fs pulses at 800 nm wavelength and kHz repetition rate. Supercontinua have also previously been generated in SMF by pumping at 1064 or 1330 nm,⁹⁵ the spectrum broadening out to longer wavelengths mainly due to stimulated Raman scattering (SRS). Then in 2000, it was observed that highly nonlinear PCF, designed with zero GVD close to 800 nm, massively broadens the spectrum of low (few nJ) energy unamplified Ti:sapphire pulses launched into just a few cm of fiber.^{21,96,97} Removal of the need for a power amplifier, the hugely increased ($\sim 100 \text{ MHz}$) repetition rate, and the spatial and temporal coherence of the light emerging from the core, makes this source unique. The broadening extends both to higher and to lower frequencies because four-wave mixing operates more efficiently than SRS when the dispersion profile is appropriately designed. This SC source has applications in optical coherence tomography,^{98,99} frequency metrology,^{100,101} and all kinds of spectroscopy. It is particularly useful as a bright lowcoherence source in measurements of group delay dispersion based on a Mach-Zehnder interferometer.

A comparison of the bandwidth and spectrum available from different broad-band light sources is shown in Fig. 20; the advantages of PCF-based SC sources are evident. Supercontinua have been generated in different PCFs at 532 nm,¹⁰² 647 nm,¹⁰³ 1064 nm,¹⁰⁴ and 1550 nm.⁸ Using inexpensive microchip lasers at 1064 or 532 nm with an appropriately designed PCF, compact SC sources are now available with important applications across many areas of science. A commercial ESM-PCF based source uses a 10-W fiber laser delivering 5 ps pulses at 50 MHz repetition rate, and produces an average spectral power density of approximately 4.5 mW/nm in the range 450 to 800 nm.¹⁰⁵ The use of multicomponent glasses such as Schott SF6 or tellurite glass allows the balance of nonlinearity and dispersion to be adjusted, as well as offering extended transparency into the infrared.¹⁰⁶

Parametric Amplifiers and Oscillators In step-index fibers the performance of optical parametric oscillators and amplifiers is severely constrained owing to the limited scope for GVD engineering. In PCF these constraints are lifted, permitting flattening of the dispersion profile and control of higher-order dispersion terms. The wide range of experimentally available group-velocity dispersion profiles has, for example, allowed studies of ultrashort pulse propagation in the 1550 nm wavelength band with flattened dispersion.^{78,79} The effects of higher-order dispersion in such PCFs are subtle.^{107,108} Parametric devices have been designed for pumping at 647, 1064, and 1550 nm, the small

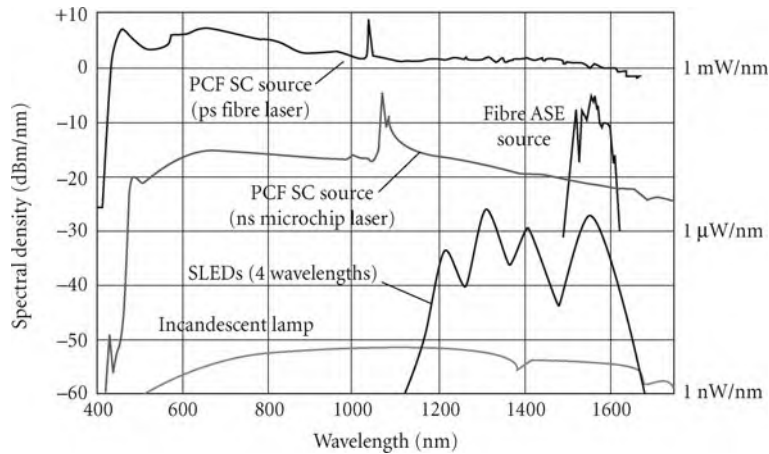


FIGURE 20 Comparison of the brightness of various broad-band light sources (SLED—superluminescent light-emitting diode; ASE—amplified spontaneous emission; SC—supercontinuum). The microchip laser SC spectrum was obtained by pumping at 1064 nm with 600 ps pulses. (Updated version of a plot by Hendrik Sabert.)

effective mode areas offering high gain for a given pump intensity, and PCF-based oscillators synchronously pumped by fs and ps pump pulses have been demonstrated at relatively low power levels.^{109–112} Dispersion-engineered PCF is being successfully used in the production of bright sources of correlated photon pairs, by allowing the signal and idler side-bands to lie well outside the noisy Raman band of the glass. In a recent example, a PCF with zero dispersion at 715 nm was pumped by a Ti:sapphire laser at 708 nm (normal dispersion).¹¹³ Under these conditions phase-matching is satisfied by signal and idler waves at 587 and 897 nm, and 10 million photon pairs per second were generated and delivered via single-mode fiber to Si avalanche detectors, producing approximately 3.2×10^5 coincidences per second for a pump power of 0.5 mW. These results point the way to practical and efficient sources entangled photon pairs that can be used as building blocks in future multiphoton interference experiments.

Soliton Self-Frequency Shift Cancellation The ability to create PCFs with negative dispersion slope at the zero dispersion wavelength (in SMF the slope is positive, i.e., the dispersion becomes more anomalous as the wavelength increases) has made it possible to observe Čerenkov-like effects in which solitons (which form on the anomalous side of the dispersion zero) shed power into dispersive radiation at longer wavelengths on the normal side of the dispersion zero. This occurs because higher-order dispersion causes the edges of the soliton spectrum to phase-match to linear waves. The result is stabilization of the soliton self-frequency shift, at the cost of gradual loss of soliton energy.¹¹⁴ The behavior of solitons in the presence of wavelength-dependent dispersion is the subject of many recent studies.¹¹⁵

Raman Scattering

The basic characteristics of glass-related Raman scattering in PCF, both stimulated and spontaneous, do not noticeably differ compared to SMF. One must of course take account of the differing proportions of light in glass and air (see section “Kerr Nonlinearities”) to calculate the effective strength of the Raman response. A very small solid glass core allows one to enhance stimulated Raman scattering, whereas in a hollow core it is strongly suppressed.

Brillouin Scattering

The periodic micro/nanostructuring in ultrasmall core glass-air PCF strongly alters the acoustic properties compared to conventional SMF.^{116–119} Sound can be guided in the core both as leaky and as tightly confined acoustic modes. In addition, the complex geometry and “hard” boundaries cause coupling between all three displacement components (radial, azimuthal, and axial), with the result that each acoustic mode has elements of both shear (S) or longitudinal (L) strain. This complex acoustic behavior strongly alters the characteristics of forward and backward Brillouin scattering.

Backward Scattering When a solid-core silica-air PCF has a core diameter of around 70 percent of the vacuum wavelength of the launched laser light, and the air-filling fraction in the cladding is very high, the spontaneous Brillouin signal displays multiple bands with Stokes frequency shifts in the 10 GHz range. These peaks are caused by discrete guided acoustic modes, each with different proportions of longitudinal and shear strain, strongly localized to the core.¹²⁰ At the same time the threshold power for stimulated Brillouin scattering increases fivefold—a rather unexpected result, since conventionally one would assume that higher intensities yield lower nonlinear threshold powers. This occurs because the effective overlap between the tightly confined acoustic modes and the optical mode is actually smaller than in a conventional fiber core; the sound field contains a large proportion of shear strain, which does not contribute significantly to changes in refractive index. This is of direct practical relevance to parametric amplifiers, which can be pumped 5 times harder before stimulated Brillouin scattering appears.

Forward Scattering The very high air-filling fraction in small-core PCF also permits sound at frequencies of a few GHz to be trapped purely in the transverse plane by phononic bandgap effects (Fig. 21). The ability to confine acoustic energy at zero axial wavevector $\beta_{ac} = 0$ means that the ratio

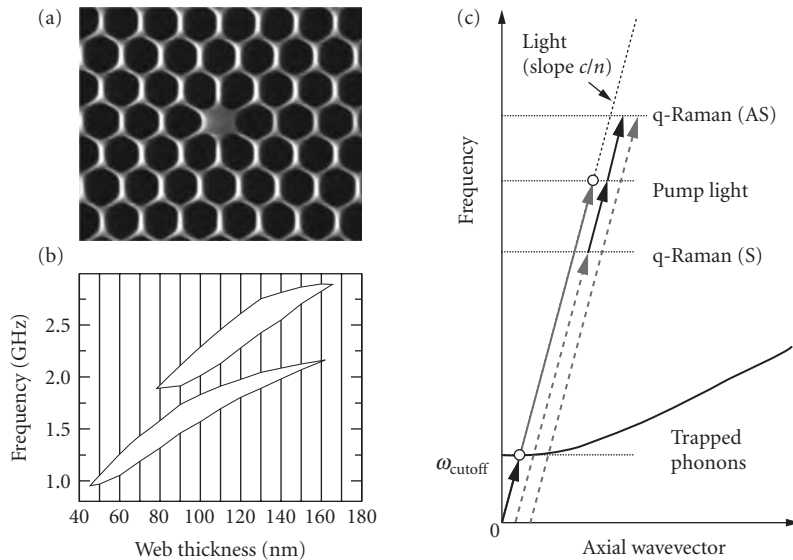


FIGURE 21 (a) Example of PCF used in studies of Brillouin scattering (core diameter 1.1 μm); (b) the frequencies of full phononic bandgaps (in-plane propagation, pure in-plane motion) in the cladding of the PCF in (b); and (c) illustrating how a trapped acoustic phonon can phase-match to light at the acoustic cutoff frequency. The result is a quasi-Raman scattering process that is automatically phase-matched. (After Ref. 121.)

of frequency ω_{ac} to wavevector β_{ac} becomes arbitrarily large as $\beta_{ac} \rightarrow 0$, and thus can easily match the value for the light guided in the fiber, c/n . This permits phase-matched interactions between the acoustic mode and two spatially identical optical modes of different frequency.¹²¹ Under these circumstances the acoustic mode has a well-defined cutoff frequency ω_{cutoff} above which its dispersion curve—plotted on an (ω, β) diagram—is flat, similar to the dispersion curve for optical phonons in diatomic lattices. The result is a scattering process that is Raman-like (i.e., the participating phonons are optical-phonon-like), even though it makes use of acoustic phonons; Brillouin scattering is turned into Raman scattering, power being transferred into an optical mode of the same order, frequency shifted from the pump frequency by the cutoff frequency. Used in stimulated mode, this effect may permit generation of combs of frequencies spaced by approximately 2 GHz at 1550 nm wavelength.

11.9 INTRAFIBER DEVICES, CUTTING, AND JOINING

As PCF becomes more widely used, there is an increasing need for effective cleaves, low loss splices, multiport couplers, intrafiber devices, and mode-area transformers. The air holes provide an opportunity not available in standard fibers: the creation of dramatic morphological changes by altering the hole size by collapse (under surface tension) or inflation (under internal overpressure) when heating to the softening temperature of the glass. Thus, not only can the fiber be stretched locally to reduce its cross-sectional area, but the microstructure can itself be radically altered.

Cleaving and Splicing

PCF cleaves cleanly using standard tools, showing slight end-face distortion only when the core crystal is extremely small (interhole spacing $\sim 1 \mu\text{m}$) and the air-filling fraction very high ($>50\%$). Solid glass end-caps can be formed by collapsing the holes (or filling them with sol-gel glass) at the fiber end to form a core-less structure through which light can be launched into the fiber. Solid-core PCF can be fusion-spliced successfully both to itself and to step-index fiber using resistive heating elements (electric-arcs do not allow sufficient control). The two fiber ends are placed in intimate contact and heated to softening point. With careful control, they fuse together without distortion. Provided the mode areas are well matched, splice losses of <0.2 dB can normally be achieved except when the core is extremely small ($<1.5 \mu\text{m}$). Fusion splicing hollow-core fiber is feasible when there is a thick solid glass outer sheath (e.g., as depicted in Fig. 10b), although very low splice losses can be obtained simply by placing identical fibers end-to-end and clamping them (the index-matching “fluid” for hollow-core PCF is vacuum).

The ability to hermetically splice gas-filled hollow-core PCF to SMF has made it possible to produce in-line gas cells for stimulated Raman scattering in hydrogen and frequency measurement and stabilization (using acetylene). These developments may lead for the first time to practical miniature gas-laser devices that could even be coiled up inside a credit card.¹²²

Mode Transformers

In many applications it is important to be able to change the mode area without losing light. This is done traditionally using miniature bulk optics—tiny lenses precisely designed to match to a desired numerical aperture and spot-size. In PCF an equivalent effect can be obtained by scanning a heat source (flame or carbon dioxide laser) along the fiber. This causes the holes to collapse, the degree of collapse depending on the dwell-time of the heat. Drawing the two fiber ends apart at the same time provides additional control. Graded transitions can fairly easily be made—mode diameter reductions as high as 5:1 have been realized with low loss.

Ferrule methods have been developed for making low loss interfaces between conventional single-mode fibers and photonic crystal fibers.¹²³ Adapted from the fabrication of PCF preforms from stacked tubes and rods, these techniques avoid splicing and are versatile enough to interface to virtually any type of index-guiding silica PCF. They are effective for coupling light into and out of all of the individual cores of a multicore fiber without input or output crosstalk. The technique also creates another opportunity—the use of taper transitions to couple light between a multimode fiber and several single-mode fibers. When the number of single-mode fibers matches the number of spatial modes in the multimode fiber, the transition can have low loss in both directions. This means that the high performance of single-mode fiber devices can be reached in multimode systems, for example a multimode fiber filter with the transmission spectrum of a single-mode fiber Bragg grating,¹²⁴ a device that has applications in earth-based astronomy, where the high throughput of a multimode fiber can be retained while unwanted atmospheric emission lines are filtered out.

A further degree of freedom may be gained by pressurizing the holes during the taper process.¹²⁵ The resulting hole inflation permits radical changes in the guidance characteristics. It is possible for example to transform a PCF, with a relatively large core and small air-filling fraction, into a PCF with a very small core and a large air-filling fraction, the transitions having very low loss.¹²⁶

Heating and stretching PCF can result in quite remarkable changes in the scale of the micro/nano structures, without significant distortion. Recently a solid-core PCF was reduced 5 times in linear scale, resulting in a core diameter of 500 nm (Fig. 22). This permitted formation of a PCF with a zero dispersion wavelength that matched the 532 nm emission wavelength of a frequency doubled Nd:YAG laser¹⁰² (this is important for supercontinuum generation—as discussed in Sec. 11.8). A further compelling advantage of the tapering approach is that it neatly side-steps the difficulty of launching light into submicron sized cores; light is launched into the entry port (in this case with core diameter 2.5 μm) and adiabatically evolves, with negligible loss, into the mode of the 500 nm core.

In-Fiber Devices

Precise use of heat and pressure induces large changes in the optical characteristics of PCF, giving rise to a whole family of new intrafiber components. Microcouplers can be made in a PCF with two optically isolated cores by collapsing the holes so as to allow the mode fields to expand and

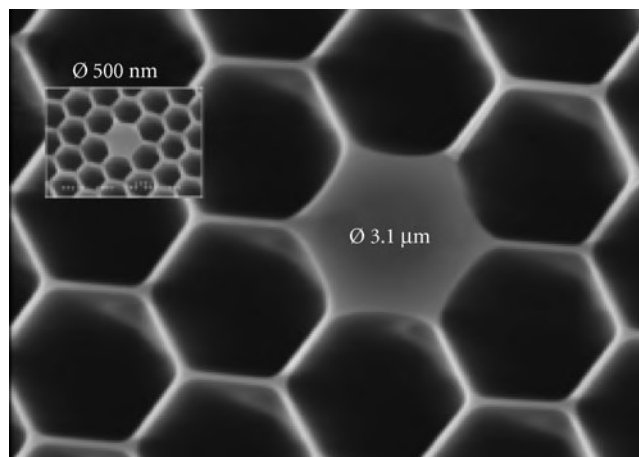


FIGURE 22 Scanning electron micrographs (depicted to the same scale) of the fiber cross sections produced by tapering a solid-core PCF. The structure is very well preserved, even down to core diameters of 500 nm.¹⁰²

interact with each other, creating local coupling.¹²⁷ Long period gratings, which scatter the core light into cladding modes within certain wavelength bands, can be made by periodic modulation of hole size.¹²⁸ By rocking a birefringent PCF to and fro, while scanning a carbon dioxide laser along it, so-called “rocking filters” can be made, which transfer power from one polarization state to the other within a narrow band of wavelengths.¹²⁹ All these components have one great advantage over equivalent devices made in conventional fiber: being formed by permanent changes in morphology, they are highly stable with temperature and over time.

11.10 CONCLUSIONS

In moving away from the constraints of conventional fibers, photonic crystal fibers have created new opportunities spanning many areas of science and technology. Some of the more important from an every widening range of applications (as of late 2006) are discussed in Ref. 1. Compact high brightness supercontinuum sources and frequency comb systems for optical clocks are already available as commercial products. Emerging applications include constrained photochemistry or biochemistry and microfluidics, biophotonic and biomedical devices, medical imaging, astronomy, particle delivery, high power fiber lasers, gas-based fiber devices, and fiber delivery of high power laser light in manufacturing. The next decade should see many of these applications mature into fields of scientific research or even commercial products.

11.11 APPENDIX

The analysis in Ref. 16 to estimate $n_{\max}$ in a triangular lattice of air holes leads to the equation:

$$w I_1(a_n w)[J_1(bu)Y_0(a_n u) - J_0(a_n u)Y_1(bu)] + u I_0(a_n w)[J_1(bu)Y_1(a_n u) - J_1(a_n u)Y_1(bu)] = 0 \quad (14)$$

where $a_n = d/2\Lambda$, $b = (\sqrt{3}/2\pi)^{1/2}$ (needed to ensure the correct value of $n_{\max}$ in the long wavelength limit) and $u^2 + w^2 = v^2$. The leading root of Eq. (14), evaluated for $a_n = 0.2$ and $n_g = 1.444$ (i.e., neglecting dispersion), yields the polynomial fit (7).

11.12 REFERENCES

1. P. St.J. Russell, “Photonic-Crystal Fibers,” *Journal of Lightwave Technology*, **24**:4729–4749, 2006.
2. J. C. Knight, T. A. Birks, P. St.J. Russell, and D. M. Atkin, “Pure Silica Single-Mode Fiber with Hexagonal Photonic Crystal Cladding,” in *Conference on Optical Fiber Communications* San Jose, California: Optical Society of America, 1996.
3. P. St.J. Russell, “Photonic Crystal Fibers,” *Science*, **299**:358–362, Jan. 2003.
4. J. C. Knight, “Photonic Crystal Fibres,” *Nature*, **424**:847–851, Aug. 2003.
5. J. C. Knight, T. A. Birks, P. St.J. Russell, and D. M. Atkin, “All-Silica Single-Mode Optical Fiber with Photonic Crystal Cladding,” *Optics Letters*, **21**:1547–1549, Oct. 1996.
6. J. C. Knight, T. A. Birks, P. St.J. Russell, and D. M. Atkin, “All-Silica Single-Mode Optical Fiber with Photonic Crystal Cladding: Errata,” *Optics Letters*, **22**:484–485, Apr. 1997.
7. R. F. Cregan, B. J. Mangan, J. C. Knight, T. A. Birks, P. St.J. Russell, P. J. Roberts, and D. C. Allan, “Single-Mode Photonic Band Gap Guidance of Light in Air,” *Science*, **285**:1537–1539, Sep. 1999.
8. V. V. R. K. Kumar, A. K. George, W. H. Reeves, J. C. Knight, P. St.J. Russell, F. G. Omenetto, and A. J. Taylor, “Extruded Soft Glass Photonic Crystal Fiber for Ultrabroad Supercontinuum Generation,” *Optics Express*, **10**:1520–1525, Dec. 2002.

9. J. C. Knight, J. Arriaga, T. A. Birks, A. Ortigosa-Blanch, W. J. Wadsworth, and P. St.J. Russell, "Anomalous Dispersion in Photonic Crystal Fiber," *IEEE Photonics Technology Letters*, **12**:807–809, Jul. 2000.
10. www.corning.com/opticalfiber/.
11. E. M. Dianov and V. M. Mashinsky, "Germania-Based Core Optical Fibers," *Journal of Lightwave Technology*, **23**:3500–3508, 2005.
12. P. St.J. Russell, "Designing Photonic Crystals," in *Electron and Photon Confinement in Semiconductor Nanostructures*, Amsterdam: IOS Press, pp. 79–103, 2003.
13. C. M. Bowden, J. P. Dowling, and H. O. Everitt, "Development and Applications of Materials Exhibiting Photonic Band Gaps," *Journal of the Optical Society of America B—Optical Physics*, **10**:280–413, 1993.
14. P. St.J. Russell, in *NATO Advanced Study Institute on Confined Electrons and Holes*, Erice, Sicily, 1993.
15. T. A. Birks, P. J. Roberts, P. St.J. Russell, D. M. Atkin, and T. J. Shepherd, "Full 2-D Photonic Band Gaps in Silica/Air Structures," *Electronics Letters*, **31**:1941–1942, 1995.
16. T. A. Birks, J. C. Knight, and P. St.J. Russell, "Endlessly Single-Mode Photonic Crystal Fiber," *Optics Letters*, **22**:961–963, Jul. 1997.
17. J. C. Knight, T. A. Birks, R. F. Cregan, P. St.J. Russell, and J. P. de Sandro, "Large Mode Area Photonic Crystal Fibre," *Electronics Letters*, **34**:1347–1348, Jun. 1998.
18. D. Mogilevtsev, T. A. Birks, and P. St.J. Russell, "Group-Velocity Dispersion in Photonic Crystal Fibers," *Optics Letters*, **23**:1662–1664, Nov. 1998.
19. A. Ortigosa-Blanch, A. Diez, M. Delgado-Pinar, J. L. Cruz, and M. V. Andres, "Ultrahigh Birefringent Non-linear Microstructured Fiber," *IEEE Photonics Technology Letters*, **16**:1667–1669, Jul. 2004.
20. B. J. Mangan, J. C. Knight, T. A. Birks, P. St.J. Russell, and A. H. Greenaway, "Experimental Study of Dual-core Photonic Crystal Fibre," *Electronics Letters*, **36**:1358–1359, Aug. 2000.
21. J. K. Ranka, R. S. Windeler, and A. J. Stentz, "Visible Continuum Generation in Air–silica Microstructure Optical Fibers with Anomalous Dispersion at 800 nm," *Optics Letters*, **25**:25–27, 2000.
22. V. N. Melekin and A. B. Manenkov, "Dielectric Tube as a Low-Loss Waveguide," *Soviet Physics—Technical Physics*, **13**:1698–1699, 1968.
23. P. Yeh and A. Yariv, "Bragg Reflection Waveguides," *Optics Communications*, **19**:427–430, 1976.
24. F. Brechet, P. Roy, J. Marcou, and D. Pagnoux, "Single-Mode Propagation in Depressed-Core-Index Photonic-Bandgap Fiber Designed for Zero-Dispersion Propagation at Short Wavelengths," *Electronics Letters*, **36**:514–515, 2000.
25. S. G. Johnson, M. Ibanescu, M. Skorobogatiy, O. Weisberg, T. D. Engeness, M. Soljacic, S. A. Jacobs, J. D. Joannopoulos, and Y. Fink, "Low-Loss Asymptotically Single-Mode Propagation in Large-Core Omniguide Fibers," *Optics Express*, **9**:748–779, 2001.
26. D. C. Allan, J. A. West, J. C. Fajardo, M. T. Gallagher, K. W. Koch, and N. F. Borrelli, "Photonic Crystal Fibers: Effective Index and Bandgap Guidance," in *Photonic Crystals and Light Localisation in the 21st Century*, C. M. Soukoulis (ed.), Kluwer Academic Publishers, Netherlands, pp. 305–320, 2001.
27. V. V. R. Kumar, A. K. George, J. C. Knight, and P. St.J. Russell, "Tellurite Photonic Crystal Fiber," *Optics Express*, **11**:2641–2645, Oct. 2003.
28. K. M. Kiang, K. Frampton, M. Monro, R. Moore, J. Tucknott, D. W. Hewak, D. J. Richardson, and H. N. Rutt, "Extruded Single-Mode Non-Silica Glass Holey Optical Fibers," *Electronics Letters*, **38**:546–547, 2002.
29. R. Bise and D. J. Trevor, "Sol-Gel Derived Microstructured Fiber: Fabrication and Characterization," in *Conference on Optical Fiber Communication*, Anaheim, Paper OWL6 2005.
30. M. C. J. Large, A. Argyros, F. Cox, M. A. van Eijkelenborg, S. Ponrathnam, N. S. Pujari, I. M. Bassett, R. Lwin, and G. W. Barton, "Microstructured Polymer Optical Fibres: New Opportunities and Challenges," *Molecular Crystals and Liquid Crystals*, **446**:219–231, 2006.
31. E. F. Chillce, C. M. B. Cordeiro, L. C. Barbosa, and C. H. B. Cruz, "Tellurite Photonic Crystal Fiber Made by a Stack-and-Draw Technique," *Journal of Non-Crystalline Solids*, **352**:3423–3428, Sep. 2006.
32. X. Feng, T. M. Monro, V. Finazzi, R. C. Moore, K. Frampton, P. Petropoulos, and D. J. Richardson, "Extruded Single-Mode, High-Nonlinearity, Tellurite Glass Holey Fibre," *Electronics Letters*, **41**:835–837, Jul. 2005.
33. D. A. Gaponov and A. S. Biryukov, "Optical Properties of Microstructure Tellurite Glass Fibres," *Quantum Electronics*, **36**:343–348, Apr. 2006.

34. T. P. White, B. Kuhlmeier, R. C. McPhedran, D. Maystre, G. Renversez, C. M. De Sterke, and L. C. Botten, "Multipole Method for Microstructured Optical Fibers: 1. Formulation," *Journal of the Optical Society of America B—Optical Physics*, **19**:2322–2330, 2002.
35. R. C. McPhedran, L. C. Botten, A. A. Asatryan, N. A. P. Nicorovici, P. A. Robinson, and C. M. De Sterke, "Calculation of Electromagnetic Properties of Regular and Random Arrays of Metallic and Dielectric Cylinders," *Physical Review E*, **60**:7614–7617, 1999.
36. P. J. Roberts and T. J. Shepherd, "The Guidance Properties of Multi-Core Photonic Crystal Fibers," *Journal of Optics a—Pure and Applied Optics*, **3**:S1–S8, 2001.
37. A. Ferrando, E. Silvestre, J. J. Miret, P. Andres, and M. V. Andres, "Full-Vector Analysis of a Realistic Photonic Crystal Fiber," *Optics Letters*, **24**:276–278, Mar. 1999.
38. D. Mogilevtsev, T. A. Birks, and P. St.J. Russell, "Localized Function Method for Modeling Defect Modes in 2-D Photonic Crystals," *Journal of Lightwave Technology*, **17**:2078–2081, Nov. 1999.
39. A. W. Snyder and J. D. Love, *Optical Waveguide Theory*, London: Chapman & Hall, 1983.
40. T. A. Birks, D. M. Bird, T. D. Hedley, J. M. Pottage, and P. St.J. Russell, "Scaling Laws and Vector Effects in Bandgap-Guiding Fibres," *Optics Express*, **12**:69–74, Jan. 2004.
41. R. D. Meade, A. M. Rappe, K. D. Brommer, J. D. Joannopoulos, and O. L. Alerhand, "Accurate Theoretical Analysis of Photonic Band-Gap Materials," *Physical Review B*, **48**:8434–8437, 1993.
42. G. J. Pearce, T. D. Hedley, and D. M. Bird, "Adaptive Curvilinear Coordinates in a Plane-Wave Solution of Maxwell's Equations in Photonic Crystals," *Physical Review B*, **71**:195108, 2005.
43. J. M. Pottage, D. M. Bird, T. D. Hedley, T. A. Birks, J. C. Knight, P. St.J. Russell, and P. J. Roberts, "Robust Photonic Band Gaps for Hollow Core Guidance in PCF Made from High Index Glass," *Optics Express*, **11**:2854–2861, Nov. 2003.
44. T. M. Monro, D. J. Richardson, N. G. R. Broderick, and P. J. Bennett, "Holey Optical Fibers: An Efficient Modal Model," *Journal of Lightwave Technology*, **17**:1093–1102, 1999.
45. C. T. Chan, Q. L. Yu, and K. M. Ho, "Order N Spectral Method for Electromagnetic Waves," *Physical Review B*, **51**:16635–16642, 1995.
46. V. Dangui, M. J. F. Digonnet, and G. S. Kino, "A Fast and Accurate Numerical Tool to Model the Modal Properties of Photonic-Bandgap Fibers," *Optics Express*, **14**:2979–2993, Apr. 2006.
47. C. Mias, J. P. Webb, and R. L. Ferrari, "Finite Element Modelling of Electromagnetic Waves in Doubly and Triply Periodic Structures," *IEEE Proceedings—Optoelectronics*, **146**:111–118, 1999.
48. A. Hochman and Y. Leviatan, "Analysis of Strictly Bound Modes in Photonic Crystal Fibers by Use of a Source-Model Technique," *Journal of the Optical Society of America A—Optics Image Science and Vision*, **21**:1073–1081, 2004.
49. J. C. Knight, T. A. Birks, P. St.J. Russell, and J. P. de Sandro, "Properties of Photonic Crystal Fiber and the Effective Index Model," *Journal of the Optical Society of America A—Optics Image Science and Vision*, **15**:748–752, Mar. 1998.
50. J. C. Knight, J. Broeng, T. A. Birks, and P. St.J. Russell, "Photonic Band Gap Guidance in Optical Fibers," *Science*, **282**:1476–1478, Nov. 1998.
51. F. Benabid, J. C. Knight, G. Antonopoulos, and P. St.J. Russell, "Stimulated Raman Scattering in Hydrogen-Filled Hollow-Core Photonic Crystal Fiber," *Science*, **298**:399–402, Oct. 2002.
52. F. Couny, F. Benabid, and P. S. Light, "Large-Pitch Kagome-Structured Hollow-Core Photonic Crystal Fiber," *Optics Letters*, **31**:3574–3576, Dec. 2006.
53. B. J. Mangan, J. Arriaga, T. A. Birks, J. C. Knight, and P. St.J. Russell, "Fundamental-Mode Cutoff in a Photonic Crystal Fiber with a Depressed-Index Core," *Optics Letters*, **26**:1469–1471, Oct. 2001.
54. W. N. MacPherson, M. J. Gander, R. McBride, J. D. C. Jones, P. M. Blanchard, J. G. Burnett, A. H. Greenaway, et al., "Remotely Addressed Optical Fibre Curvature Sensor Using Multicore Photonic Crystal Fibre," *Optics Communications*, **193**:97–104, Jun. 2001.
55. J. Limpert, N. D. Robin, I. Manek-Honninger, F. Salin, F. Roser, A. Liem, T. Schreiber, et al., "High-Power Rod-Type Photonic Crystal Fiber Laser," *Optics Express*, **13**:1055–1058, Feb. 2005.
56. W. J. Wadsworth, R. M. Percival, G. Bouwmans, J. C. Knight, and P. St.J. Russell, "High Power Air-Clad Photonic Crystal Fibre Laser," *Optics Express*, **11**:48–53, Jan. 2003.
57. M. T. Myaing, J. Y. Ye, T. B. Norris, T. Thomas, J. R. Baker, W. J. Wadsworth, G. Bouwmans, J. C. Knight, and P. St.J. Russell, "Enhanced Two-Photon Biosensing with Double-Clad Photonic Crystal Fibers," *Optics Letters*, **28**:1224–1226, Jul. 2003.

58. P. J. Roberts, F. Couny, H. Sabert, B. J. Mangan, D. P. Williams, L. Farr, M. W. Mason, et al., "Ultimate Low Loss of Hollow-Core Photonic Crystal Fibres," *Optics Express*, **13**:236–244, Jan. 2005.
59. F. Benabid, G. Bouwmans, J. C. Knight, P. St.J. Russell, and F. Couny, "Ultrahigh Efficiency Laser Wavelength Conversion in a Gas-Filled Hollow Core Photonic Crystal Fiber by Pure Stimulated Rotational Raman Scattering in Molecular Hydrogen," *Physical Review Letters*, **93**:123903 Sep. 2004.
60. G. Humbert, J. C. Knight, G. Bouwmans, P. St.J. Russell, D. P. Williams, P. J. Roberts, and B. J. Mangan, "Hollow Core Photonic Crystal Fibers for Beam Delivery," *Optics Express*, **12**:1477–1484, Apr. 2004.
61. P. J. Roberts, F. Couny, H. Sabert, B. J. Mangan, T. A. Birks, J. C. Knight, and P. St.J. Russell, "Loss in Solid-Core Photonic Crystal Fibers Due to Interface Roughness Scattering," *Optics Express*, **13**:7779–7793, Oct. 2005.
62. J. A. West, C. Smith, N. F. Borrelli, D. C. Allan, and K. W. Koch, "Surface Modes in Air-Core Photonic Band-Gap Fibers," *Optics Express*, **12**:1485–1496, 2004.
63. A. Argyros, T. A. Birks, S. G. Leon-Saval, C. M. B. Cordeiro, F. Luan, and P. St.J. Russell, "Photonic Bandgap with an Index Step of One Percent," *Optics Express*, **13**:309–314, Jan. 2005.
64. J. C. Knight, F. Luan, G. J. Pearce, A. Wang, T. A. Birks, and D. M. Bird, "Solid Photonic Bandgap Fibres and Applications," *Japanese Journal of Applied Physics Part 1—Regular Papers Short Notes & Review Papers*, **45**:6059–6063, Aug. 2006.
65. F. Luan, A. K. George, T. D. Hedley, G. J. Pearce, D. M. Bird, J. C. Knight, and P. St.J. Russell, "All-Solid Photonic Bandgap Fiber," *Optics Letters*, **29**:2369–2371, Oct. 2004.
66. N. M. Litchinitser, S. C. Dunn, B. Usner, B. J. Eggleton, T. P. White, R. C. McPhedran, and C. M. De Sterke, "Resonances in Microstructured Optical Waveguides," *Optics Express*, **11**:1243–1251, 2003.
67. A. Argyros, T. A. Birks, S. G. Leon-Saval, C. M. B. Cordeiro, and P. St.J. Russell, "Guidance Properties of Low-Contrast Photonic Bandgap Fibres," *Optics Express*, **13**:2503–2511, Apr. 2005.
68. G. Bouwmans, L. Bigot, Y. Quiquempois, F. Lopez, L. Provino, and M. Douay, "Fabrication and Characterization of an All-Solid 2D Photonic Bandgap Fiber with a Low-Loss Region (20 dB/km) around 1550 nm," *Optics Express*, **13**:8452–8459, 2005.
69. M. J. Steel, T. P. White, C. M. De Sterke, R. C. McPhedran, and L. C. Botten, "Symmetry and Degeneracy in Microstructured Optical Fibers," *Optics Letters*, **26**:488–490, 2001.
70. A. Ortigosa-Blanch, J. C. Knight, W. J. Wadsworth, J. Arriaga, B. J. Mangan, T. A. Birks, and P. St.J. Russell, "Highly Birefringent Photonic Crystal Fibers," *Optics Letters*, **25**:1325–1327, Sep. 2000.
71. H. Kubota, S. Kawanishi, S. Koyanagi, M. Tanaka, and S. Yamaguchi, "Absolutely Single Polarization Photonic Crystal Fiber," *IEEE Photonics Technology Letters*, **16**:182–184, Jan. 2004.
72. C. Kerbage, P. Steinvurzel, P. Reyes, P. S. Westbrook, R. S. Windeler, A. Hale, and B. J. Eggleton, "Highly Tunable Birefringent Microstructured Optical Fiber," *Optics Letters*, **27**:842–844, May. 2002.
73. P. J. Roberts, D. P. Williams, H. Sabert, B. J. Mangan, D. M. Bird, T. A. Birks, J. C. Knight, and P. St.J. Russell, "Design of Low-Loss and Highly Birefringent Hollow-Core Photonic Crystal Fiber," *Optics Express*, **14**:7329–7341, Aug. 2006.
74. D. R. Chen and L. F. Shen, "Ultrahigh Birefringent Photonic Crystal Fiber with Ultralow Confinement Loss," *IEEE Photonics Technology Letters*, **19**:185–187, Jan.–Feb. 2007.
75. A. Michie, J. Canning, K. Lyytikainen, M. Aslund, and J. Digweed, "Temperature Independent Highly Birefringent Photonic Crystal Fibre," *Optics Express*, **12**:5160–5165, Oct. 2004.
76. T. Martynkien, M. Szpulak, and W. Urbanczyk, "Modeling and Measurement of Temperature Sensitivity in Birefringent Photonic Crystal Hole Fibers," *Applied Optics*, **44**:7780–7788, Dec. 2005.
77. D. H. Kim and J. U. Kang, "Sagnac Loop Interferometer Based on Polarization Maintaining Photonic Crystal Fiber with Reduced Temperature Sensitivity," *Optics Express*, **12**:4490–4495, Sep. 2004.
78. W. H. Reeves, J. C. Knight, P. St.J. Russell, and P. J. Roberts, "Demonstration of Ultra-Flattened Dispersion in Photonic Crystal Fibers," *Optics Express*, **10**:609–613, Jul. 2002.
79. W. H. Reeves, D. V. Skryabin, F. Biancalana, J. C. Knight, P. St.J. Russell, F. G. Omenetto, A. Efimov, and A. J. Taylor, "Transformation and Control of Ultra-Short Pulses in Dispersion-Engineered Photonic Crystal Fibres," *Nature*, **424**:511–515, Jul. 2003.
80. G. Bouwmans, F. Luan, J. C. Knight, P. St.J. Russell, L. Farr, B. J. Mangan, and H. Sabert, "Properties of a Hollow-Core Photonic Bandgap Fiber at 850 nm Wavelength," *Optics Express*, **11**:1613–1620, Jul. 2003.
81. K. Kurokawa, K. Tajima, and K. Nakajima, "10-GHz 0.5-ps Pulse Generation in 1000-nm Band in PCF for High-Speed Optical Communication," *Journal of Lightwave Technology*, **25**:75–78, Jan. 2007.

82. K. Kurokawa, K. Tajima, K. Tsujikawa, and K. Nakagawa, "Penalty-Free Dispersion-Managed Soliton Transmission Over a 100-km Low-Loss PCF," *Journal of Lightwave Technology*, **24**:32–37, 2006.
83. G. J. Pearce, J. M. Pottage, D. M. Bird, P. J. Roberts, J. C. Knight, and P. St.J. Russell, "Hollow-Core PCF for Guidance in the Mid to Far Infrared," *Optics Express*, **13**:6937–6946, Sep. 2005.
84. J. D. Shephard, W. N. MacPherson, R. R. J. Maier, J. D. C. Jones, D. P. Hand, M. Mohebbi, A. K. George, P. J. Roberts, and J. C. Knight, "Single-Mode Mid-IR Guidance in a Hollow-Core Photonic Crystal Fiber," *Optics Express*, **13**:7139–7144, Sep. 2005.
85. P. J. Roberts, D. P. Williams, B. J. Mangan, H. Sabert, F. Couny, W. J. Wadsworth, T. A. Birks, J. C. Knight, and P. St.J. Russell, "Realizing Low Loss Air Core Photonic Crystal Fibers by Exploiting an Antiresonant Core Surround," *Optics Express*, **13**:8277–8285, Oct. 2005.
86. M. D. Nielsen, N. A. Mortensen, M. Albertsen, J. R. Folkenberg, A. Bjarklev, and D. Bonacinni, "Predicting Macrobending Loss for Large-Mode Area Photonic Crystal Fibers," *Optics Express*, **12**:1775–1779, 2004.
87. J. Laegsgaard, N. A. Mortensen, J. Riishede, and A. Bjarklev, "Material Effects in Air-Guiding Photonic Bandgap Fibers," *Journal of the Optical Society of America B—Optical Physics*, **20**:2046–2051, 2003.
88. M. Onishi, T. Okuno, T. Kashiwada, S. Ishikawa, N. Akasaka, and M. Nishimura, "Highly Nonlinear Dispersion-Shifted Fibers and Their Application to Broadband Wavelength Converter," *Optical Fiber Technology*, **4**:204–214, 1998.
89. P. Petropoulos, H. Ebendorff-Heidepriem, V. Finazzi, R. Moore, K. Frampton, D. J. Richardson, and M. Monro, "Highly Nonlinear and Anomously Dispersive Lead Silicate Glass Hole Fibers," *Optics Express*, **11**:3568–3573, 2003.
90. F. Luan, J. C. Knight, P. St.J. Russell, S. Campbell, D. Xiao, D. T. Reid, B. J. Mangan, D. P. Williams, and P. J. Roberts, "Femtosecond Soliton Pulse Delivery at 800 nm Wavelength in Hollow-Core Photonic Band-gap Fibers," *Optics Express*, **12**:835–840, Mar. 2004.
91. C. J. Hensley, D. G. Ouzounov, A. L. Gaeta, N. Venkataraman, M. T. Gallagher, and K. W. Koch, "Silica-Glass Contribution to the Effective Nonlinearity of Hollow-Core Photonic Band-Gap Fibers," *Optics Express*, **15**:3507–3512, Mar. 2007.
92. A. Efimov, A. J. Taylor, F. G. Omenetto, A. V. Yulin, N. Y. Joly, F. Biancalana, D. V. Skryabin, J. C. Knight, and P. St.J. Russell, "Time-Spectrally-Resolved Ultrafast Nonlinear Dynamics in Small-Core Photonic Crystal Fibers: Experiment and Modelling," *Optics Express*, **12**:6498–6507, Dec. 2004.
93. G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed., Academic Press, San Diego, CA, 2007.
94. R. R. Alfano, *The Supercontinuum Laser Source*, Springer-Verlag, New York, 1989.
95. S. V. Chernikov, Y. Zhu, J. R. Taylor, and V. P. Gapontsev, "Supercontinuum Self-Q-Switched Ytterbium Fiber Laser," *Optics Letters*, **22**:298–300, 1997.
96. J. M. Dudley, G. Genty, and S. Coen, "Supercontinuum Generation in Photonic Crystal Fiber," *Reviews of Modern Physics*, **78**:1135–1184, Oct.–Dec. 2006.
97. W. J. Wadsworth, A. Ortigosa-Blanch, J. C. Knight, T. A. Birks, T. P. M. Man, and P. St.J. Russell, "Supercontinuum Generation in Photonic Crystal Fibers and Optical Fiber Tapers: A Novel Light Source," *Journal of the Optical Society of America B—Optical Physics*, **19**:2148–2155, Sep. 2002.
98. G. Humbert, W. J. Wadsworth, S. G. Leon-Saval, J. C. Knight, T. A. Birks, P. St.J. Russell, M. J. Lederer, et al., "Supercontinuum Generation System for Optical Coherence Tomography Based on Tapered Photonic Crystal Fibre," *Optics Express*, **14**:1596–1603, Feb. 2006.
99. I. Hartl, X. D. Li, C. Chudoba, R. K. Ghanta, T. H. Ko, J. G. Fujimoto, J. K. Ranka, and R. S. Windeler, "Ultra-high-Resolution Optical Coherence Tomography Using Continuum Generation in an Air-Silica Microstructure Optical Fiber," *Optics Letters*, **26**:608–610, May. 2001.
100. H. Hundertmark, D. Kracht, D. Wandt, C. Fallnich, V. V. R. K. Kumar, A. K. George, J. C. Knight, and P. St.J. Russell, "Supercontinuum Generation with 200 pJ Laser Pulses in an Extruded SF6 Fiber at 1560 nm," *Optics Express*, **11**:3196–3201, Dec. 2003.
101. R. Holzwarth, T. Udem, T. W. Haensch, J. C. Knight, W. J. Wadsworth, and P. St.J. Russell, "Optical Frequency Synthesizer for Precision Spectroscopy," *Physical Review Letters*, **85**:2264–2267, Sep. 2000.
102. S. G. Leon-Saval, T. A. Birks, W. J. Wadsworth, P. St.J. Russell, and M. W. Mason, "Supercontinuum Generation in Submicron Fibre Waveguides," *Optics Express*, **12**:2864–2869, Jun. 2004.
103. S. Coen, A. H. L. Chau, R. Leonhardt, J. D. Harvey, J. C. Knight, W. J. Wadsworth, and P. St.J. Russell, "Supercontinuum Generation by Stimulated Raman Scattering and Parametric Four-Wave Mixing in Photonic Crystal Fibers," *Journal of the Optical Society of America B—Optical Physics*, **19**:753–764, Apr. 2002.

104. W. J. Wadsworth, N. Joly, J. C. Knight, T. A. Birks, F. Biancalana, and P. St.J. Russell, "Supercontinuum and Four-Wave Mixing with Q-Switched Pulses in Endlessly Single-Mode Photonic Crystal Fibres," *Optics Express*, **12**:299–309, Jan. 2004.
105. www.fianium.com.
106. F. G. Omenetto, N. A. Wolchover, M. R. Wehner, M. Ross, A. Efimov, A. J. Taylor, V. Kumar, et al., "Spectrally Smooth Supercontinuum from 350 nm to 3 μ m in Sub-Centimeter Lengths of Soft-Glass Photonic Crystal Fibers," *Optics Express*, **14**:4928–4934, May 2006.
107. M. Yu, C. J. McKinstrie, and G. P. Agrawal, "Modulational Instabilities in Dispersion-Flattened Fibers," *Physical Review E*, **52**:1072–1080, 1995.
108. A. Y. H. Chen, G. K. L. Wong, S. G. Murdoch, R. Leonhardt, J. D. Harvey, J. C. Knight, W. J. Wadsworth, and P. St.J. Russell, "Widely Tunable Optical Parametric Generation in a Photonic Crystal Fiber," *Optics Letters*, **30**:762–764, Apr. 2005.
109. Y. J. Deng, Q. Lin, F. Lu, G. P. Agrawal, and W. H. Knox, "Broadly Tunable Femtosecond Parametric Oscillator Using a Photonic Crystal Fiber," *Optics Letters*, **30**:1234–1236, May 2005.
110. J. Lasri, P. Devgan, R. Y. Tang, J. E. Sharping, and P. Kumar, "A Microstructure-Fiber-Based 10-GHz Synchronized Tunable Optical Parametric Oscillator in the 1550-nm Regime," *IEEE Photonics Technology Letters*, **15**:1058–1060, Aug. 2003.
111. J. E. Sharping, M. Fiorentino, P. Kumar, and R. S. Windeler, "Optical Parametric Oscillator Based on Four-Wave Mixing in Microstructure Fiber," *Optics Letters*, **27**:1675–1677, Oct. 2002.
112. J. D. Harvey, R. Leonhardt, S. Coen, G. K. L. Wong, J. C. Knight, W. J. Wadsworth, and P. St.J. Russell, "Scalar Modulation Instability in the Normal Dispersion Regime by Use of a Photonic Crystal Fiber," *Optics Letters*, **28**:2225–2227, Nov. 2003.
113. J. Fulconis, O. Alibart, W. J. Wadsworth, P. St.J. Russell, and J. G. Rarity, "High Brightness Single Mode Source of Correlated Photon Pairs Using a Photonic Crystal Fiber," *Optics Express*, **13**:7572–7582, Sep. 2005.
114. D. V. Skryabin, F. Luan, J. C. Knight, and P. St.J. Russell, "Soliton Self-Frequency Shift Cancellation in Photonic Crystal Fibers," *Science*, **301**:1705–1708, Sep. 2003.
115. N. Y. Joly, F. G. Omenetto, A. Efimov, A. J. Taylor, J. C. Knight, and P. St.J. Russell, "Competition Between Spectral Splitting and Raman Frequency Shift in Negative-Dispersion Slope Photonic Crystal Fiber," *Optics Communications*, **248**:281–285, Apr. 2005.
116. V. Laude, A. Khelif, S. Benchane, M. Wilm, T. Sylvestre, B. Kibler, A. Mussot, J. M. Dudley, and H. Maillotte, "Phononic Band-Gap Guidance of Acoustic Modes in Photonic Crystal Fibers," *Physical Review B*, **71**:045107, 2005.
117. S. Guenneau and A. B. Movchan, "Analysis of Elastic Band Structures for Oblique Incidence," *Archive for Rational Mechanics and Analysis*, **171**:129–150, 2004.
118. P. St.J. Russell, E. Marin, A. Diez, S. Guenneau, and A. B. Movchan, "Sonic Band Gaps in PCF Preforms: Enhancing the Interaction of Sound and Light," *Optics Express*, **11**:2555–2560, Oct. 2003.
119. P. St.J. Russell, "Light in a Tight Space: Enhancing Matter-Light Interactions Using Photonic Crystals," in *Conference on Nonlinear Optics*, Hawaii, pp. 377–379, 2002.
120. P. Dainese, P. St.J. Russell, N. Joly, J. C. Knight, G. S. Wiederhecker, H. L. Fragnito, V. Laude, and A. Khelif, "Stimulated Brillouin Scattering from Multi-GHz-Guided Acoustic Phonons in Nanostructured Photonic Crystal Fibres," *Nature Physics*, **2**:388–392, Jun. 2006.
121. P. Dainese, P. St.J. Russell, G. S. Wiederhecker, N. Joly, H. L. Fragnito, V. Laude, and A. Khelif, "Raman-Like Light Scattering from Acoustic Phonons in Photonic Crystal Fiber," *Optics Express*, **14**:4141–4150, May 2006.
122. F. Benabid, F. Couny, J. C. Knight, T. A. Birks, and P. St.J. Russell, "Compact, Stable and Efficient All-Fibre Gas Cells Using Hollow-Core Photonic Crystal Fibres," *Nature*, **434**:488–491, Mar. 2005.
123. S. G. Leon-Saval, T. A. Birks, N. Y. Joly, A. K. George, W. J. Wadsworth, G. Kakarantzas, and P. St.J. Russell, "Splice-Free Interfacing of Photonic Crystal Fibers," *Optics Letters*, **30**:1629–1631, Jul. 2005.
124. S. G. Leon-Saval, T. A. Birks, J. Bland-Hawthorn, and M. Englund, "Multimode Fiber Devices with Single-Mode Performance," *Optics Letters*, **30**:2545–2547, 2005.
125. T. A. Birks, G. Kakarantzas, P. St.J. Russell, and D. F. Murphy, "Photonic Crystal Fiber Devices," *Proceedings of the Society of Photo-Instrumentation Engineers*, **4943**:142–151, 2002.

126. W. J. Wadsworth, A. Witkowska, S. Leon-Saval, and T. A. Birks, "Hole Inflation and Tapering of Stock Photonic Crystal Fibers," *Optics Express*, **13**:6541–6549, 2005.
127. G. Kakarantzas, T. E. Dimmick, T. A. Birks, R. Le Roux, and P. St.J. Russell, "Miniature All-Fiber Devices Based on CO₂ Laser Microstructuring of Tapered Fibers," *Optics Letters*, **26**:1137–1139, Aug. 2001.
128. G. Kakarantzas, T. A. Birks, and P. St.J. Russell, "Structural Long-Period Gratings in Photonic Crystal Fibers," *Optics Letters*, **27**:1013–1015, Jun. 2002.
129. G. Kakarantzas, A. Ortigosa-Blanch, T. A. Birks, P. St.J. Russell, L. Farr, F. Couny, and B. J. Mangan, "Structural Rocking Filters in Highly Birefringent Photonic Crystal Fiber," *Optics Letters*, **28**:158–160, Feb. 2003.

INFRARED FIBERS

James A. Harrington

*Rutgers University
Piscataway, New Jersey*

12.1 INTRODUCTION

Infrared (IR) optical fibers may be defined as fiber optics transmitting radiation with wavelengths greater than approximately $2\text{ }\mu\text{m}$. The first IR fibers were fabricated in the mid-1960s from chalcogenide glasses such as arsenic trisulfide and had losses in excess of 10 dB/m .¹ During the mid-1970s, the interest in developing an efficient and reliable IR fiber for short-haul applications increased, partly in response to the need for a fiber to link broadband, long-wavelength radiation to remote photodetectors in military sensor applications. In addition, there was an ever-increasing need for a flexible fiber delivery system for transmitting CO_2 laser radiation in surgical applications. Around 1975, a variety of IR materials and fibers were developed to meet these needs. These included the heavy metal fluoride glass (HMFG) and polycrystalline fibers as well as hollow rectangular waveguides. While none of these fibers had physical properties even approaching those of conventional silica fibers, they were nevertheless useful in lengths less than 2 to 3 m for a variety of IR sensor and power delivery applications.²

IR fiber optics may logically be divided into three broad categories: glass, crystalline, and hollow waveguides. These categories may be further subdivided based on fiber material, structure, or both, as shown in Table 1. Over the past 30 years many novel IR fibers have been made in an effort to fabricate a fiber optic with properties as close as possible to those of silica, but only a relatively small number have survived. A good source of general information on these various IR fiber types may be found in the literature.²⁻⁶ In this review only the best, most viable, and, in most cases, commercially available IR fibers are discussed. In general, both the optical and mechanical properties of IR fibers remain inferior to those of silica fibers, and therefore the use of IR fibers is still limited primarily to nontelecommunication, short-haul applications requiring only tens of meters of fiber rather than the kilometer lengths common to telecommunication applications. The short-haul nature of IR fibers results from the fact that most IR fibers have losses in the range of a few decibels per meter. An exception is fluoride glass fibers, which can have losses as low as a few decibels per kilometer. In addition, IR fibers are much weaker than silica fiber and, therefore, more fragile. These deleterious features have slowed the acceptance of IR fibers and restricted their use today to applications in chemical sensing, thermometry, and laser power delivery.

A key feature of current IR fibers is their ability to transmit longer wavelengths than most oxide glass fibers can. In some cases the transmittance of the fiber can extend well beyond $20\text{ }\mu\text{m}$, but most applications do not require the delivery of radiation longer than about $12\text{ }\mu\text{m}$. In Fig. 1 we give the

TABLE 1 Categories of IR Fibers with a Common Example to Illustrate Each Subcategory

Main	Subcategory	Examples
Glass	Heavy metal fluoride (HMFG)	ZrF ₄ -BaF ₂ -LaF ₃ -AlF ₃ -NAF(ZBLAN)
	Germanate	GeO ₂ -PbO
	Chalcogenide	As ₂ S ₃ and AsGeTeSe
Crystal	Polycrystalline (PC)	AgBrCl
	Single crystal (SC)	Sapphire
Hollow waveguide	Metal/dielectric film	Hollow glass waveguide
	Refractive index <1	Hollow sapphire at 10.6 μm

attenuation values for some of the most common IR fibers as listed in Table 1. From the data it is clear that there is a wide variation in range of transmission for the different IR fibers and that there is significant extrinsic absorption that degrades the overall optical response. Most of these extrinsic bands can be attributed to various impurities, but, in the case of the hollow waveguides, they are due to interference effects resulting from the thin film coatings used to make the guides.

Some of the other physical properties of IR fibers are listed in Table 2. For comparison, the properties of silica fibers are also listed. The data in Table 2 and in Fig. 1 reveal that, compared to silica, IR fibers usually have higher losses, larger refractive indices and dn/dT values, lower melting or softening points, and greater thermal expansion. For example, chalcogenide and polycrystalline Ag halide fibers have refractive indices greater than 2. This means that the Fresnel loss exceeds 20 percent for two fiber ends. The higher dn/dT and low melting or softening point lead to thermal lensing and low laser-induced damage thresholds for some of the fibers. Finally, a number of these fibers do not have cladding analogous to clad oxide glass fibers. Nevertheless, core-only IR fibers such as sapphire and chalcogenide fibers can still be useful because their refractive indices are sufficiently high. For these high-index fibers, the energy is largely confined to the core of the fiber as long as the unprotected fiber core does not come in contact with an absorbing medium.¹²

The motivation to develop a viable IR fiber stems from many proposed applications. A summary of the most important current and future applications and the associated candidate IR fiber that will best meet each need is given in Table 3. We may note several trends from this table. The first is that

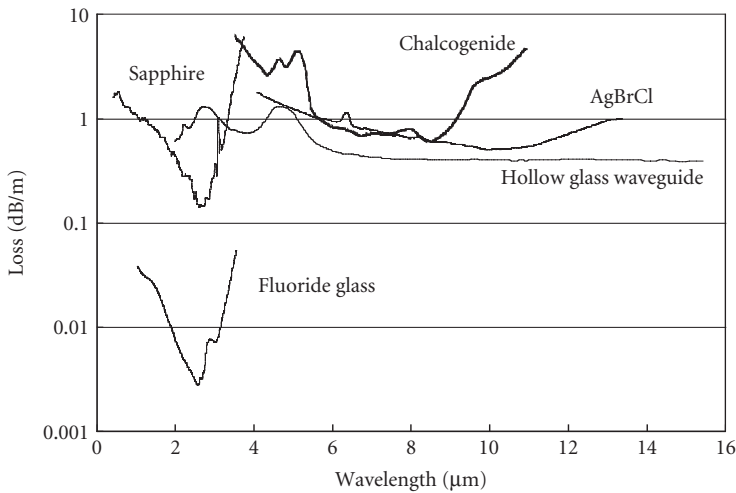


FIGURE 1 Composite loss spectra for some common IR fiber optics: ZBLAN fluoride glass,⁷ SC sapphire,⁸ chalcogenide glass,⁹ PC AgBrCl,¹⁰ and hollow glass waveguide.¹¹

TABLE 2 Selected Physical Properties of Key IR Fibers Compared to Conventional Silica Fiber

Property	Glass			Crystal		Hollow
	Silica	HMFG ZBLAN	Chalcogenide AsGeSeTe	PC AgBrCl	SC Sapphire	Hollow Silica Waveguide
Glass transition or melting point, °C	1175	265	245	412	2030	150 (usable T)
Thermal conductivity, W/m °C	1.38	0.628	0.2	1.1	36	1.38
Thermal expansion coefficient, 10^{-6} °C ⁻¹	0.55	17.2	15	30	5	0.55
Young's modulus, GPa	70.0	58.3	21.5	0.14	430	70.0
Density, g/cm ³	2.20	4.33	4.88	6.39	3.97	2.20
Refractive index	1.455	1.499	2.9	2.2	1.71	NA
(λ , μm)	(0.70)	(0.589)	(10.6)	(10.6)	(3.0)	
dn/dT , 10^{-5} °C ⁻¹	+1.2	-1.5	+10	-1.5	+1.4	NA
(λ , μm)	(1.06)	(1.06)	(10.6)	(10.6)	(1.06)	
Fiber transmission range, μm	0.24–2.0	0.25–4.0	4–11	3–16	0.5–3.1	0.9–25
Loss* at 2.94 μm , dB/m	~800	0.08	5	3	0.4	0.5
Loss* at 10.6 μm , dB/m	NA	NA	2	0.5	NA	0.4

*Typical measured loss.

NA = not applicable.

TABLE 3 Examples of IR Fiber Candidates for Various Sensor and Power Delivery Applications

Application	Comments	Suitable IR fibers
Fiber-optic chemical sensors	Evanescent wave principle—liquids	AgBrCl, sapphire, chalcogenide, HMFG
	Hollow core waveguides—gases	Hollow glass waveguides
Radiometry	Blackbody radiation, temperature measurements	Hollow glass waveguides, AgBrCl, chalcogenide, sapphire
Er:YAG laser power delivery	3- μm transmitting fibers with high damage threshold	Hollow glass waveguides, sapphire, germanate glass
CO ₂ laser power delivery	10- μm transmitting fibers with high damage threshold	Hollow glass waveguides
Thermal imaging	Coherent bundles	HMFG, chalcogenide
Fiber amplifiers and lasers	Doped IR glass fibers	HMFG, chalcogenide

hollow waveguides are an ideal candidate for laser power delivery at all IR laser wavelengths. The air core of these special fibers or waveguides gives an inherent advantage over solid-core fibers, whose damage threshold is frequently very low for these IR-transmissive materials. The high refractive index of chalcogenide fibers is ideal for chemical sensing via evanescent wave coupling of a small portion of the light from the core into an IR-absorbing medium. For the measurement of temperature through the simple transmission of blackbody radiation, IR fibers that transmit beyond about 8 μm , such as the Ag halide, chalcogenide, and hollow waveguides, are excellent candidates for use in measuring temperatures below 50°C. This is because the peak for room-temperature blackbody radiation is about 10 μm .

12.2 NONOXIDE AND HEAVY-METAL OXIDE GLASS IR FIBERS

There are two IR-transmitting glass fiber systems that are relatively similar to conventional silica-containing glass fibers. One is the HMFG and the other is heavy-metal germanate glass fibers based on GeO₂. The germanate glass fibers generally do not contain fluoride compounds; instead, they

contain heavy metal oxides to shift the IR absorption edge to longer wavelengths. The advantage of germanate fibers over HMFG fibers is that germanate glass has a higher glass transition temperature and, therefore, higher laser damage thresholds. But the level of loss for the HMFG fibers is lower. Finally, chalcogenide glass fibers made from chalcogen elements such as As, Ge, S, and Te contain no oxides or halides, making them a good choice for nonlaser power delivery applications.

HMFG Fibers

Poulain et al.¹³ discovered HMFGs or fluoride glasses accidentally in 1975 at the University of Rennes. In general, the typical fluoride glass has a glass transition temperature T_g four times less than that of silica, is considerably less stable than silica, and has failure strains of only a few percent compared to greater than 5 percent for silica. While an enormous number of multicomponent fluoride glass compositions have been fabricated, comparably few have been drawn into fiber. This is because the temperature range for fiber drawing is normally too small in most HMFGs to permit fiberization of the glass. The most popular HMFGs for fabrication into fibers are the fluorozirconate and fluoroaluminate glasses, of which the most common are $\text{ZrF}_4\text{-BaF}_2\text{-LaF}_3\text{-AlF}_3\text{-NaF}$ (ZBLAN) and $\text{AlF}_3\text{-ZrF}_4\text{-BaF}_2\text{-CaF}_2\text{-YF}_3$, respectively. The key physical properties of these glasses are summarized in Table 4. An important feature of the fluoroaluminate glass is its higher T_g , which largely accounts for the higher laser damage threshold for the fluoroaluminate glasses compared to ZBLAN at the Er:YAG laser wavelength of 2.94 μm .

The fabrication of HMFG fiber is similar to any glass fiber drawing technology except that the preforms are made using some type of melt-forming method rather than by a vapor deposition process as is common with silica fibers. Specifically, a casting method based on first forming a clad glass tube and then adding the molten core glass is used to form either multimode or single-mode fluorozirconate fiber preforms. The cladding tube is made either by a rotational casting technique in which the tube is spun in a metal mold or by merely inverting and pouring out most of the molten cladding glass contained in a metal mold to form a tube.¹⁴ The cladding tube is then filled with a higher-index core glass. Other preform fabrication techniques include rod-in-tube and crucible techniques. The fluoroaluminate fiber preforms have been made using an unusual extrusion technique in which core and cladding glass plates are extruded into a core-clad preform.¹⁵ All methods, however, involve fabrication from the melted glass rather than from the more pristine technique of vapor deposition used to form SiO_2 -based fibers. This process creates inherent problems such as the formation of bubbles, core-cladding interface irregularities, and small preform sizes. Most HMFG fiber drawing is done using preforms rather than the crucible method. A ZBLAN preform is drawn at about 310°C in a controlled atmosphere (to minimize contamination by moisture or oxygen impurities, which can significantly weaken the fiber) using a narrow heat zone compared to silica. Either ultraviolet (UV) acrylate or Teflon coatings are applied to the fiber. In the case of Teflon, heat-shrink Teflon (FEP) is generally applied to the glass preform prior to the draw.

The attenuation in HMFG fibers is predicted to be about 10 times less than that for silica fibers.¹⁶ Based on extrapolations of the intrinsic losses resulting from Rayleigh scattering and multiphonon

TABLE 4 Fluorozirconate vs. Fluoroaluminate Glasses

Property	Fluorozirconate (ZBLAN)	Fluoroaluminate ($\text{AlF}_3\text{-ZrF}_4\text{-BaF}_2\text{-CaF}_2\text{-YF}_3$)
Glass transition temperature, °C	265	400
Durability	Medium	Excellent
Loss at 2.94 μm , dB/m	0.01	0.1
Er:YAG laser peak output energy, mJ	300 (300- μm core)	850 (500- μm core)

Comparison between fluorozirconate and fluoroaluminate glasses of some key properties that relate to laser power transmission and durability of the two HMFG fibers. Other physical properties are relatively similar.

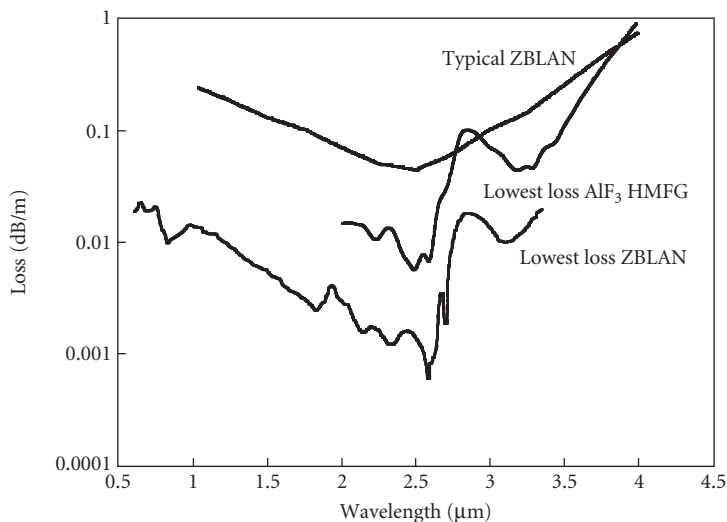


FIGURE 2 Losses in the best BTRL⁷ and typical (Infrared Fiber Systems, Silver Spring, Maryland) ZBLAN fluoride glass fibers compared to those for fluoroaluminate glass fibers.¹⁵

absorption, the minimum in the loss curves or V-curves is projected to be about 0.01 dB/km at 2.55 μm . Recent refinements of the scattering loss have modified this value slightly to be 0.024 dB/km, or about 8 times less than that for silica fiber.⁷ In practice, however, extrinsic loss mechanisms still dominate fiber loss. In Fig. 2, losses for two ZBLAN fibers are shown. The data from British Telecom (BTRL) represents state-of-the-art fiber 110 m in length.⁷ The other curve is more typical of commercially available (Infrared Fiber Systems, Silver Spring, Maryland) ZBLAN fiber. More recently commercially available fiber (IRphotonics, Montreal, CA) has become available with a loss of 0.02 dB/m at 3 μm .¹⁷ The lowest measured loss for a BTRL 60-m-long fiber is 0.45 dB/km at 2.3 μm . Some of the extrinsic absorption bands that contribute to the total loss shown in Fig. 2 for the BTRL fiber are Ho^{3+} (0.64 and 1.95 μm), Nd^{3+} (0.74 and 0.81 μm), Cu^{2+} (0.97 μm), and OH^- (2.87 μm). Scattering centers such as crystals, oxides, and bubbles have also been found in the HMFG fibers. In their analysis of the data in Fig. 2, the BTRL group separated the total minimum attenuation coefficient (0.65 dB/km at 2.59 μm) into an absorptive loss component equal to 0.3 dB/km and a scattering loss component equal to 0.35 dB/km. The losses for the fluoroaluminate glass fibers are also shown for comparison in Fig. 2.¹⁵ Clearly, the losses are not as low as for the BTRL-ZBLAN fiber, but the AlF_3 -based fluoride fibers do have the advantage of higher glass transition temperatures and therefore are better candidates for laser power delivery.

The reliability of HMFG fibers depends on protecting the fiber from attack by moisture and on pretreatment of the preform to reduce surface crystallization. In general, the HMFGs are much less durable than oxide glasses. The leach rates for ZBLAN glass range between 10^{-3} and 10^{-2} g/cm²/day. This is about five orders of magnitude higher than the leach rate for Pyrex glass. The fluoroaluminate glasses are more durable, with leach rates that are more than three times lower than those for the fluorozirconate glasses. The strength of HMFG fibers is less than that of silica fibers. From Table 2 we see that Young's modulus E for fluoride glass is 51 GPa compared to 73 GPa for silica glass. Taking the theoretical strength to be about one-fifth that of Young's modulus gives a theoretical value of strength of 11 GPa for fluoride glass. The largest bending strength measured has been about 1.4 GPa, well below the theoretical value. To estimate the bending radius R , we may use the approximate expression $R = 1.198r(E/\sigma_{\text{max}})$, where σ_{max} is the maximum fracture stress and r is the fiber radius.¹⁸

Germanate Fibers

Heavy metal oxide glass fibers based on GeO_2 have recently shown great promise as an alternative to HMFG fibers for 3- μm laser power delivery.¹⁹ Today, GeO_2 -based glass fibers are composed of GeO_2 (30–76 percent)–RO (15–43 percent)–XO (3–20 percent), where R represents an alkaline earth metal and X represents an element of Group IIIA.²⁰ In addition, small amounts of heavy metal fluorides may be added to the oxide mixture.²¹ The oxide-only germanate glasses have glass transition temperatures as high as 680°C, excellent durability, and a relatively high refractive index of 1.84. In Fig. 3, loss data is given for a typical germanate glass fiber. While the losses are not as low as they are for the fluoride glasses shown in Fig. 2, these fibers have an exceptionally high damage threshold at 3 μm . Specifically, over 20 W (2 J at 10 Hz) of Er:YAG laser power has been launched into these fibers.

Chalcogenide Fibers

Chalcogenide glass fibers were drawn into essentially the first IR fiber in the mid-1960s.¹ Chalcogenide fibers fall into three categories: sulfide, selenide, and telluride.²² One or more chalcogen elements are mixed with one or more elements such as As, Ge, P, Sb, Ga, Al, Si, and so on to form a glass having two or more components. From the data in Table 2 we see that the glasses have low softening temperatures more comparable to those of fluoride glass than to those of oxide glasses. Chalcogenide glasses are very stable, durable, and insensitive to moisture. A distinctive difference between these glasses and the other IR fiber glasses is that they do not transmit well in the visible region and their refractive indices are quite high. Additionally, most of the chalcogenide glasses, except for As_2S_3 , have a rather large value of dn/dT .⁹ This fact limits the laser power handling capability of the fibers. In general, chalcogenide glass fibers have proven to be an excellent candidate for evanescent wave fiber sensors and for IR fiber image bundles.²³

Chalcogenide glass is made by combining highly purified (>6 nines purity) raw elements in an ampoule that is heated in a rocking furnace for about 10 hours. After melting and mixing, the glass is quenched and a glass preform is fabricated using rod-in-tube or rotational casting methods. Preform

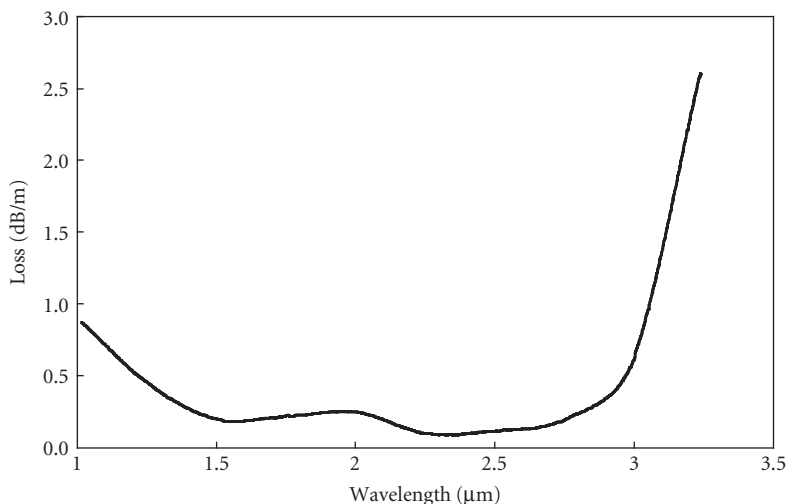


FIGURE 3 Germanate glass fiber manufactured by Infrared Fiber Systems, Silver Spring, Maryland.

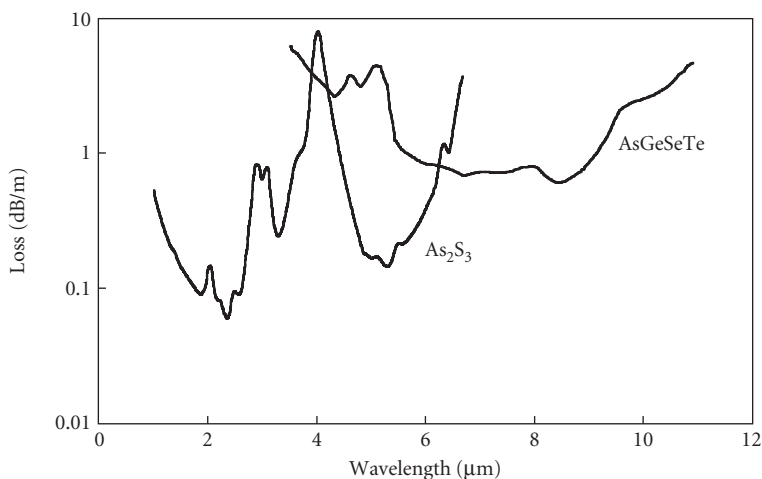


FIGURE 4 Two common chalcogenide glass fibers: As_2S_3 and an AsGeSeTe fiber.⁹ Note the many impurity bands pervasive in these fiber systems.

fiber draws involve drawing a core-clad preform or a core-only preform. For the core-only preform draw, either a soft chalcogenide cladding can be extruded over the fiber as it is drawn or the preform can be Teflon clad. Crucible drawing is also possible.

The losses for the most important chalcogenide fibers are given in Fig. 4. Arsenic trisulfide (As_2S_3) fiber, one of the simplest and oldest chalcogenide fibers, has a transmission range from 0.7 to about 6 μm .²² This fiber is red in color and therefore transmits furthest into the visible region but cuts off in the long-wavelength end well before the heavier chalcogenide fibers.⁹ Longer wavelengths are transmitted through the addition of heavier elements like Te, Ge, and Se, as shown in Fig. 4. A key feature of essentially all chalcogenide glasses is the strong extrinsic absorption resulting from the bonding of contaminants such as hydrogen, H_2O , and OH^- to the elemental cations. In particular, absorption peaks between 4.0 and 4.6 μm are due to S-H or Se-H bonds, and those at 2.78 and 6.3 μm are due to OH^- (2.78 μm) and/or molecular water. The hydride impurities are often especially strong and can be deleterious when these fibers are used in chemical sensing applications where the desired chemical signature falls in the region of extrinsic absorption. Another important feature of most of the chalcogenide fibers is that their losses are in general much higher than those of the fluoride glasses. In fact, at the important CO_2 laser wavelength of 10.6 μm , the lowest loss is still above 1 dB/m for the Se-based fibers.²² More recently single-mode chalcogenide fibers have been drawn from preforms made both using rotational casting and rod-in-tube methods.²⁴ It is also possible to form a chalcohalide glass which is composed of both chalcogen elements and a halide usually I^- . One example is the TeSeAsI glass which has been drawn into fiber by Lucas' group in Rennes, France.²⁵

12.3 CRYSTALLINE FIBERS

Crystalline IR fibers are an attractive alternative to glass IR fibers because most nonoxide crystalline materials can transmit longer-wavelength radiation than IR glasses and, in the case of sapphire, exhibit some superior physical properties as well.² The disadvantage is that crystalline fibers are difficult to fabricate. There are two types of crystalline fiber: single-crystal (SC)⁸ and polycrystalline (PC).^{26,27} Historically, the first crystalline fiber made was hot-extruded KRS-5 fiber fabricated at Hughes Research Labs in 1975,²⁸ KRS-5 or TlBrI was chosen because it is very ductile and because

it can transmit beyond the 20- μm range required for the intended military surveillance satellite application. In fact, crystalline fibers such as KRS-5 and other halide crystals were initially thought to hold great potential as next-generation ultra-low-loss fibers because their intrinsic loss was predicted to be as low as 10^{-3} dB/m.²⁸ Unfortunately, this loss was not only never achieved but not even approached experimentally.

PC Fibers

There are many halide crystals that have excellent IR transmission, but only a few have been fabricated into fiber optics. The technique used to make PC fibers is hot extrusion. As a result, only the silver and thallium halides have the requisite physical properties (such as ductility, low melting point, and independent slip systems) to be successfully extruded into fiber. In the hot extrusion process, a single-crystal billet or preform is placed in a heated chamber and the fiber is extruded to net shape through a diamond or tungsten carbide die at a temperature equal to about half the melting point. The final PC fibers are usually from 500 to 900 μm in diameter with no buffer jacket. The polycrystalline structure of the fiber consists of grains on the order of 10 μm or larger in size. The billet may be clad using the rod-in-tube method. In this method, a mixed silver halide such as AgBrCl is used as the core and then a lower-index tube is formed using a Cl⁻-rich AgBrCl crystal. The extrusion of a core-clad fiber is not as easy to achieve as it is in glass drawing, but Artjushenko et al.¹⁰ at the General Physics Institute (GPI) in Moscow have achieved clad Ag halide fibers with losses nearly as low as those for the core-only Ag halide fiber. Single-mode PC fibers have been extruded from a rod-in-tube preform by Katzir and his group at Tel Aviv University. These fibers have a core diameter of 50 to 60 μm and they are single-mode at 10.6 μm .²⁹ Today, the PC Ag halide fibers represent the best PC fibers. KRS-5 is no longer a viable candidate due largely to the toxicity of Tl and the greater flexibility of the Ag halide fibers.

The losses for the Ag halide fibers are shown in Fig. 5. Both the core-only and core-clad fibers are shown, and, as with the other IR fibers, we again see that there are several extrinsic absorption bands. Water is often present at 3 and 6.3 μm and there is sometimes an SO_4^- absorption near 9.6 μm . Furthermore, we note the decreasing attenuation as the wavelength increases. This is a result of λ^{-2}

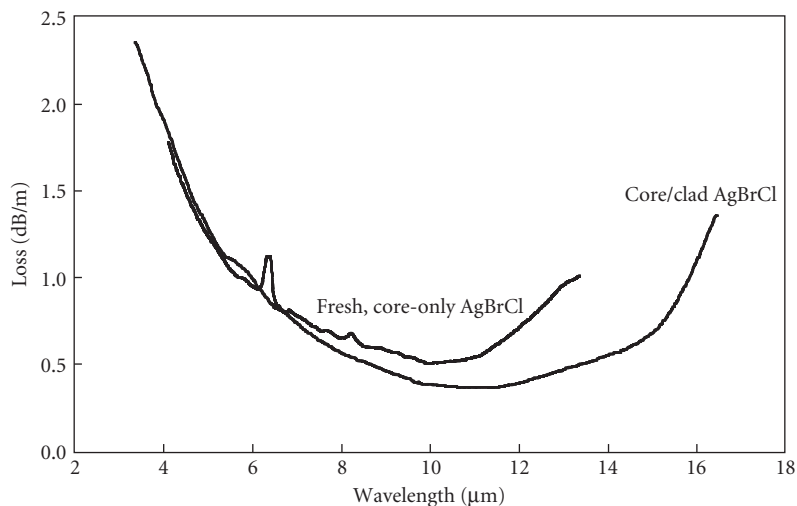


FIGURE 5 Losses in a typical PC silver halide fiber²⁶ compared to those in recently developed core-clad silver halide fiber.¹⁰

scattering from strain-induced defects in the extruded fiber. An important feature of the data is that the loss at 10.6 μm can be as low as 0.2 dB/m for the core-only fiber and these fibers will transmit to almost 20 μm . These fibers have been used to transmit about 100 W of CO_2 laser power, but the safe limit seems to be 20 to 25 W.³⁰ This is due to the low melting point of the fibers.

There are several difficulties in handling and working with PC fibers. One is an unfortunate aging effect in which the fiber transmission is observed to decrease over time.³¹ Normally the aging loss, which increases uniformly over the entire IR region, is a result of strain relaxation and possible grain growth as the fiber is stored. Another problem is that Ag halides are photosensitive; exposure to visible or UV radiation creates colloidal Ag, which in turn leads to increased losses in the IR. Finally, AgBrCl is corrosive to many metals. Therefore, the fibers should be packaged in dark jackets and connectorized with materials such as Ti, Au, or ceramics.

The mechanical properties of these ductile fibers are quite different from those of glass fibers. The fibers are weak, with ultimate tensile strengths of about 80 MPa for a 50–50 mixture of AgBrCl. However, the main difference between the PC and glass fibers is that the PC fibers plastically deform well before fracture. This plastic deformation leads to increased loss as a result of increased scattering from separated grain boundaries. Therefore, in use, the fibers should not be bent beyond their yield point; too much bending can lead to permanent damage and a region of high loss in the fiber.

SC Fibers

Meter-long lengths of SC fibers have been made from only a small number of the over 80 IR transmissive crystalline materials. Initially some SC fibers were grown by zone-refining methods from the same metal halides used to extrude PC fibers. The idea was that removal of the grain boundaries in the PC fibers would improve the optical properties of the fiber. This did not occur, so most of the crystalline materials chosen today for SC fiber fabrication have been oxides. Compared to halides, oxide materials like Al_2O_3 (sapphire) have the advantage of high melting points, chemical inertness, and the ability to be conveniently melted and grown in air. Currently, sapphire is the most popular SC fiber.^{8,32,33}

Sapphire is an insoluble, uniaxial crystal (trigonal structure) with a melting point of over 2000°C. It is an extremely hard and robust material with a usable fiber transmission from about 0.5 to 3.2 μm . Other important physical properties shown in Table 2 include a refractive index equal to 1.75 at 3 μm , a thermal expansion about 10 times higher than that of silica, and a Young's modulus approximately six times greater than that of silica. These properties make sapphire an almost ideal IR fiber candidate for applications less than about 3.2 μm . In particular, this fiber has been used to deliver over 10 W of average power from an Er:YAG laser operating at 2.94 μm .³⁴

Sapphire fibers are fabricated using either the edge-defined, film-fed growth (EFG) or the laser-heated pedestal growth (LHPG) techniques.³⁵ In either method, some or all of the starting sapphire material is melted and an SC fiber is pulled from the melt. In the EFG method, a capillary tube is used to conduct the molten sapphire to a seed fiber, which is drawn slowly into a long fiber. Multiple capillary tubes, which also serve to define the shape and diameter of the fiber, may be placed in one crucible of molten sapphire so that many fibers can be drawn at one time. The LHPG process is a crucibleless technique in which a small molten zone at the tip of an SC sapphire source rod (< 2 mm diameter) is created using a CO_2 laser. A seed fiber slowly pulls the SC fiber as the source rod continuously moves into the molten zone to replenish the molten material. Both SC fiber growth methods are very slow (several millimeters per minute) compared to glass fiber drawing. The EFG method, however, has an advantage over LHPG methods because more than one fiber can be continuously pulled at a time. LHPG methods, however, have produced the cleanest and lowest-loss fibers owing to the fact that no crucible is used that can contaminate the fiber. The sapphire fibers grown by these techniques are unclad, pure Al_2O_3 with the C axis usually aligned along the fiber axis. Fiber diameters range from 100 to 300 μm and lengths are generally less than 2 m. Postcladding techniques mostly involve a Teflon coating using heat-shrink tubing.

The optical properties of the as-grown sapphire fibers are normally inferior to those of the bulk starting material. This is particularly evident in the visible region and is a result of color-center-type

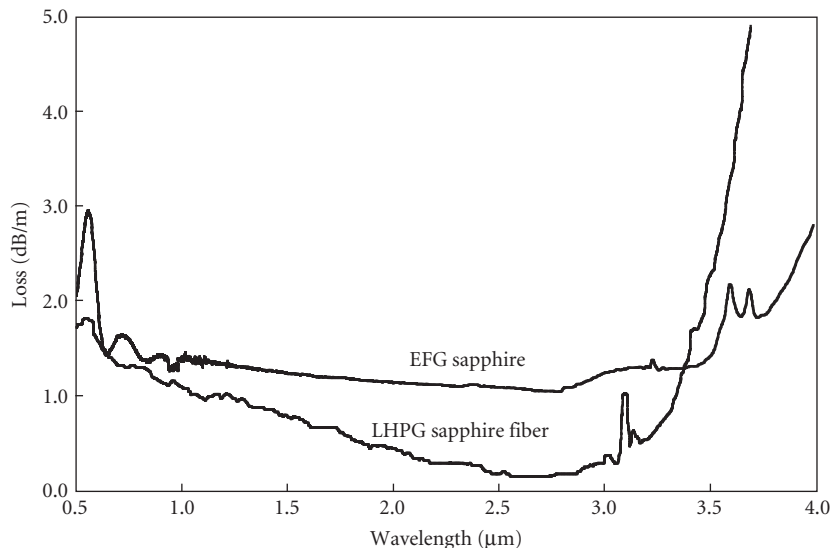


FIGURE 6 SC sapphire fibers grown by the EFG³⁵ (Saphikon, Inc., Milford, New Hampshire) and LHPG³⁴ methods.

defect formation during the fiber drawing. These defects and the resulting absorption can be greatly reduced if the fibers are postannealed in air or oxygen at about 1000°C. In Fig. 6, the losses for LHPG fiber grown at Rutgers University⁸ and EFG fiber grown by Photran, Inc. (Milford, New Hampshire) are shown. Both fibers have been annealed at 1000°C to reduce short-wavelength losses. We see that the LHPG fiber has the lowest overall loss. In particular, LHPG fiber loss at the important Er:YAG laser wavelength of 2.94 μm is less than 0.3 dB/m, compared to the intrinsic value of 0.15 dB/m. There are also several impurity absorptions beyond 3 μm that are believed to be due to transition metals like Ti or Fe. Sapphire fibers have been used at temperatures of up to 1400°C without any change in their transmission.

12.4 HOLLOW WAVEGUIDES

The first optical-frequency hollow waveguides were similar in design to microwave guides. Garmire et al.³⁶ made a simple rectangular waveguide using aluminum strips spaced 0.5 mm apart by bronze shim stock. Even when the aluminum was not well polished, these guides worked surprisingly well. Losses at 10.6 μm were well below 1 dB/m, and Garmire early demonstrated the high power handling capability of an air-core guide by delivering over 1 kW of CO₂ laser power through this simple structure. These rectangular waveguides, however, never gained much popularity, primarily because their overall dimensions (about 0.5 $\times$ 10 mm) were quite large in comparison to circular-cross-section guides and also because the rectangular guides cannot be bent uniformly in any direction. As a result, hollow circular waveguides with diameters of 1 mm or less fabricated using metal, glass, or plastic tubing are the most common guides today. In general, hollow waveguides are an attractive alternative to conventional solid-core IR fibers for laser power delivery because of the inherent advantage of their air core. Hollow waveguides not only enjoy the advantage of high laser power thresholds but also low insertion loss, no end reflection, ruggedness, and small beam divergence. A disadvantage, however, is a loss on bending, which varies as $1/R$ where R is the bending radius.

In addition, the losses for these guides vary as $1/a^3$ where a is the radius of the bore; therefore the loss can be arbitrarily small for a sufficiently large core. The bore size and bending radius dependence of all hollow waveguides are characteristics of these guides not shared by solid-core fibers. Initially these waveguides were developed for medical and industrial applications involving the delivery of CO_2 laser radiation, but more recently they have been used to transmit incoherent light for broadband spectroscopic and radiometric applications.^{37–39} Today they are one of the best alternatives for power delivery in IR laser surgery and industrial laser delivery systems, with losses as low as 0.1 dB/m and transmitted CW laser powers as high as 2.7 kW.⁴⁰

Hollow core waveguides may be grouped into two categories: (1) those whose inner core materials have refractive indices greater than 1 (leaky guides) and (2) those whose inner wall materials have refractive indexes less than 1 [attenuated total reflectance (ATR) guides]. Leaky or $n > 1$ guides have metallic and dielectric films deposited on the inside of metallic,⁴¹ plastic,⁴² or glass¹¹ tubing. ATR guides are made from dielectric materials with refractive indices of less than 1 in the wavelength region of interest.⁴³ Therefore, $n < 1$ guides are fiber-like in that the core index ($n \approx 1$) is greater than the clad index. Hollow sapphire fibers operating at 10.6 μm ($n = 0.67$) are an example of this class of hollow guide.⁴⁴

Hollow Metal and Plastic Waveguides

The earliest circular-cross-section hollow guides were formed using metallic and plastic tubing as the structural members. Miyagi and colleagues in Japan used sputtering methods to deposit Ge,⁴⁵ ZnSe, and ZnS⁴¹ coatings on aluminum mandrels. Then a final layer of Ni was electroplated over these coatings before the aluminum mandrel was removed by chemical leaching. The final structure was then a flexible Ni tube with optically thick dielectric layers on the inner wall to enhance the reflectivity in the infrared. Croitoru and colleagues⁴⁶ at Tel Aviv University applied Ag followed by AgI coatings on the inside of polyethylene and Teflon tubing to make a very flexible waveguide. Similar Ag and Ag halide coatings were deposited inside Ag tubes by Morrow and colleagues.⁴⁷

Hollow Glass Waveguides

The most popular structure today is the hollow glass waveguide (HGW) developed initially at Rutgers University.⁴⁸ The advantage of glass tubing is that it is much smoother than either metal and plastic tubing and, therefore, the scattering losses are less. HGWs are fabricated using wet chemistry methods to first deposit a layer of Ag on the inside of silica glass tubing and then to form a dielectric layer of AgI over the metallic film by converting some of the Ag to AgI. The silica tubing used has a polymer coating of UV acrylate or polyimide on the outside surface to preserve the mechanical strength. The thickness of the AgI is optimized to give high reflectivity at a particular laser wavelength or range of wavelengths. Using these techniques, HGWs have been fabricated with lengths as long as 13 m and bore sizes ranging from 250 to 1300 μm .

The spectral loss for an HGW with a 530- μm bore is given in Fig. 2. This HGW was designed for an optimal response at 10 μm . The peaks at about 3 and 5 μm are not absorption peaks but rather interference bands due to thin-film optical effects. For broadband applications and shorter-wavelength applications, a thinner AgI coating would be used to shift the interference peaks to shorter wavelengths. In fact, for the thinnest AgI films these HGWs not only transmit IR radiation but also shorter wavelengths down to nearly 0.5 μm .⁴⁹ For such HGWs the optical response will be nearly flat without interference bands in the far IR fiber region of the spectrum. The data in Fig. 7 shows the straight loss measured using a CO_2 and Er:YAG laser for different bore sizes. An important feature of this data is the $1/a^3$ dependence of loss on bore size predicted by the theory of Marcatili and Schmeltzer.⁵⁰ In general, the losses are less than 0.5 dB/m at 10 μm for bore sizes larger than $\sim 400 \mu\text{m}$. Furthermore, the data at 10.6 μm agrees well with the calculated values, but at 3 μm the measured losses are somewhat above those predicted by Marcatili and Schmeltzer. This is a result of increased scattering at the shorter wavelengths from the metallic and dielectric films. The bending loss depends

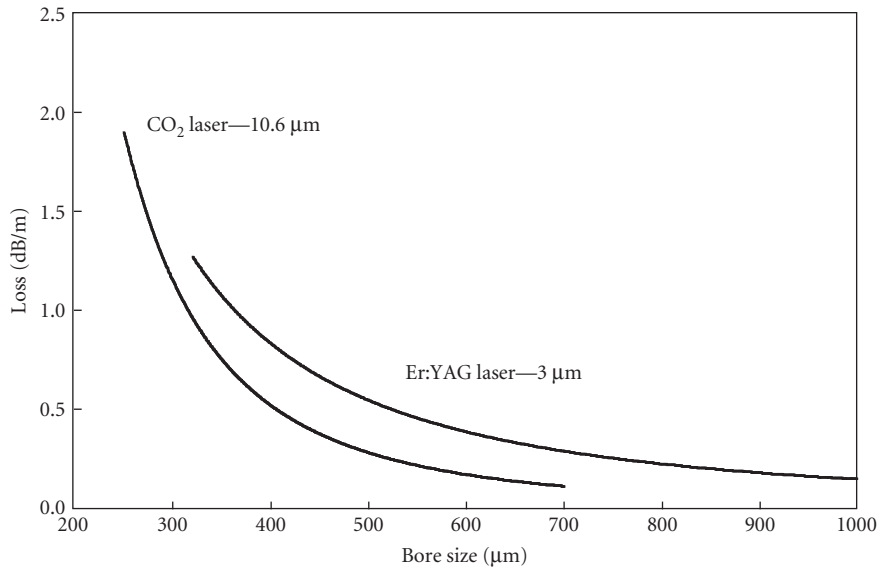


FIGURE 7 Straight losses measured in hollow glass waveguides with Ag/AgI films. The guide labeled “CO₂ laser” was designed for optimal transmission at 10.6 μm while that labeled “Er:YAG laser” was designed for optimal transmission at 3 μm. Note that the loss varies approximately as $1/a^3$.

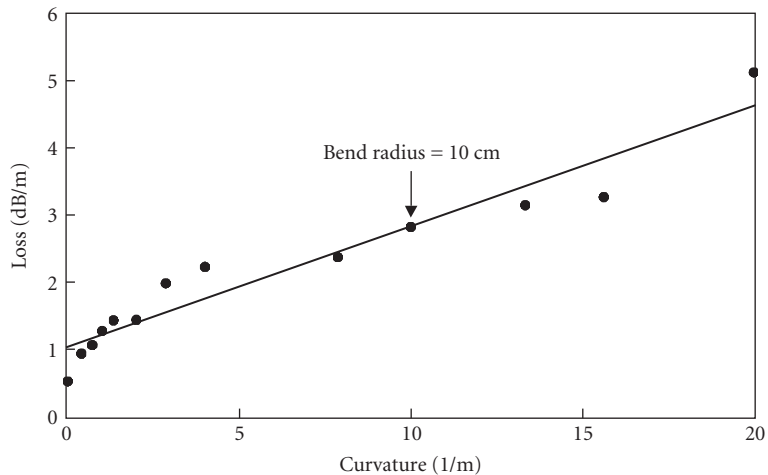


FIGURE 8 Additional loss on bending an HGW with a 530-μm bore, measured at 10.6 μm. The loss is seen to increase as the curvature increases.

on many factors such as the quality of the films, the bore size, and the uniformity of the silica tubing. A typical bending loss curve for an HGW with a 530-μm bore measured with a CO₂ laser is given in Fig. 8. The losses are seen to increase linearly with increasing curvature as predicted. It is important to note that while there is an additional loss on bending for any hollow guide, it does not necessarily mean that this restricts the use of hollow guides in power delivery or sensor applications. Normally

most fiber delivery systems have rather large bend radii and therefore a minimal amount of the guide is under tight bending conditions and the bending loss is low. From the data in Fig. 8 one can calculate the bending loss contribution for an HGW link by assuming some modest bends over a small section of guide length. An additional important feature of hollow waveguides is that they are nearly single mode. This is a result of the strong dependence of loss on the fiber mode parameter. That is, the loss of high-order modes increases as the square of the mode parameter, so even though the guides are very multimode, in practice only the lowest-order modes propagate. This is particularly true for the small-bore ($<300\text{ }\mu\text{m}$) guides, in which virtually only the lowest-order HE_{11} mode is propagated.

HGWs have been used quite successfully in IR laser power delivery and, more recently, in some sensor applications. Modest CO_2 and Er:YAG laser powers below about 80 W can be delivered without difficulty. At higher powers, water-cooling jackets have been placed around the guides to prevent laser damage. The highest CO_2 laser power delivered through a water-cooled hollow metallic waveguide with a bore of $1800\text{ }\mu\text{m}$ was 2700 W, and the highest power through a water-cooled HGW with a $700\text{-}\mu\text{m}$ bore was 1040 W.⁵¹ Sensor applications include gas and temperature measurements. A coiled HGW filled with gas can be used in place of a more complex and costly White cell to provide an effective means for gas analysis. Unlike evanescent wave spectroscopy, in which light is coupled out of a solid-core-only fiber into media in contact with the core, all of the light is passing through the gas in the hollow guide cell, making this a sensitive, quick-response fiber sensor. Temperature measurements may be aided by using an HGW to transmit blackbody radiation from a remote site to an IR detector. Such an arrangement has been used to measure jet engine temperatures.

12.5 SUMMARY AND CONCLUSIONS

During the past 30 years of the development of IR fibers, there has been a great deal of fundamental research designed to produce a fiber with optical and mechanical properties close to those of silica. We can see that today we are still far from that Holy Grail, but some viable IR fibers have emerged that, as a class, can be used to address some of the needs for a fiber that can transmit greater than $2\text{ }\mu\text{m}$. Yet we are still limited with the current IR fiber technology by high loss and low strength. Nevertheless, more applications are being found for IR fibers as users become aware of their limitations and, more importantly, how to design around their properties.

There are two near-term applications of IR fibers: laser power delivery and sensors. An important future application for these fibers, however, may be more in active fiber systems like the Er- and Pr-doped fluoride fibers and emerging doped chalcogenide fibers. In regards to power delivery fibers, the best choice seems to be hollow waveguides for CO_2 lasers and SC sapphire, germanate glass, or HGWs for Er:YAG laser delivery. Chemical, temperature, and imaging bundles make use mostly of solid-core fibers. Evanescent wave spectroscopy (EWS) using chalcogenide and fluoride fibers is quite successful. A distinct advantage of an IR-fiber EWS sensor is that the signature of the analyte is often very strong in the infrared or fingerprint region of the spectrum. Temperature sensing generally involves the transmission of blackbody radiation. IR fibers can be very advantageous at low temperatures, especially near room temperature, where the peak in the blackbody radiation is near $10\text{ }\mu\text{m}$. Finally, there is an emerging interest in IR imaging using coherent bundles of IR fibers. Several thousand chalcogenide fibers have been bundled by Amorphous Materials (Garland, Texas) to make an image bundle for the 3- to $10\text{-}\mu\text{m}$ region.

12.6 REFERENCES

1. N. S. Kapany and R. J. Simms, "Recent Developments of Infrared Fiber Optics," *Infrared Phys.* **5**:69 (1965).
2. J. A. Harrington, *Infrared Fiber Optics and Their Applications*, SPIE Press, Bellingham, WA, 2004.
3. T. Katsuyama and H. Matsumura, *Infrared Optical Fibers*, Adam Hilger, Bristol, UK, 1989.
4. I. Aggarwal and G. Lu, *Fluoride Glass Optical Fiber*, Academic Press, San Diego, 1991.

5. P. France, M. G. Drexhage, J. M. Parker, M. W. Moore, S. F. Carter, and J. V. Wright, *Fluoride Glass Optical Fibers*, CRC Press, Boca Raton, Florida, 1990.
6. J. Sanghera and I. Aggarwal, *Infrared Fiber Optics*, CRC Press, Boca Raton, Florida, 1998.
7. S. F. Carter, M. W. Moore, D. Szebesta, D. Ransom, and P. W. France, "Low Loss Fluoride Fibre by Reduced Pressure Casting," *Electron. Lett.* **26**:2115–2117 (1990).
8. R. Nubling and J. A. Harrington, "Optical Properties of Single-Crystal Sapphire Fibers," *Appl. Opt.* **36**:5934–5940 (1997).
9. J. Nishii, S. Morimoto, I. Inagawa, R. Iizuka, T. Yamashita, and T. Yamagishi, "Recent Advances and Trends in Chalcogenide Glass Fiber Technology: A Review," *J. Non-Cryst. Sol.* **140**:199–208 (1992).
10. V. Artjushenko, V. Ionov, K. I. Kalaidjian, A. P. Kryukov, E. F. Kuzin, A. A. Lerman, A. S. Prokhorov, E. V. Stepanov, K. Bakhshpour, K. B. Moran, and W. Neuberger, "Infrared Fibers: Power Delivery and Medical Applications," *Proc. SPIE* **2396**:25–36 (1995).
11. Y. Matsuura, T. Abel, and J. A. Harrington, "Optical Properties of Small-Bore Hollow Glass Waveguides," *Appl. Opt.* **34**:6842–6847 (1995).
12. P. Kaiser, A. C. Hart Jr., and L. L. Blyler, "Low Loss FEP-Clad Silica Fibers," *Appl. Opt.* **14**:156 (1975).
13. M. Poulain, M. Chanthanasinh, and J. Lucas, "New Fluoride Glasses," *Mat. Res. Bull.* **12**:151–156 (1977).
14. D. Tran, G. H. Sigel, and B. Bendow, "Heavy Metal Fluoride Glasses and Fibers: A Review," *J. Light-Wave Technol.* **LT-2**:566–586 (1984).
15. K. Itoh, K. Miura, M. Masuda, M. Iwakura, and T. Yamagishi, "Low-Loss Fluorozirco-Aluminate Glass Fiber," in *Proceedings of 7th International Symposium on Halide Glass*, Center for Advanced Materials Technology, Monash University, Lorne, Australia, 1991, pp. 2.7–2.12.
16. P. W. France, S. F. Carter, M. W. Moore, and C. R. Day, "Progress in Fluoride Fibres for Optical Communications," *Brit. Telecom Tech. J.* **5**:28–44 (1987).
17. F. Sequin, M. Saad, P. Orsini, and D. Baierl, "Fluoride Glass Fiber for Reliable Er:YAG and Er, Cr:YSGG Laser Power Delivery," I. Gannot, ed., *Proc. SPIE* **6852**, (2008).
18. M. J. Matthewson, C. R. Kurkjian, and S. T. Gulati, "Strength Measurement of Optical Fibers by Bending," *J. Am. Cer. Soc.* **69**:815–821 (1986).
19. S. Kobayashi, N. Shibata, S. Shibata, and T. Izawa, "Characteristics of Optical Fibers in Infrared Wavelength Region," *Rev. Electrical Comm. Lab.* **26**:453–467, 1978.
20. D. Tran, Heavy Metal-Oxide Glass Optical Fibers for Use in Laser Medical Surgery, U.S. Patent 5,274,728, December 28, 1993.
21. G. Cao, F. Lin, H. Hu, and F. Gan, "A New Fluorogermanate Glass," *J. Non-Cryst. Solids* **326 & 327**:170–176, (2003).
22. Y. Kanamori, Y. Terunuma, and T. Miyashita, "Preparation of Chalcogenide Optical Fiber," *Rev. Electrical Comm. Lab.* **32**:469–477 (1984).
23. J. Nishii, T. Yamashita, T. Yamagishi, C. Tanaka, and H. Sone, "As₂S₃ Fibre for Infrared Image Bundle," *Int. J. Optoelectron.* **7**:209–216 (1992).
24. C. Boussard-Pledel, V. S. Shiryayev, P. Houizot, T. Jouan, J. L. Adam, and J. Lucas, "Single-Mode Infrared Fibers Based on TeAsSe Glass Systems," *Mater. Sci. Eng. B (Solid-State Materials for Advanced Technology)* **127**:138–143 (2006).
25. C. Blanchetiere, K. LeFoulgoc, H. L. Ma, X. H. Zhang, and J. Lucas, "Tellurium Halide Glass Fibers: Preparation and Application," *J. Non-Cryst. Solids* **184**:200–203 (1995).
26. V. G. Artjushenko, L. N. Butvina, V. V. Vojtsekhovskiy, E. M. Dianov, and J. G. Kolesnikov, "Mechanisms of Optical Losses in Polycrystalline KRS-5 Fibers," *J. Lightwave Technol.* **LT-4**:461–465 (1986).
27. A. Sa'ar, F. Moser, S. Akselrod, and A. Katzir, "Infrared Optical Properties of Polycrystalline Silver Halide Fibers," *Appl. Phys. Lett.* **49**:305–307 (1986).
28. D. A. Pinnow, A. L. Gentile, A. G. Standlee, A. J. Timper, and L. M. Hobrock, "Polycrystalline Fiber Optical Waveguides for Infrared Transmission," *Appl. Phys. Lett.* **33**:28–29 (1978).
29. S. Shalem, A. Tsun, E. Rave, A. Millo, I. Nagli, and A. Katzir, "Silver Halide Single-Mode Fibers for the Middle Infrared," *Appl. Phys. Lett.* **87**:91103-1–91103-3 (2005).
30. K. Takahashi, N. Yoshida, and M. Yokota, "Optical Fibers for Transmitting High-Power CO₂ Laser Beam," *Sumitomo Electric Tech. Rev.* **23**:203–210 (1984).

31. J. A. Wysocki, R. G. Wilson, A. G. Standlee, A. C. Pastor, R. N. Schwartz, A. R. Williams, G.-D. Lei, and L. Kevan, "Aging Effects in Bulk and Fiber TlBr-TlI," *J. Appl. Phys.* **63**:4365–4371 (1988).
32. D. H. Jundt, M. M. Fejer, and R. L. Byer, "Characterization of Single-Crystal Sapphire Fibers for Optical Power Delivery Systems," *Appl. Phys. Lett.* **55**:2170–2172 (1989).
33. R. S. F. Chang, V. Phomsakha, and N. Djeu, "Recent Advances in Sapphire Fibers," *Proc. SPIE* **2396**:48–53 (1995).
34. R. Nubling and J. A. Harrington, "Single-Crystal LHPG Sapphire Fibers for Er:YAG Laser Power Delivery," *Appl. Opt.* **37**:4777–4781 (1998).
35. H. E. LaBelle, "EFG, the Invention and Application to Sapphire Growth," *J. Cryst. Growth* **50**:8–17 (1980).
36. E. Garmire, T. McMahon, and M. Bass, "Flexible Infrared Waveguides for High-Power Transmission," *J. Quant. Elect.* **QE-16**:23–32 (1980).
37. S. J. Saggese, J. A. Harrington, and G. H. Sigel Jr., "Attenuation of Incoherent Infrared Radiation in Hollow Sapphire and Silica Waveguides," *Opt. Lett.* **16**:27–29 (1991).
38. M. Saito, Y. Matsuura, M. Kawamura, and M. Miyagi, "Bending Losses of Incoherent Light in Circular Hollow Waveguides," *J. Opt. Soc. Am. A* **7**:2063–2068 (1990).
39. M. Saito and K. Kikuchi, "Infrared Optical Fiber Sensors," *Opt. Rev.* **4**:527–538 (1997).
40. A. Hongo, K. Morosawa, K. Matsumoto, T. Shiota, and T. Hashimoto, "Transmission of Kilowatt-Class CO₂ Laser Light Through Dielectric-Coated Metallic Hollow Waveguides for Material Processing," *Appl. Opt.* **31**:5114–5120 (1992).
41. Y. Matsuura, M. Miyagi, and A. Hongo, "Fabrication of Low-Loss Zinc-Selenide Coated Silver Hollow Waveguides for CO₂ Laser Light," *J. Appl. Phys.* **68**:5463–5466 (1990).
42. M. Alaluf, J. Dror, R. Dahan, and N. Croitoru, "Plastic Hollow Fibers as a Selective Infrared Radiation Transmitting Medium," *J. Appl. Phys.* **72**:3878–3883 (1992).
43. C. C. Gregory and J. A. Harrington, "Attenuation, Modal, Polarization Properties of $n < 1$, Hollow Dielectric Waveguides," *Appl. Opt.* **32**:5302–5309 (1993).
44. J. A. Harrington and C. C. Gregory, "Hollow Sapphire Fibers for the Delivery of CO₂ Laser Energy," *Opt. Lett.* **15**:541–543 (1990).
45. M. Miyagi, Y. Shimada, A. Hongo, K. Sakamoto, and S. Nishida, "Fabrication and Transmission Properties of Electrically Deposited Germanium-Coated Waveguides for Infrared Radiation," *J. Appl. Phys.* **60**:454–456 (1986).
46. O. Morhaim, D. Mendlovic, I. Gannot, J. Dror, and N. Croitoru, "Ray Model for Transmission of Infrared Radiation through Multibent Cylindrical Waveguides," *Opt. Eng.* **30**:1886–1891 (1991).
47. P. Bhardwaj, O. J. Gregory, C. Morrow, G. Gu, and K. Burbank, "Performance of a Dielectric-Coated Monolithic Hollow Metallic Waveguide," *Mat. Lett.* **16**:150–156 (1993).
48. T. Abel, J. Hirsch, and J. A. Harrington, "Hollow Glass Waveguides for Broadband Infrared Transmission," *Opt. Lett.* **19**:1034–1036 (1994).
49. K. E. Sui, W. Shi, X. L. Tang, X. S. Shu, K. Iwai, and M. Miyagi, "Optical Properties of AgI/Ag Infrared Hollow Fiber in the Visible Wavelength Region," *Opt. Lett.* **33**:318–320 (2008).
50. E. A. J. Marcatili and R. A. Schmeltzer, "Hollow Metallic and Dielectric Waveguides for Long Distance Optical Transmission and Lasers," *Bell Syst. Tech. J.* **43**:1783–1809 (1964).
51. R. K. Nubling and J. A. Harrington, "Hollow-Waveguide Delivery Systems for High-Power, Industrial CO₂ Lasers," *Appl. Opt.* **34**:372–380 (1996).

This page intentionally left blank

SOURCES, MODULATORS, AND DETECTORS FOR FIBER OPTIC COMMUNICATION SYSTEMS

Elsa Garmire

*Dartmouth College
Hanover, New Hampshire*

13.1 INTRODUCTION

Optical communication systems utilize fiber optics to transmit the light that carries the signals. Such systems require optoelectronic devices as sources and detectors of such light, and they need modulators to impress the telecommunication signals onto the light. This chapter outlines the basics of these devices. Characteristics of devices designed for both high-performance, high-speed telecommunication systems (*telecom*) and low-cost, more modest performance local area networks (LAN) and data communication systems (*datacom*) are presented. Sources for telecom are edge-emitting laser diodes (LDs), including double heterostructure (DH), quantum well (QW), strained layer (SL), distributed feedback (DFB), and distributed Bragg reflector (DBR) laser diodes. Operating characteristics of these edge-emitting LDs include threshold, light-out versus current-in, spatial, and spectral characteristics. The transient response includes relaxation oscillations, turn-on delay, and modulation response. The noise characteristics are described by relative intensity noise (RIN), signal-to-noise ratio (SNR), mode partition noise (in multimode LDs), and phase noise (which determines linewidth). Frequency chirping broadens the linewidth, described in the small- and large-signal regime; external optical feedback may profoundly disturb the stability of the LDs and may lead to coherence collapse.

Semiconductor lasers usually have a laser cavity in the plane of the semiconductor device, with light emitted out through a cleaved edge in an elliptical output pattern. This output is not ideally suited to coupling into fibers, which have circular apertures. Low-cost systems, such as datacom, put a premium on simplicity in optical design. These systems typically use multimode fibers and surface-emitting, light-emitting diodes (LEDs). The LEDs are less temperature dependent than LDs and are more robust, but they typically are slower and less efficient. Those LEDs applicable to fiber optics are described here, along with their operating and transient response characteristics. Edge-emitting LEDs have some niche fiber-optic applications and are briefly described along with the latest advance in LEDs, which is to incorporate a resonant cavity to narrow both the linewidth and increase the output efficiency that can couple into a fiber.

Vertical cavity surface-emitting lasers (VCSELs) have vertical laser cavities that emit light vertically out of the plane of the semiconductor device. Fibers couple more easily to these surface-emitting sources, but VCSEL performance is usually degraded compared to that of the edge-emitting sources. This chapter outlines typical VCSEL designs (material, optical, and electrical); their spatial, spectral, and

polarization characteristics; and their light-out versus current-in characteristics. The most common VCSELs are based on gallium arsenide, operating from 750 to 980 nm, but longer wavelength VCSELs are becoming more practical with GaInNAs quantum wells on GaAs or five-component InAlGaAsP on indium phosphide (InP).

The most common modulators used in fiber-optic systems today are external lithium niobate modulators. These are usually used in Y-branch interferometers, creating intensity modulation as a result of phase modulation by the electro-optic effect. These modulators are introduced in this chapter along with a discussion of high-speed modulation, losses, and polarization dependence, as well as a brief description of optical damage and other modulator geometries. These devices provide chirpfree modulation that can be made very linear for applications such as cable TV.

A recent development in waveguide modulators is the multimode interference (MMI) modulator, based on interference between multiple in-plane spatial modes. This leads to ultra compact modulators, because of much reduced optical path length requirements.

Until recently, lithium niobate was the electro-optic material of choice. Recently polymers have become a viable option for some applications, particularly low-cost. These modulators can be ultra-high-speed and have less insertion loss because of their lower refractive index. Initial challenges of long-term reliability are being met with additional research.

An alternative modulator uses semiconductors, particularly QWs, which allow for more compact devices and monolithic integration. Typically, these are intensity modulators using electroabsorption. By careful design, the chirp in these modulators can be controlled and even used to counteract pulse spreading from chromatic dispersion in fibers. The quantum-confined Stark effect (QCSE) is described, along with the *pin* waveguides used as modulators and techniques for their integration with lasers. Their operating characteristics as intensity modulators, their chirp, and improvements available by using strained QWs are presented.

Some semiconductor modulators use phase change rather than absorption change. The electro-optic effect in III-V semiconductors is discussed, along with the enhanced refractive index change that comes from the QCSE, termed *electrorefraction*. Particularly large refractive index changes occur if available quantum well states are filled by electrons. Phase-change modulators based on this principle can be used in interferometers to yield intensity modulators.

Detectors used in fiber systems are primarily *pin* diodes, although short descriptions of avalanche photodetectors (APDs) and metal-semiconductor-metal (MSM) detectors are provided. The geometry, sensitivity, speed, dark current, and noise characteristics of the most important detectors used in fiber systems are described.

Most of the devices discussed in this chapter are based on semiconductors, and their production relies on the ability to tailor the material to design specifications through *epitaxial growth*. This technology starts with a bulk crystal substrate (usually the binary compounds GaAs or InP) and employs the multilayered growth upon this substrate of a few micrometers of material with a different composition, called a *heterostructure*. Ternary layers substitute a certain fraction x for one of the two binary components.

Thus, $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is a common ternary alloy used in laser diodes. Another common ternary is $\text{In}_x\text{Ga}_{1-x}\text{As}$. Layers are *lattice-matched* when the ternary layers have the same size lattice as the binary; otherwise, the epitaxial layer will have *strain*. Lattice-matched epitaxial layers require the substituting atom to have approximately the same size as the atom it replaces. This is true of Al and Ga, so that $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ternary layers are lattice-matched to GaAs. The lowest-cost lasers are those based on GaAs substrates with $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ternary layers surrounding the active layer. These lasers operate at wavelengths near the bandgap of GaAs, about 850 nm, and are typically used in low-cost data communications (as well as in CD players).

The wavelengths required for laser sources in telecommunications applications are those at which the fiber has the lowest loss and/or dispersion, traditionally 1.55 and 1.3 μm . There is no binary semiconductor with a bandgap at these wavelengths, nor is there a lattice-matched ternary. The $\text{In}_x\text{Ga}_{1-x}\text{As}$ ternary will be strained under compression when it is grown on either GaAs or InP, because indium is a much bigger atom than gallium, and arsenic is much bigger than phosphorous. The way to eliminate this strain is to use a fourth small atom to reduce the size of the lattice back to that of the binary. This forms a *quaternary*. The heterostructure most useful for fiber optics applications is based on InP substrates.

The quaternary $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ is commonly used, with the compositions x and y chosen to simultaneously provide the desired wavelength and lattice match. These quaternary heterostructures are the basis for much of the long-wavelength technology: sources, modulators, and detectors.

Earlier volumes of this *Handbook* discuss the basics of lasers (Chap. 16, "Lasers," Vol. II), LEDs (Chap. 17, "Light-Emitting Diodes," Vol. II), modulators (Pt. 3, "Modulators," Vol. V), and detectors (Pt. 5, "Detectors," Vol. II). The reader is referred there for general information. This chapter is specific to characteristics that are important for fiber communication systems.

13.2 DOUBLE HETEROSTRUCTURE LASER DIODES

Telecommunication sources are usually edge-emitting lasers, grown with an active laser layer that has a bandgap near either 1.55 or 1.3 μm . The active layer has a quaternary composition consisting of $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$, grown lattice-matched to InP. The materials growth and fabrication technology had to be developed specifically for telecommunication applications and is now mature. These LDs are more temperature sensitive than GaAs lasers, and this fact has to be incorporated into their use. For telecom applications they are often attached to a thermoelectric cooler and typically provided with a monitoring photodiode in the laser package in order to provide a signal for temperature and/or current control.

Today's telecom systems use single-mode fibers, which require lasers with a single spatial mode. In order to avoid dispersion over long distances, a single frequency mode is necessary. These requirements constrain the geometry of laser diodes used for telecom applications, as discussed in the next section. Following sections discuss the operating characteristics of these LDs and their transient response and noise characteristics, both as isolated diodes and when subject to small reflections from fiber facets. The modulation characteristics of these diodes are discussed, along with frequency chirping. Advanced laser concepts, such as quantum well lasers, strained layer lasers, and lasers with distributed reflection (DFB and DBR lasers) are also introduced.

A typical geometry of an edge-emitting InGaAsP/InP laser is shown in Fig. 1. The active quaternary laser region is shown cross-hatched. It is from this region that light will be emitted. Light travels back and forth between cleaved mirror facets, confined to the active InGaAsP region by the buried heterostructure, and is emitted out of the cross-hatched region, where it diffracts to the far field. The current is confined to the stripe region by the current-blocking *npn* structure on either side. Traditionally, the active regions have uniform composition and are lattice-matched to the substrate. More advanced laser diodes, often used for telecom applications, have active regions containing one or more quantum wells and may be grown to incorporate internal strain in the active region. Both these characteristics are described in Sec. 13.6.

The design of a double heterostructure laser diode requires optimization of the issues discussed in the following subsections.

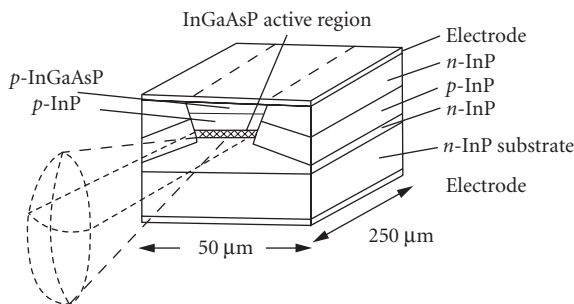


FIGURE 1 Typical geometry for an edge-emitting long-wavelength laser diode, as used in telecommunication systems.

Injection of a Population Inversion into the Active Region

Stimulated emission requires a forward-biased diode, created by sandwiching the active region between p and n layers. Electrons are injected into the active region from the n side and holes are injected from the p side; they become *free carriers*. Efficient electrical injection requires high-quality ohmic contacts attached to the n and p layers; electrical current through the junction then drives the laser.

Confinement of Carriers within the Plane of the Active Layer

Confinement is achieved by growing the n and p layers on either side of the active region with a larger bandgap, as shown in Fig. 2a. In quaternary lasers, wider bandgap material is provided by decreasing x and y relative to their values in the active region. Stimulated emission during electron-hole recombination in the narrow bandgap active layer provides the laser light. The thinner the active layer, the higher its gain. When the active layer thickness is as small as a few 10s of nanometers, the free electron and hole energy levels become quantized in the growth direction, and the active layer becomes a *quantum well*. Quantum wells have higher gain than bulk semiconductor active layers, and thus one or more QWs are often used as the active layers (see Sec. 13.6).

Confinement of Light Near the Active Layer

Stimulated emission gain is proportional to the product of the carrier and photon densities, so that edge-emitting lasers require the highest possible light intensity. This is done by containing the light in an optical waveguide, with a typical near-field light profile, as shown in Fig. 2b. To achieve optical confinement, the layers surrounding the waveguide must have lower refractive indices. It is fortunate that higher-bandgap materials that confine carriers also have smaller refractive index, and so the active layer automatically becomes a waveguide.

Proper optical confinement requires a single waveguide mode. This means that the waveguide layer must be thinner than the cutoff value for higher-order modes. The waveguide thickness d_g must be small enough that

$$d_g k_o \sqrt{n_g^2 - n_c^2} \equiv V < \pi \quad (1)$$

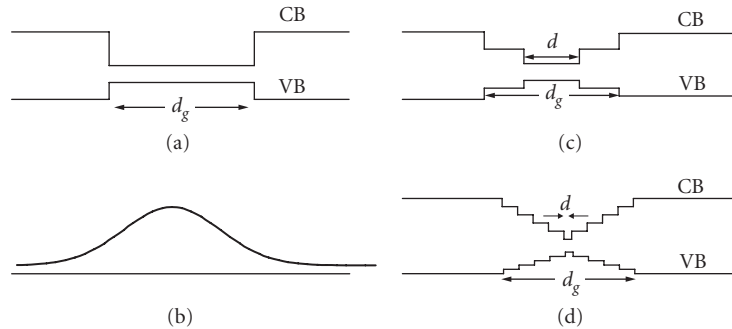


FIGURE 2 Typical diode laser energy band structures and guided mode profile: (a) conduction band (CB) and valence band (VB) of double heterostructure; (b) corresponding near-field spatial profile for light guided in layer of width d_g ; (c) separate confinement heterostructure (SCH); and (d) graded index separate confinement heterostructure (GRINSCH).

where n_g is the refractive index of the waveguide layer (usually the active layer), n_c is the refractive index of the surrounding cladding (usually the p and n layers), and $k_o = 2\pi/\lambda$, where λ is the free-space wavelength of the laser light. Typically, $n_g - n_c \sim 0.2$ and $d_g < 0.56 \mu\text{m}$ for $\lambda = 1.3 \mu\text{m}$. The parameter V is usually introduced to characterize a waveguide.

If the waveguide is too thin, however, the waveguided optical mode spreads out beyond the waveguide layer. The fraction of optical power Γ_g (called the *waveguide confinement factor*) that remains in the waveguide layer of thickness d_g is given approximately by¹

$$\Gamma_g = \frac{V^2}{V^2 + 2} \quad (2)$$

As d_g becomes small, the confinement factor becomes small. When the carriers are confined in very thin layers, such as in quantum wells, the electrical carrier confinement layer cannot serve as an effective optical waveguide because the confinement factor is too small. Then a thicker waveguide region is used to confine the photons, and the carriers are separately confined in a thinner active region, a geometry called a *separate confinement heterostructure* (SCH), as shown in Fig. 2c. In this case the optical confinement factor, defined by the fraction of photons in the active layer of thickness d_a , is $\Gamma = \Gamma_g (d/d_g)$.

The light can be more effectively focused into a thin active layer by grading the refractive index in the separate confinement region, called a *graded index SCH* (GRINSCH) laser, shown in Fig. 2d. This graded refractive index is produced by growing material with varying bandgaps within the waveguide layer. Grading can be achieved by several discrete layers, as shown, or by grading many ultrathin layers with slight compositional differences. In either case, the focusing property of a GRINSCH structure can be approximated by fitting the graded refractive index to a parabolic refractive index profile $n(x)$ such that $n(x)^2 = n_g^2(1 - x^2/x_0^2)$, where x_0 is related to the curvature of the refractive index near $x = 0$: $x_0 = (n_g/n'')^{1/2}$, where $n'' \equiv \partial^2 n / \partial x^2$ near $x = 0$. The mode guided by this profile has a gaussian beam-intensity profile $I(x) = I_o \exp(-x^2/w^2)$, where $w^2 = x_0^2/k_g$ and $k_g = 2\pi n_g/\lambda$.

Limiting Carrier Injection to Stripe Geometry

Lasers are most efficient when the drive current is limited to the width of the optically active laser area. This requires defining a narrow stripe geometry electrode by means of a window etched in an isolating oxide layer or by ion implantation to render either side of the stripe resistive. More complex laser structures, such as those used in telecommunication applications, often define the conductive stripe electrode by using current-blocking *npn* layers grown on either side of the electrode, as shown in Fig. 1. The *npn* layers, consisting of back-to-back diodes, do not conduct current.

Injected carriers do not usually need lateral confinement, except to achieve the highest possible efficiency. Lateral free-carrier confinement may occur as a by-product of lateral optical confinement, which is discussed next.

Lateral Confinement of Light

The simplest laser diode structures do not specifically confine light laterally, except as the result of the stripe geometry carrier injection. These are called *gain-guided* because high gain in the stripe region, due to the presence of free carriers, introduces a complex refractive index that guides the light laterally. Gain-guided LDs tend to be multimode (both lateral spatial modes and longitudinal frequency modes) unless the stripe is very narrow ($10 \mu\text{m}$). In this case, the spatial far-field pattern has “rabbit ears,” a double-lobed far-field pattern that is typically not very useful for coupling into single-mode fibers. Thus, gain-guided LDs are not usually used for telecommunications.

High-quality single-mode LDs for telecom applications typically require a real refractive index difference laterally across the laser. The lowest threshold lasers use *buried heterostructure* (BH) lasers, the geometry shown in Fig. 1. After most of the layers are grown, the sample is taken out of the

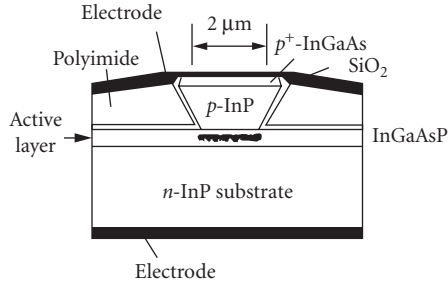


FIGURE 3 Geometry for a ridge waveguide (RWG) laser.

growth chamber and a stripe geometry mesa is etched. Then the sample is returned to the growth chamber, and one or more cladding layers with lower refractive index (higher bandgap) are grown, typically InP, as shown in Fig. 1. When the regrowth is planar, these are called *planar buried heterostructure* (PBH) lasers. The result is a real refractive index guide in the lateral dimension. The width of these *index-guided* laser stripes may be anything from $1\ \mu\text{m}$ to more than $10\ \mu\text{m}$, depending on the refractive index difference between the active stripe and the lateral cladding material. Equation (1), which specifies the condition for single-mode, applies here, with d_g as the width of the lateral index guide and n_c defined by the regrown material. A typical lateral width for low-threshold BH lasers is $3\ \mu\text{m}$.

A laser geometry that is much simpler to fabricate and has a higher reliability in production than that of BH lasers is the *ridge waveguide* (RWG) laser, shown in Fig. 3. Light is confined to the region under the p -InP etched mesa by strip loading, which increases the effective refractive index in the waveguide region under the etched mesa. The fabrication starts with the growth of a separate confinement heterostructure (sometimes with the addition of a thin etch-stop layer just after the top waveguide layer), followed by a stripe mesa etch down to the waveguide layer, finishing with planarization and contacting to the stripe. The etch leaves a ridge of p -cladding material above the waveguide layer, which causes *strip loading*, raising the effective refractive index locally in the stripe region, thereby creating lateral confinement of the light. Although the RWG laser is attractive because it requires only a single epitaxial growth, its threshold current is relatively high.

Retroreflection of Guided Light Along the Stripe

Light is usually reflected back and forth inside the laser cavity by Fresnel reflection from cleaved end facets. Since the waveguide refractive index is $n_g \sim 3.5$, the natural Fresnel reflectivity at an air interface, $R = [(n_g - 1)/(n_g + 1)]^2$, is ~ 0.3 . This rather low reflectivity means that LDs are high gain, requiring enough amplification that 70 percent of the light is regenerated on each pass through the active medium.

Relying on Fresnel reflection means that both facets emit light. The light emitted out the back facet may be recovered by depositing a high-reflectivity multilayer coating on the back facet, as is typically done in most telecom lasers. Sometimes a coating is also provided on the front facet to alter its reflectivity, typically to lower it, which increases the output power (as long as the gain is high enough to overcome the large loss upon reflection). The reflectivities must be such that the laser can obey the laser operating condition, which states that in a single round-trip through a laser of length L , the increase in optical power from gain must balance the reduction from finite reflectivity, so that their product is unity. That is,

$$R_1 R_2 \exp(2g_L L) = 1 \quad (3)$$

where R_1 and R_2 are the reflectivities of the two facets and g_L is the *modal gain per unit length* (as experienced by the waveguided laser mode), with a subscript L to represent that the gain is measured with respect to length. If $R_1 = R_2 = 0.3$, then $g_L L = 1.2$. Typical laser diodes have lengths of 400 μm , so $g_L \sim 30 \text{ cm}^{-1}$.

In-plane retroreflection can also be achieved by using distributed feedback created from a grating fabricated on top of the active layer. This method enables the construction of *distributed feedback* (DFB) lasers and *distributed Bragg reflector* (DBR) lasers, which are discussed in Sec. 13.5.

Mounting so that Light Is Edge-Emitted

Because the light is emitted out of the facet laterally, there must be a clear optical path for the light as it exits the laser. In many cases, the light is mounted with the active layer down, very close to its copper (or diamond) heat sink, in order to maximize cooling.² In this case, the laser chip must be placed at the very edge of the heat-sink block, as shown in Fig. 4a.

In some cases, the laser is mounted with its active region up with its substrate next to the heat sink. The edge alignment is not so critical in this case, but of course the laser light will still be emitted in a direction parallel to the plane of the heat sink. Because the thermal conductivity of the heat sink is much higher than that of the substrate, only the lowest threshold lasers, operating at moderate power levels, are operated with the active region up.

Suitable Packaging in Hermetic Enclosure

Water vapor can degrade bare facets of a semiconductor laser when it is operating; therefore, LDs are usually passivated (i.e., their facets are coated with protective layers) and/or they are placed in sealed packages. The LD may be placed in a standard three-pin semiconductor device package, such as a TO-46 can with an optical window replacing the top of the can, as shown in Fig. 4a. The LD will be situated near the package window because the light diverges rapidly after it is emitted from the laser facet. The package window should be antireflection coated because any light reflected back into the laser can have serious consequences on the stability of the output (see Sec. 13.5).

Many high-end applications require an on-chip power monitor and/or a controllable thermoelectric cooler. In this case a more complex package will be used, typically a 14-pin “butterfly” package, often aligned to a fiber pigtail, such as shown in Fig. 4b. In the less expensive datacom applications, nonhermetic packages may be acceptable with proper capping and passivation of the laser surfaces.

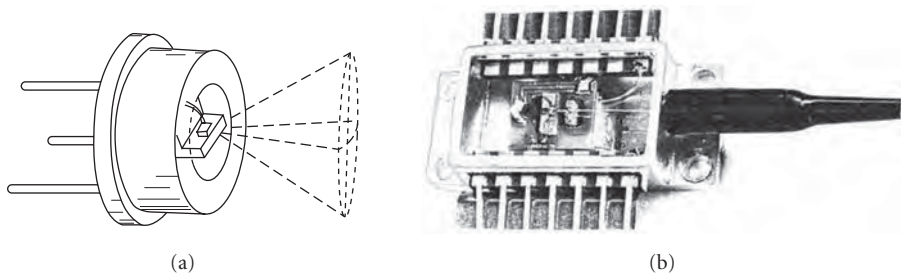


FIGURE 4 Packaging laser diodes: (a) typical hermetically sealed package showing heat sink and emission pattern for a laser diode with its active region placed down on a copper (or diamond) heat sink and (b) typical butterfly package, showing laser in the middle, monitoring photodiode (behind), and fiber alignment chuck in front, all mounted on a thermoelectric cooler. (Photo provided by Spectra-Diode Laboratories.)

Fiber Pigtail Connection

Because light diverges at a rather large angle as it comes out of an edge-emitting laser (as discussed later), it is often desirable to use a laser provided with a *fiber pigtail*, which is a prealigned length of fiber that can be spliced or connected to the telecom fiber in the field. There will be an inevitable reduction in output power (compared to that of a laser with no pigtail) because of finite coupling efficiency into the pigtail, but the output will be immediately useful in a telecom system. The alternative to a fiber pigtail is the use of a microlens, often a graded index (GRIN) lens, discussed in Chap. 18 in this volume.

Long Life

Early lasers showed degradation with running time, but those problems have been solved, and the LDs used in telecom systems should last hundreds of thousands of hours. However, long life requires that care be taken in their use because large reverse-bias static voltages can break down the *pn* diode. Thus, protection from electrostatic shock while handling and from reflected reverse-bias electrical currents during operation should be maintained. In addition, if LDs are driven with too much forward-bias current, the optical output can be so large that the light may damage the facet out of which it is emitted. Since the threshold is strongly temperature dependent, a laser driven at constant current that becomes too cold can emit too much light, with resulting optical damage. Thus, many telecom lasers have monitoring photodiodes to control the laser output and ensure that it stays within acceptable bounds.

13.3 OPERATING CHARACTERISTICS OF LASER DIODES

The principles of semiconductor laser operation are shown in Chap. 19 in Vol. II of this *Handbook*. A forward-biased *pn* junction injects carriers into the active region. As the drive current increases, the carrier density in the active region increases. This reduces the absorption from an initially high value (at thermal equilibrium the absorption coefficient $\alpha \approx 500 \text{ cm}^{-1}$) to zero, at which point the active layer becomes transparent at the prospective laser wavelengths. An active layer is characterized by its carrier density at transparency, N_{tr} . Typically, $N_{tr} \approx 10^{18} \text{ cm}^{-3}$. Above this carrier density, stimulated emission occurs, with a gain proportional to the diode carrier density above transparency. The gain depends on the detailed device design, taking into account the issues enumerated in the preceding section and the materials involved. The gain is sizeable only in direct-band semiconductors (semiconductors based on the III-V or II-VI columns of the periodic table).

Laser Threshold

Threshold is given by the requirement that the round-trip optical gain due to stimulated emission must equal the round-trip optical loss due to the sum of the transmission out the end facets and any residual distributed loss. Gain occurs only for light that is actually in the active region, and not for the fraction of waveguided light that extends outside the active region. Typically, the *local* gain per unit length G_L is defined as that experienced locally by light inside the active region. (The *modal* gain per unit length is $g_L = \Gamma G_L$.) Near transparency, the local gain depends linearly on carrier density N : $G_L = a_L (N - N_{tr})/N_{tr}$, where a_L is the proportionality constant in units of length ($a_L \equiv N \partial G_L / \partial N$ near N_{tr}). Assuming G_L is linear in N , an unpumped region ($N = 0$) has $G_L = -a_L$, which is its loss per unit length. Typically, $a_L \sim 250 \text{ cm}^{-1}$.

The current density (J) is related to the carrier density through $J = eNd/\tau$, where τ is the lifetime of the electron-hole pairs. The transparency current density is 1200 A/cm^2 when $d = 0.15 \text{ }\mu\text{m}$,

$N_{tr} = 10^{18} \text{ cm}^{-3}$, and $\tau = 2 \text{ ns}$. The threshold condition on gain can be found by taking the natural logarithm of Eq. (3) to obtain $g_{L,th} = G_{L,th}$, $\Gamma = \alpha_i + \alpha_m$, where α_m is the mirror reflectivity amortized over length, such that $2\alpha_m L = \ln(1/R_1 R_2)$; and α_i represents any internal losses for the laser mode, also amortized over length. Combining these relations, along with the fact that the current I in a LD with stripe width w and length L is $I = JwL$, gives

$$I_{th} = I_{tr} + \frac{ewN_{tr}}{\tau a_L} \left[\frac{1}{2} \ln \left(\frac{1}{R_1 R_2} \right) + \alpha_i L \right] d \left(1 + \frac{2}{V^2} \right) \quad (4)$$

where the waveguide V parameter is from Eq. (1) with $d_g = d$. Note that the threshold current is independent of device length L when the internal losses are small. The longer the spontaneous lifetime, the lower the threshold current density (although this may lengthen the turn-on time, as discussed in Sec. 13.4). Finally, as expected by the relation between current and current density, a thinner stripe width w will lower the threshold current (consistent with appropriate spatial output, as discussed later). The current density at transparency N_{tr} is a basic property of the gain curve of the active region. It is smaller for QW lasers (Sec. 13.6) than for thicker active regions.

Because V is linearly proportional to d , there is an optimal active layer thickness, a trade-off between increasing the carrier density as much as possible, but not so much as to lose optical confinement. The optimum thickness for $\lambda = 1.3\text{-}\mu\text{m}$ LDs is $\sim 0.15 \mu\text{m}$, and comparable for $\lambda = 1.55\text{-}\mu\text{m}$ LDs (0.15 to $0.18 \mu\text{m}$). Threshold currents for broad-area DH lasers can be under 500 A/cm^2 at $\lambda = 1.3 \mu\text{m}$ and $\sim 1000 \text{ A/cm}^2$ at $\lambda = 1.55 \mu\text{m}$. Confining carriers and light separately can beat this requirement, a trick used in designing QW lasers.

Light Out versus Current In (the L-I Curve)

Below laser threshold only spontaneous emission is observed, which is the regime of the LED, as discussed in Sec. 13.8. In the spontaneous regime, the output varies linearly with input current and is emitted in all directions within the active region. As a result, a negligible amount of light is captured by the single-mode fiber of telecom below threshold.

Above threshold, the electrical power is converted to optical power. In general, the light will come out of both facets, so the amount of light reflected out the front facet depends on the rear facet reflectivity. When a 100 percent reflective mirror is deposited on the back facet, the optical power at photon energy $h\nu$ (wavelength $\lambda = c/\nu$) emitted out the front facet is

$$P_{out} = \frac{h\nu}{e} \frac{\alpha_m}{\alpha_m + \alpha_i} (I - I_{th} - I_L) \eta_i \quad (5)$$

where η_i is the *internal quantum efficiency*, which is the fraction of injected carriers that recombine by radiative recombination (usually close to unity in a well-designed semiconductor laser), and I_L is any leakage current. This equation indicates a linear dependence between light out and current above threshold (for constant quantum efficiency). The power out will drop by a factor of 2 if the back facet has a reflectivity equal to that of the front facet, since half the light will leave out the back.

From Eq. (5) can be calculated the *external slope efficiency* of the LD, given by $\partial P_{out} / \partial I$. This allows the *differential quantum efficiency* η_D to be calculated in terms of the power out (both facets):

$$\eta_D \equiv \frac{e}{h\nu} \frac{\partial P_{out}}{\partial I} = \eta_i \frac{\alpha_m}{\alpha_m + \alpha_i} \quad (6)$$

The quantity η_D is also called the *external quantum efficiency*. This efficiency depends on intrinsic loss, possibly due to scattering from roughness in the edges of the waveguide. In long-wavelength lasers, this loss is primarily due to intervalence band absorption.

The *internal* quantum efficiency η_i depends on the way carriers recombine. The rate of carrier loss is the sum of spontaneous processes, expressed in terms of carrier density divided by a lifetime τ_e , and stimulated emission, expressed in terms of local gain per unit time G_T and local photon density P :

$$R(N) = \frac{N}{\tau_e} + G_T(N)P \quad (7)$$

The spontaneous carrier lifetime is given by

$$\frac{1}{\tau_e} = A_{nr} + BN + CN^2 \quad (8)$$

The term BN is due to spontaneous radiative recombination; its dependence on N results from needing the simultaneous presence of both an electron and hole (which have the same charge densities because of charge neutrality in undoped active regions). The nonradiative recombination terms that decrease the quantum efficiency below unity are a constant term A_{nr} (accounting for all background nonradiative recombination) and an Auger recombination term (with coefficient C) that depends on the square of the carrier density (Auger processes involve several carriers simultaneously). This term is particularly important in long-wavelength lasers where the Auger coefficient C is large.

Stimulated emission is accounted for by gain in the time domain G_T , which depends on N (approximately linearly near threshold). The group velocity v_g converts gain per unit length G_L into a rate G_T (gain per unit time), $G_T \equiv v_g G_L$. We can define a_T in the time domain by $a_T = v_g a_L$ so that $G_T = a_T (N - N_{tr})/N_{tr}$.

The *internal* quantum efficiency in a laser is the fraction of the recombination processes that emit light:

$$\eta_i = \frac{BN^2 + G_T(N)P}{A_{nr}N + BN^2 + CN^3 + G_T(N)P} \quad (9)$$

At high powers in well-designed lasers, the internal quantum efficiency approaches 1.

Figure 5 shows a typical experimental result³ for the light out of a LD as a function of applied current (the so-called *L-I* curve). It can be seen that the linear relation between light out and current

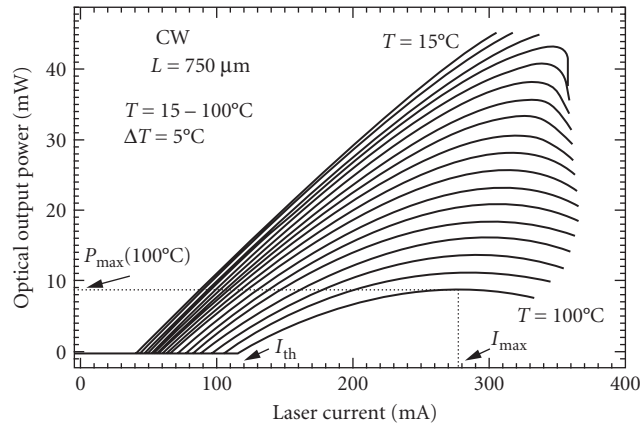


FIGURE 5 Typical experimental result for light out versus current in (the *L-I* curve). These results are for diodes operating at 1.3 μm , consisting of strained layer multiple quantum well InGaAsP lasers measured at a series of elevated temperatures.³

saturates as the current becomes large enough, particularly at high temperatures. The decrease in external slope efficiency with increasing current has contributions from the increase in leakage current with injection current, from junction heating that reduces recombination lifetime and increases threshold current, and because the internal absorption increases with injection current.

When there is more than one mode (longitudinal or transverse) in the LD, the L - I curve has *kinks* at certain current levels. These are slight abrupt reductions in light out as the current increases. After a kink the external slope efficiency may be different, along with different spatial and spectral features of the laser. These multimode lasers may be acceptable for low-cost communication systems, but high-quality systems require singlemode lasers that do not exhibit such kinks in their L - I curves.

Temperature Dependence of Laser Properties

Long-wavelength lasers are typically more sensitive to temperature than are GaAs lasers. This sensitivity is usually expressed as an experimentally measured exponential dependence of threshold on temperature T through $I_{th}(T) = I_0 \exp(T/T_0)$, where T_0 is a characteristic temperature (in kelvin) that expresses the measured thermal sensitivity. This formula is valid only over a limited temperature range, because it has no real physical derivation, but it has proved convenient and is often quoted. The data in Fig. 5 correspond to $T_0 \approx 80$ K.

The mechanisms for LD sensitivity to temperature depend on the material system. In InP-based long-wavelength DH lasers, T_0 is dominated by Auger recombination. However, in short-wavelength GaAs lasers and in strained layer QWs, Auger recombination is suppressed, T_0 is higher and is attributed to intervalence band absorption and/or carrier leakage over the heterostructure barrier, depending on the geometry. Typical long-wavelength DH lasers have T_0 in the range of 50 to 70 K. Typical strained layer QW lasers have T_0 in the range of 70 to 90 K, although higher T_0 can be achieved by incorporating aluminum in the barriers, with as high as 143 K reported.⁴ This temperature dependence limits the maximum optical power that can be obtained because of the phenomenon of *thermal runaway*, as shown at the highest temperatures in Fig. 5. While increasing the current usually increases the power, the junction temperature also increases (due to ohmic losses), so the threshold may increase and the output power may tend to decrease.

Various means for increasing T_0 have been explored; the most effective solution is an active layer of tensile strained QWs (discussed in Sec. 13.6). This has increased T_0 from approximately 50 K to as high as 140 K, comparable to that measured in GaAs. In double heterostructures, losses by carrier leakage can be reduced with a dual active region for double carrier confinement, which was demonstrated to achieve T_0 values as high as 180 K in 1.3- μ m InP lasers.⁵

The temperature dependence of long-wavelength lasers may limit their performance at high temperatures, which in turn limits where they can be used in the field. In practice, many long-wavelength lasers require thermoelectric coolers, particularly at higher power levels.

Spatial Characteristics of Emitted Light

Light is emitted out the facet of the laser diode after it has been waveguided in both directions. It will diverge by diffraction, more strongly in the out-of-plane dimension, where it has been more strongly guided. The diffracting output is sketched in Fig. 1. The spatial characteristics of the output can be estimated by fitting the guided light to a gaussian profile and then calculating the far-field pattern. The *out-of-plane* near-field profile for the lowest-order mode in an optical confinement layer of width d_g has been fit to a gaussian distribution $\exp(-x^2/w^2)$ with⁶

$$w = d_g \left(0.321 + \frac{2.1}{V^{3/2}} + \frac{4}{V^6} \right) \quad \text{for } 1.8 < V < 6 \quad (10)$$

where V is from Eq. (1). The far-field diffraction angle θ_{ff} has a slightly different gaussian fit, with a half-angle given by $\tan \theta_{ff} = \lambda/\pi w_o$, with⁶

$$w_o = d_g \left(0.31 + \frac{3.15}{V^{3/2}} + \frac{2}{V^6} \right) \quad \text{for } 1.5 < V < 6 \quad (11)$$

Experimental data can be compared to the gaussian beam formulation by remembering that the full-width half-maximum power FWHM $= w(2 \ln 2)^{1/2}$. For a typical strongly index-guided buried heterostructure laser, the far-field FWHM angle out-of-plane is approximately 1 rad and in-plane is approximately 1/2 rad. These angles are independent of current for index-guided lasers. Separate confinement heterostructure (SCH) lasers can have smaller out-of-plane beam divergences, more typically approximately 30°.

Single-mode lasers that are index-guided in the lateral direction (buried heterostructure and ridge waveguide) will obey the preceding equations, with lateral divergence angles varying from 30° to 10°, depending on design. This beam width will also be independent of current. When LDs are gain-guided laterally, the spatial variation of the gain leads to a complex refractive index and a curved wavefront. The result is that the equivalent gaussian lateral beam appears to have been emitted from somewhere inside the laser facet. The out-of-plane beam, however, is still index-guided and will appear to be emitted from the end facet. This means that the output of a gain-guided laser has *astigmatism*, which must be compensated for by a suitably designed external lens if the laser is to be focused effectively into a fiber (as discussed elsewhere in this *Handbook*).

If the laser emits a diverging gaussian beam with waist w , a lens can be used to focus it into a fiber. An effective thin lens of focal length f placed a distance d_1 after the laser facet will focus to a new waist w' given by

$$w'^2 = w^2 \frac{f^2}{b^2 + X_1^2} \quad (12)$$

where $b \equiv \pi w^2/\lambda$ and $X_1 \equiv d_1 - f$.

The new waist will be located a distance d_2 from the lens, where

$$X_2 = X_1 \frac{f^2}{X_1^2 + b^2} \quad (13)$$

and $X_2 \equiv d_2 - f$.

This new waist must be matched to the spatial mode of the fiber. Because of the large numerical aperture of LD emission, simple lenses may exhibit severe spherical aberration and typical coupling efficiencies may be only a few percent. Fiber systems usually utilize pigtailed fiber, butt-coupled as close as possible to the laser, without any intervening lens. Alternatively, a ball lens may be melted directly onto a fiber tip and placed near the laser facet. Sometimes *graded index* (GRIN) lenses are used to improve coupling into fibers.

Gain-guided LDs with electrode stripe widths of more than 5 μm usually emit multiple in-plane spatial modes that interfere laterally, producing a spatial output with multiple maxima and nulls. Such spatial profiles are suitable for multimode fiber applications, but cannot be efficiently coupled into single-mode fibers. They will diffract at an angle given by setting w equal to the minimum near-field feature size. If the stripe is narrow enough, gain-guided LDs are always single-mode, but these devices have a double-lobed far-field spatial profile (from the complex refractive index in the gain medium) that cannot be conveniently coupled into single-mode fibers.

Spectral Characteristics of Laser Light

In principle, a Fabry-Perot laser has many frequency modes with frequencies ν_m , given by requiring standing waves within the laser cavity. Since the m th mode obeys $m\lambda/2n = L$, where n is the refractive index experienced by the guided laser mode, then $\nu_m = mc/2nL$. Taking the differential, the frequency

difference between modes is $\Delta\nu = c/2n_{\text{eff}}L$, where the *effective group refractive index* $n_{\text{eff}} = n + v(\partial n/\partial v)$. For typical semiconductor lasers, $n = 3.5$ and $n_{\text{eff}} = 4$, so that when $L = 250\text{ }\mu\text{m}$, the frequency difference between modes is $\Delta\nu = 150\text{ GHz}$, and since $\Delta\lambda = (\lambda^2/c)\Delta\nu$, when $\lambda = 1.5\text{ }\mu\text{m}$, the wavelength spacing is $\Delta\lambda \approx 1\text{ nm}$.

At any given instant in time, a single spatial mode emits only one spectral mode. However, in multimode lasers, considerable *mode hopping* occurs, in which the LD jumps from one spectral mode to another very rapidly. Most spectral measurements are time averages and do not resolve this mode hopping, which can occur in nanoseconds or less. Explanations for the mode hopping typically involve *spatial hole burning* or *spectral hole burning*. Hole burning occurs when the available carrier density is momentarily depleted, either spatially or spectrally. At that time an adjacent mode with a different (longitudinal or lateral) spatial profile providing a different resonance wavelength may be more advantageous for laser action. Thus, the laser jumps to this new mode. The competition for available gain between different modes often induces semiconductor lasers to emit light at several frequencies.

One way to provide a single spectral mode is to ensure a single (lateral) spatial mode. It has been found that single spatial mode lasers usually have single spectral modes, at least at moderate power levels. However, this may change upon modulation. The only way to *ensure* a single-frequency LD is to ensure a single longitudinal mode by using distributed feedback, as discussed in Sec. 13.7.

Polarization

The light emitted from a typical LD is usually linearly polarized in the plane of the heterostructure. While gain in the semiconductor medium has no favored polarization dependence, the *transverse electric* (TE) *waveguide mode* (polarized in-plane) is favored for two reasons. First, the TE mode is slightly more confined than the *transverse magnetic* (TM) mode (polarized out-of-plane). Second, the Fresnel reflectivity off the cleaved end facets is strongly polarization sensitive. As waveguided light travels along the active stripe region, it can be considered to follow a zig-zag path, being totally internally reflected by the cladding layers. The total internal reflection angle for these waves is about 10° off the normal to the cleaved facets of the laser. This is enough to cause the TM waveguide mode to experience less reflectivity, while the TE-polarized mode experiences higher reflectivity and has a lower threshold. Thus, laser light from LDs is traditionally polarized in the plane of the junction.

The introduction of strain (Sec. 13.6) in the active layer changes the polarization properties, and the particular polarization will depend on the details of device geometry. In addition, DFB and DBR lasers (Sec. 13.7) do not have strong polarization preferences, and they must be carefully designed and fabricated if well-defined single polarization is required.

13.4 TRANSIENT RESPONSE OF LASER DIODES

When laser diodes are operated by direct current, the output is constant and follows the L - I curve discussed previously. When the LD is rapidly switched, however, there are transient phenomena that must be taken into account. Such considerations are important for any high-speed communication system, especially digital systems. The study of these phenomena comes from solving the semiconductor rate equations.⁷

Turn-On Delay

When a semiconductor laser is turned on abruptly by applying forward-biased current to the diode, it takes time for the carrier density to reach its threshold value and for the photon density to build up, as shown in the experimental data of Fig. 6.⁸ This means that a laser has an unavoidable turn-on time. The delay time depends on applied current and on carrier lifetime, which depends on carrier density

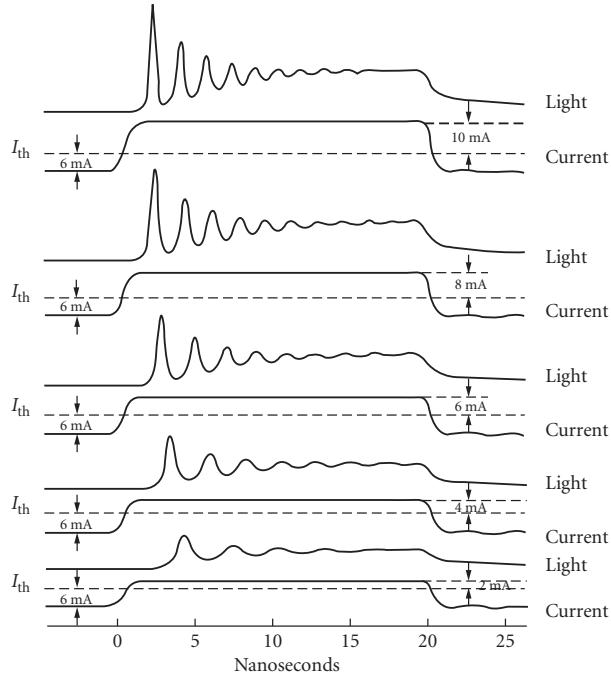


FIGURE 6 Experimental example of turn-on delay and relaxation oscillations in a laser diode when the operating current is suddenly switched from 6 mA below the threshold current of 177 mA to varying levels above threshold (from 2 to 10 mA). The GaAs laser diode was 50 μm long, with a SiO_2 -defined stripe 20 μm wide. Light output and current pulse are shown for each case.⁸

N , as shown in Eq. (8). Using a differential analysis, the turn-on time for a laser that is switched from an initial current I_i just below threshold to I just above threshold is

$$\tau_d = \tau'(N_{th}) \frac{I_{th} - I_i}{I - I_{th}} \quad (14)$$

where $\tau'(N)$ is a differential carrier recombination given by $1/\tau' = A_{nr} + 2BN + 3CN^2$. When the LD is initially off ($I_i = 0$) and $I \gg I_{th}$, the turn-on delay has an inverse current dependence: $\tau_d = \tau_e(N_{th}) I_{th}/I$.

When radiative recombination dominates, then $1/\tau_e \approx BN$ and $1/\tau' \approx 2BN \approx 2/\tau_e$, as seen by comparing the middle terms of Eqs. (11) and (22). For a 1.3- μm laser, $A_{nr} = 10^8/\text{s}$, $B = 10^{-10} \text{ cm}^3/\text{s}$, $C = 3 \times 10^{-29} \text{ cm}^6/\text{s}$, and $N_{th} \approx N_{tr} = 10^{18} \text{ cm}^{-3}$. Thus, $\tau_e = 5 \text{ ns}$ and a typical turn-on time at 1.5 times threshold current is 3 ns. The increase in delay time as the current approaches threshold is clearly seen in the data of Fig. 6. To switch a laser rapidly, it is necessary to switch it from just below threshold to far above threshold. However, Fig. 6 shows that under these conditions there are large transient oscillations, discussed in the following section.

Relaxation Oscillations

An important characteristic of the output of any rapidly switched laser (not just LDs) is the *relaxation oscillations* that can be observed in Fig. 6. These overshoots and undershoots occur as the photon and

carrier dynamics are coming into equilibrium. Such oscillations are characteristic of the nonlinear coupled laser rate equations and can be found by simple perturbation theory. These relaxation oscillations have a radian frequency Ω_R given, to first order, by⁹

$$\Omega_R^2 = \frac{1 + \chi}{\tau_e \tau_p} \frac{I - I_{th}}{I_{th}} \quad (15)$$

where I is the current, I_{th} is the current at threshold, τ_p is the photon lifetime in the cavity, given by $\nu_g \tau_p = (\alpha_i + \alpha_m)^{-1}$, and $\chi = \Gamma a_L \nu_g \tau_p = \Gamma a_T \tau_p = \Gamma [a_L / (\alpha_i + \alpha_m)]$, where a_L was previously defined and is approximately the unpumped absorption loss in the active medium. The factor χ is then the ratio of the unpumped absorption loss to the cavity loss, typically 1 to 3 for semiconductor lasers. It can also be shown that $\chi = I_{tr} / (I_{th} - I_{tr})$, where I_{tr} is the current at transparency.

At 50 percent above threshold, and when $\chi \approx 1$, the time between successive relaxation oscillation maxima is approximately the geometric mean of the carrier and photon lifetimes: $\Omega_R^2 \approx 1 / \tau_e \tau_p$. Typical LD numbers are $\tau_e = 10$ ns and $\tau_p = 3$ ps, so at 1.5 times threshold current, the relaxation oscillation frequency is $f_R = \Omega_R / 2\pi = 1$ GHz, and the time between relaxation oscillation peaks is 1 ns.

The decay rate of these relaxation oscillations γ_R is given by

$$2\tau_e \gamma_R = 1 + \chi \frac{I - I_{th}}{I_{tr}} = 1 + (1 + \chi) \frac{I - I_{th}}{I_{th}} = 1 + \Omega_R^2 \tau_e \tau_p \quad (16)$$

and is roughly 2 ns at twice threshold for typical heterostructure lasers. At 1.5 times threshold and when $\chi \approx 1$, Eq. (16) gives $\gamma_R \approx 1 / \tau_e$. The relaxation oscillations last approximately as long as the spontaneous emission lifetime of the carriers.

This first-order analysis has several assumptions that do not seriously affect the relaxation oscillation frequency, but will overestimate the time that relaxation oscillations are present. The analysis ignores *gain saturation*, which reduces gain with increased photon density P and is important at high optical powers. It also ignores the rate of spontaneous emission in the cavity, R_{sp} , which is important at small optical powers. Finally, it ignores the impact of changing carrier density on spontaneous emission lifetime. A more typical experimental decay rate for lasers at 1.3 μm wavelength is $\gamma \approx 3 / \tau_e$. A more exact theoretical formulation can be found in Ref. 10.

The number of relaxation oscillations (before they die out) in an LD at 1.5 times threshold is roughly $\Omega_R / \gamma_R \propto (\tau_e / \tau_p)^{1/2}$. Shorter carrier lifetimes mean fewer relaxation oscillations, because carriers reach steady state more rapidly. Shorter carrier lifetimes also mean faster turn-on times. Carrier lifetimes are shortened by high carrier densities, an important regime for high-speed semiconductor lasers; high carrier densities can be achieved by using as small an active region as possible (such as QWs) and by reducing the reflectivity of the laser facets to raise the threshold carrier density.

The relaxation oscillations disappear if the current is just at threshold. However, we've also seen that under this situation the turn-on time becomes very long. It is more advantageous to turn the laser on fast, suffering the relaxation oscillations and using LDs designed to achieve a high decay rate, which means using LDs with the highest possible relaxation oscillation frequency.

The relaxation oscillation can also be expressed in terms of the photon density P inside the laser cavity, or in terms of the power out, P_{out} :

$$\Omega_R^2 = g_T g'_T P = \frac{g'_T P}{\tau_m} = g'_T \left(\frac{P_{out}}{h\nu V_a} \frac{\alpha_m + \alpha_i}{\alpha_m} \right) \quad (17)$$

where g_T is the modal gain per unit time and $g'_T \equiv \partial g_T / \partial N = \Gamma a_T / N_{tr}$. This formulation makes use of the relationship between output power and internal photon density in a cavity of volume V_a :

$$P_{out} = h\nu \left(\frac{P}{\tau_m} \right) V_a \quad (18)$$

where $\tau_m = (v_g \alpha_m)^{-1}$ is the time it takes light to bleed out the mirror. Note that the relaxation oscillation frequency increases as the photon density increases, confirming that smaller laser dimensions are better for high-speed modulation.

Relaxation oscillations can be avoided by biasing the laser just below threshold (letting the laser “simmer”). Fiber optic systems are designed to reduce the impact of these relaxation oscillators, whether they are digital telecommunication systems, or high-speed analog CATV systems. Many communication applications avoid these issues by operating the LD at steady state with DC current applied and use external modulators.

Modulation Response and Gain Saturation

The *modulation response* describes the amplitude of the modulated optical output as a function of frequency under small-signal current modulation. Laser diodes have a resonance in the modulation response at the relaxation oscillation frequency, as indicated by the experimental data in Fig. 7.¹¹ It is more difficult to modulate the laser above the relaxation oscillation frequency. Carrying out a small-signal expansion of the rate equations around photon density P , the modulation response (in terms of current density J) is¹²

$$\frac{\partial P}{\partial J} = \frac{(1/ed)(g'_T P + \beta_{sp}/\tau_e)}{(g'_T P/\tau_p + \beta_{sp}/\tau_e \tau_p - \omega^2) + j\omega(g'_T P + 1/\tau_e)} \quad (19)$$

where β_{sp} is the fraction of spontaneous emission that radiates into the mode. This modulation response has the form of a second-order low-pass filter. Resonance occurs when $\omega^2 \approx g'_T P/\tau_p = \Omega_R^2$ (from Eq. (17), with negligible internal loss); that is, at the relaxation oscillation frequency.

The modulation response at a frequency well below the relaxation oscillation frequency can be expressed as $\partial P_{out}/\partial I$ using Eq. (19) in the limit when $\omega \rightarrow 0$. From $\partial P/\partial J = \tau_p/ed$ and using Eq. (18), the low frequency modulation response is

$$\frac{\partial P_{out}}{\partial I} = \frac{h\nu}{\tau_m} V_a \frac{\partial P}{\partial J} \cdot \frac{1}{wL} = \frac{h\nu}{e} \cdot \frac{\tau_p}{\tau_m} = \frac{h\nu}{e} \cdot \frac{\alpha_m}{\alpha_m + \alpha_i} \quad (20)$$

which is expected from Eq. (5) for $\eta_i \rightarrow 1$.

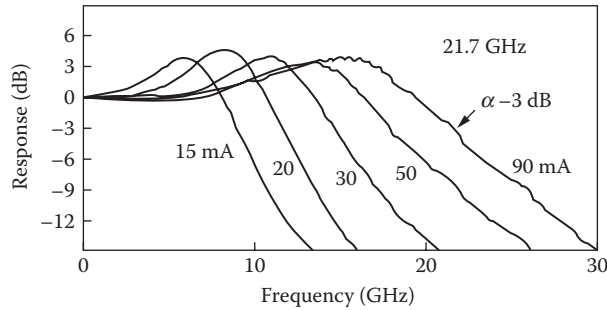


FIGURE 7 Measured small-signal modulation response of a high-speed DFB laser at several bias levels. Zero-dB modulation response is defined in terms of the low-frequency modulation response, Eq. (32).¹¹

The 3-dB modulation radian frequency bandwidth ω_B can be expressed in terms of the relaxation oscillation parameters by¹³

$$\omega_B^2 = \Omega_R^2 - \gamma_R^2 + 2\sqrt{\Omega_R^2(\Omega_R^2 + \gamma_R^2 + \gamma_R^4)} \quad (21)$$

The parameters are strongly power dependent and ω_B increases with optical power. When $\gamma_R \ll \Omega_R$, which means the time carriers remain in the active region is longer than the time photons remain in the cavity, the 3-dB bandwidth $\omega_B \approx \sqrt{3}\Omega_R \propto \sqrt{P}$. At high optical powers the presence of gain saturation (reduced gain at high optical power densities) must be included; the modulation bandwidth saturates, and the limiting value depends on the way that the gain saturates with photon density.

Using a heuristic expression for modal gain saturation of the form $g_T(N, P) = g'_T(N - N_o)/(1 + P/P_s)^{1/2}$, the limiting value of the modulation bandwidth at high optical powers is $\omega_B^2 = 3g'_T P_s / 2\tau_p$, where P_s is the saturation photon density. A typical modulation bandwidth for an InGaAsP laser operating at 1.55- μm wavelength is $\omega_B = 20$ to 40 GHz.

Frequency Chirping

When the carrier density in the active region is rapidly changed, the refractive index n also changes rapidly, causing a frequency shift proportional to $\partial n / \partial t$. This broadens the laser linewidth from its original width of approximately 100 MHz into a double-peaked profile with a gigahertz linewidth, as shown in the experimental results of Fig. 8.¹⁴ The frequency spread is directly proportional to the dependence of the refractive index n on carrier density N . This is a complex function that depends on wavelength and degree of excitation, but for simplicity a Taylor expansion around the steady-state carrier density N_o can be assumed: $n = n_o + n_1(N - N_o)$, where $n_1 \equiv \partial n / \partial N$. The (normalized) ratio of this slope to that of the modal gain per unit length g_L is called the cavity *linewidth enhancement factor* β_c :

$$\beta_c \equiv -2k_o \frac{\partial n / \partial N}{\partial g_L / \partial N} = -2k_o \frac{n_1}{g'_L} \quad (22)$$

Sometimes the linewidth enhancement factor is called the α -factor. Typical values lie between 2 and 6.

The magnitude of the frequency spread between the double lobes of a chirped pulse, $2\delta\omega_{CHP}$ can be estimated in the small-signal and large-signal regimes from analyzing the time dependence of a modulated pulse in terms of the sum of all frequency components, shown next.¹⁵

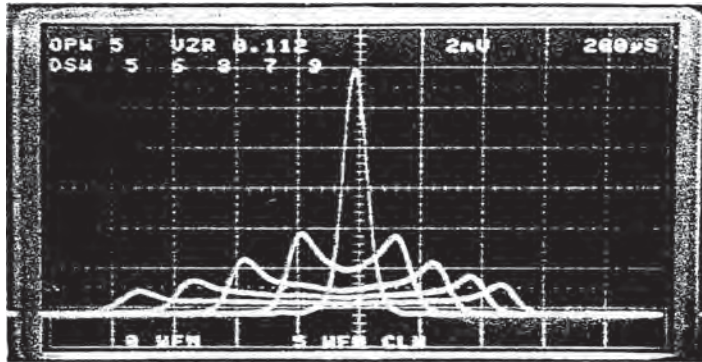


FIGURE 8 Time-averaged power spectra of 1.3- μm InGaAsP laser under sinusoidal modulation at 100 MHz. Horizontal scale is 0.05 nm per division. Spectrum broadens with increase in modulation current due to frequency chirping.¹⁴

Small Signal Modulation For a modulation frequency ω_m that is less than the relaxation oscillation frequency, and assuming that the carrier lifetime is longer than the photon lifetime, $\gamma_R \ll \Omega_R$, a small modulation current I_m will cause a frequency chirp of magnitude

$$\delta\omega_{\text{CH}} = \frac{\beta_c I_m h\nu}{2eP_{\text{out}}} \left(\frac{\alpha_m}{\alpha_m + \alpha_i} \right) \sqrt{\omega_m^2 + \gamma_p^2} \quad (23)$$

where $\gamma_p = N/P(\beta_{\text{sp}}/\tau_c) - (\partial g_T/\partial P)P$ (remembering that $\partial g_T/\partial P$ is negative). The origin of chirp is the linewidth enhancement factor β_c . It will be largest for gain-guided devices where β_c is the maximum. The chirp will be smaller in lasers with $\alpha_m \ll \alpha_p$ such as will occur for long lasers, where mirror loss is amortized over a longer length, but such lasers will have a smaller differential quantum efficiency and smaller relaxation oscillation frequency. Typical chirp bandwidths (using $\delta\lambda_{\text{CH}} = \lambda^2 \delta\omega_{\text{CH}}/2\pi c$) at 25-mA modulation current can vary from $\delta\lambda_{\text{CH}} = 0.2$ nm for gain-guided lasers to $\delta\lambda_{\text{CH}} = 0.03$ nm for ridge waveguide lasers. At a wavelength of 1 micrometer in frequency units this would be 60 GHz for gain-guided lasers and 9 GHz for ridge waveguide lasers.

Large Signal Modulation There is a transient frequency shift during large-signal modulation given by

$$\delta\omega_{\text{CH}} = \frac{\beta_c}{2} \left(\frac{1}{P} \frac{\partial P}{\partial t} \right) \quad (24)$$

When a gaussian shape pulse is assumed as $\exp(-t^2/T^2)$, the chirp becomes $\delta\omega_{\text{CH}} \approx \beta_c/T$.

The importance of the linewidth enhancement factor β_c is evident from this section; its existence will inevitably broaden modulated laser linewidths. Because it is a basic characteristic of the laser medium itself, there are no design freedoms to reduce chirp under large signal modulation.

13.5 NOISE CHARACTERISTICS OF LASER DIODES

Noise in LDs results from fluctuations in spontaneous emission and from the carrier generation-recombination process (shot noise). To analyze the response of LDs to noise, one starts with rate equations, introduces Langevin noise sources as small perturbations, and linearizes (performs a small-signal analysis). Finally, one solves in the frequency domain using Fourier analysis.^{16,17} Only the results are given here.

Relative Intensity Noise

Noise at a given frequency is described in terms of *relative intensity noise* (RIN), defined by

$$\text{RIN} = \frac{\Delta P^2}{P_T^2} \quad (25)$$

where ΔP^2 is the photon noise spectral density (noise per unit frequency interval), and P_T is the total photon number, $P_T = PV_a$. The solution to the analysis for RIN in an LD is

$$\text{RIN} = \frac{2\beta_{\text{sp}} I_{\text{th}}}{ePV_a} \cdot \frac{1/\tau'^2 + \omega^2 + (\partial g_T/\partial N)^2 P/(\beta_{\text{sp}} V_a)}{[(\Omega_R - \omega)^2 + \gamma_R^2][(\Omega_R + \omega)^2 + \gamma_R^2]} \quad (26)$$

where β_{sp} is the fraction of spontaneous emission emitted into the laser cavity. As before, the photon density P can be related to the optical power out both facets by $P_{\text{out}} = (h\nu) PV_a/\tau_m$. Note the significant enhancement of noise near the relaxation oscillation frequency $\omega = \Omega_R$, where the noise has

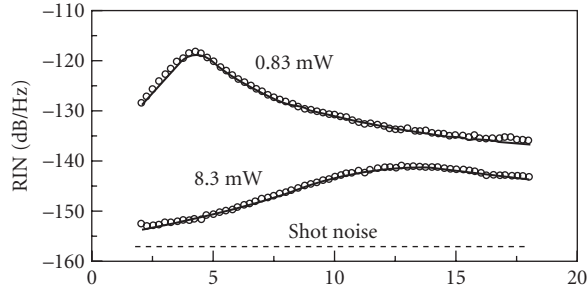


FIGURE 9 Measured relative intensity noise as a function of frequency in a multiple quantum well 1.5- μm laser diode, for optical power near threshold and high above threshold. The shot noise level for the higher power measurement is also shown.¹⁸

its maximum value. An example of RIN as a function of frequency ω is shown in Fig. 9,¹⁸ for both low power and high power, showing that the RIN goes up as the total optical power decreases.

At low frequencies, and for $\gamma_R \ll \Omega_R$, the noise is proportional to the inverse fourth power of the relaxation oscillation frequency. Clearly, it is advantageous to use as high a relaxation oscillation frequency as possible to reduce RIN. Since the relaxation oscillation frequency is proportional to the square root of the power P , the RIN increases as $1/P^3$ as the power decreases. Inserting the expression for Ω_R into Eq. (26) gives

$$\text{RIN}_{\text{lf}} = \frac{2\beta_{\text{sp}} I_{\text{th}} \tau_m^2 V_a^2}{e P_T^3 (\partial g_T / \partial N)^2} \left[\frac{1}{\tau'^2} + \left(\frac{\partial g_T}{\partial N} \right)^2 \frac{P}{\beta_{\text{sp}} V_a} \right] \quad (27)$$

where τ' is the differential carrier recombination time and P_T is the total photon number. Usually, the first term dominates. It can be seen that the volume of the active laser region V_a should be as small as possible, consistent with maintaining a significant power out.

Signal-to-Noise Ratio

The *signal-to-noise ratio* (SNR) can be found in terms of the relaxation oscillation parameters using the simplified expression for RIN (which assumes $\tau_e \Omega_R \gg \gamma_R \Omega_R \gg 1$) and the total photon number:

$$\text{SNR}^2 = \frac{2\gamma_R e}{\beta_{\text{sp}} I_{\text{th}}} P_T = \frac{2\gamma_R e \tau_m}{\beta_{\text{sp}} I_{\text{tr}} h\nu} P_{\text{out}} \quad (28)$$

As expected, the SNR ratio increases linearly with output power and increases as the amount of spontaneous emission decreases. This expression can be simplified further by recalling that $\gamma_R \approx \tau_e/3$.

Gain saturation at high optical powers eventually limits the SNR to about 30 dB; while at powers of a few milliwatts it is 20 dB, with intensity fluctuations typically close to 1 percent.

Mode Partition Noise in Multimode Laser Diodes

The preceding discussion of noise holds qualitatively for multimode lasers as long as all the laser modes are included. However, measurements made on any one mode show much more noise, particularly at low frequencies. This is due to the mode-hopping discussed previously, and is referred

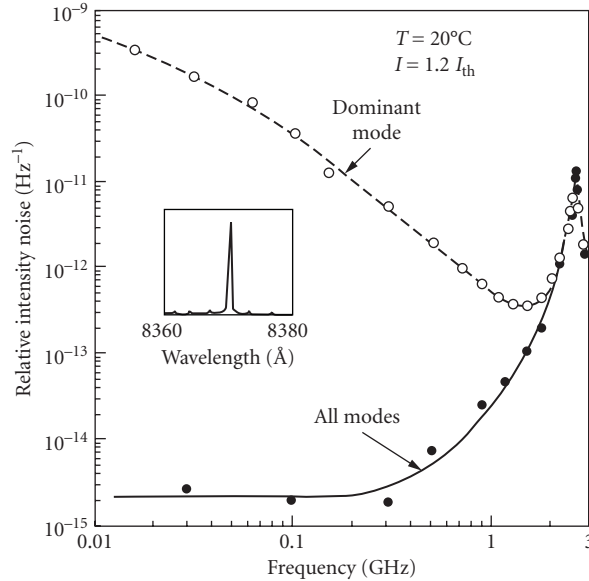


FIGURE 10 Effect of mode partition noise on relative intensity noise in multimode lasers. Experimentally observed intensity-noise spectra in all modes (solid curve) or in dominant mode (dashed curve). Inset shows spectrum of average mode power.¹⁹

to as *mode partition noise*. That is, the power partitions itself between different laser modes in a way that keeps the overall intensity relatively constant, as shown by the solid line in Fig. 10. The power in each mode is not a steady function of time, because the power distribution among the modes changes with time. Whenever the distribution changes, the power output undergoes fluctuation, leading to a noise term on a nominally stable output. This leads to the enhanced RIN on the dominant mode in Fig. 10.¹⁹ Even an output whose spectrum looks nominally single-mode, as shown in the inset of Fig. 10, can have a large RIN on the dominant mode. This is because the spectrum is time averaged. A sidemode does not contain 5 percent of the power all the time; for example, it contains 100 percent of the power for 5 percent of the time. This causes the very large RIN observed. The solution to avoiding this noise is to insist on a single longitudinal mode by using distributed feedback. Since lasers for telecommunication applications are typically single mode, we will not consider mode partition noise further. It becomes important for data communications based on multimode lasers, however.

Phase Noise—Linewidth

The fundamental linewidth of a laser is given by the stochastic process of spontaneous emission, as first derived by Schawlow and Townes in the very early days of lasers. In a laser diode, additional noise enters from the stochastic process of carrier injection. Because the refractive index is a function of the carrier density, changes in carrier density cause changes in refractive index, which in turn create phase noise.

The formula for the radian frequency linewidth of a LD includes the *linewidth enhancement factor* β_c :

$$\delta\omega = (1 + \beta_c^2)\delta\omega_0 \quad (29)$$

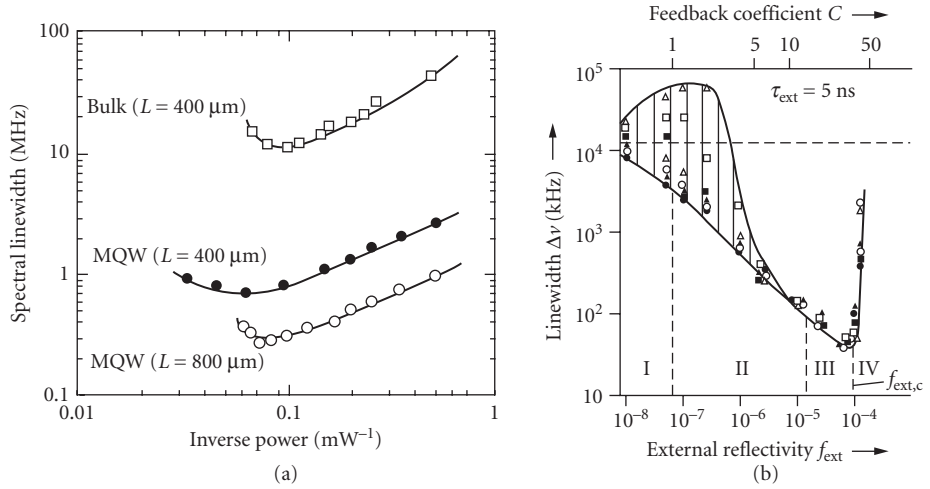


FIGURE 11 Linewidth of a laser diode: (a) DFB lasers as a function of inverse power, comparing bulk active regions and multiple quantum well active regions²⁰ and (b) Fabry-Perot laser as a function of feedback parameter, with corresponding feedback regimes.²¹

where $\delta\omega_o$ the original Schawlow-Townes linewidth is given by

$$\delta\omega_o = \frac{\beta_{sp} I_{th} h\nu \tau_m}{2eP_{out}} \quad (30)$$

Typical value of the linewidth enhancement factor is $\beta_c = 5$. The linewidth decreases inversely as the laser power increases. However, as shown in the experimental data in Fig. 11a,²⁰ at high enough power (above 10 mW) the linewidth narrowing saturates at approximately 1 to 10 MHz and then begins to broaden again at even higher power levels. It is also possible to reduce the linewidth by using QWs and increasing the cavity length (to decrease N_{th} and increase P).

External Optical Feedback and Coherence Collapse

Laser diodes are extremely sensitive to weak time-delayed feedback, such as from reflections off the front ends of fiber pigtails. These fed-back signals can result in mode hopping, strong excess noise, and chaotic behavior in the *coherence collapse* regime. Some of the features of feedback-induced noise are outlined here.

Various regimes of feedback can be characterized by a *feedback parameter* C :

$$C = \sqrt{f_{ext}} C_e \frac{\tau_{ext}}{\tau_L} \sqrt{1 + \beta_c^2} \quad (31)$$

where f_{ext} is the fraction of the emitted power that is externally reflected back into the laser. The external coupling factor $C_e = (1 - R)/\sqrt{R}$, where R is the reflectivity of the laser facet and β_c is the linewidth enhancement factor. The external round-trip time delay is τ_{ext} and the laser round-trip time is τ_L . Figure 11b²¹ gives an example of the linewidth of a semiconductor laser versus the external coupling factor C in which various feedback regimes are indicated.

Regimes of Feedback The following provides a useful classification scheme:²²

Regime I ($C < 1$): At the lowest levels of feedback, narrowing or broadening of the emission line is observed, depending on the phase of the feedback.

Regime II ($C > 1$): At higher levels of feedback, mode hopping between different external cavity modes may appear.

Regime III ($C \gg 1$): Further increasing the levels of feedback, the laser is observed to operate in a mode with the smallest linewidth.

Regime IV (coherence collapse): At yet higher feedback levels, satellite modes appear, separated from the main mode by the relaxation oscillation frequency. These grow as the feedback increases and the laser line eventually broadens, due to the collapse of the coherence of the laser. This regime does not depend on distance from the laser to the reflector.

Regime V (external cavity laser): This regime of stable operation can be reached only with an antireflection-coated laser output facet to ensure a two-mirror cavity with the largest possible coupling back into the laser, and is not of concern here.

A quantitative discussion of these regimes follows.²¹ Assume that the coupling efficiency from the laser into the fiber is η . Because feedback requires a double pass, the fraction of emitted light fed back into the laser is $f_{\text{ext}} = \eta^2 R_c$, where R_c is the reflectivity from the end of the fiber. The external reflection changes the overall cavity reflectivity and therefore the threshold gain, depending on its phase relative to the phase inside the cavity. Possible modes are defined by the threshold gain and the requirement that an effective round-trip phase $= m\pi$. But a change in the threshold gain also changes the refractive index and the phase through the linewidth enhancement factor β_c .

Regime I For very weak feedback, there is only one solution when the laser phase is set equal to $m\pi$, so that the laser frequency ω is at most slightly changed and its linewidth $\Delta\omega_0$ will be narrowed or broadened, as the external reflection adds to or subtracts from the output of the laser. The linewidth lies between extremes:

$$\frac{\Delta\omega_0}{(1+C)^2} \ll \Delta\omega \ll \frac{\Delta\omega_0}{(1-C)^2} \quad (32)$$

The system performance moves toward regime II as $C \rightarrow 1$. Note that at $C = 1$ the maximum value predicts an infinite linewidth. This indicates that even very small feedback can cause wide spectral response, as long as $C \sim 1$.

Regime II For higher feedback with $C > 1$, several solutions with the round-trip laser phase $= m\pi$ may exist. Linewidth broadening occurs because the single external cavity mode now has split into a dual mode, accompanied by considerable phase noise. Mode hopping gives linewidth broadening and intensity noise. This is a low-frequency noise with a cutoff frequency of about 10 MHz.

Regime III With increasing feedback, the mode splitting increases up to a frequency $\Delta\omega = 1/\tau_{\text{ext}}$ and the cutoff frequency for mode hopping noise decreases until only one of the split modes survives. To understand which mode survives, it is important to realize that in regime III, stable and unstable modes alternate with increasing phase. Because $\beta_c \neq 0$, the mode with the best phase stability (corresponding to the minimum linewidth mode) does not coincide with the mode with minimum threshold gain. The minimum linewidth mode predominates, due to its phase stability. This mode remains relatively stable and has the same emission frequency as the laser without feedback. This mode predominates as long as the inverse of the linewidth of the solitary laser is larger than the external cavity round-trip time. In this regime, the laser is stably phase locked by the feedback.

Regime IV The stable linewidth of regime III collapses as the fraction of power fed back f_{ext} increases to a critical value. There is considerable discussion of the physical mechanism that leads to

this coherence collapse. The existence of this regime has been demonstrated by simulation, through numerical solution of the rate equations. Fitting theoretical analyses to experimental results indicates that the onset of coherence collapse occurs when the feedback factor is larger than a critical value given by²³

$$C \geq C_{\text{crit}} = 2\gamma_R \tau_{\text{ext}} \frac{1 + \beta_c^2}{\beta_c^2} \quad (33)$$

where γ_R is the damping rate of the relaxation oscillations. As the feedback level approaches the critical value C , undamped relaxation oscillations appear and oscillations of carrier density through the linewidth enhancement factor β_c induce the phase of the field to oscillate. This analytical result assumes that $\tau_{\text{ext}} \gg 1/\gamma_R$. An approximate solution for γ_R gives $C_{\text{crit}} = \tau_{\text{ext}}/\tau_e \times (1 + \beta_c^2)/\beta_c^2$.

Cavity Length Dependence and RIN In some regimes the stable regions depend on the length of the external cavity, that is, the distance from the extra reflection to the laser diode. These regions have been mapped out for two different laser diodes, as shown in Fig. 12.²⁴ The qualitative dependence on the distance of the laser to the reflection should be similar for all LDs.

The RIN is low for weak to moderate levels of feedback but increases tremendously in regime IV. The RIN and the linewidth are strongly related (see Fig. 11); the RIN is suppressed in regimes III and V.

Low-Frequency Fluctuations When a laser operating near threshold is subject to a moderate amount of feedback, chaotic behavior evolves into *low-frequency fluctuations* (LFF). During LFF the average laser intensity shows sudden dropouts, from which it gradually recovers, only to drop out again after some variable time, typically on the order of tens of external cavity round-trips. This occurs in regimes of parameter space where at least one stable external cavity mode exists, typically at the transition between regimes IV and V. Explanations differ as to the cause of LFF, but they appear to originate in strong-intensity pulses that occur during the build-up of average intensity, as a form of mode locking, being frustrated by the drive toward maximum gain. Typical frequencies for LFF are 20 to 100 MHz, although feedback from reflectors very close to the laser has caused LFF at frequencies as high as 1.6 GHz.

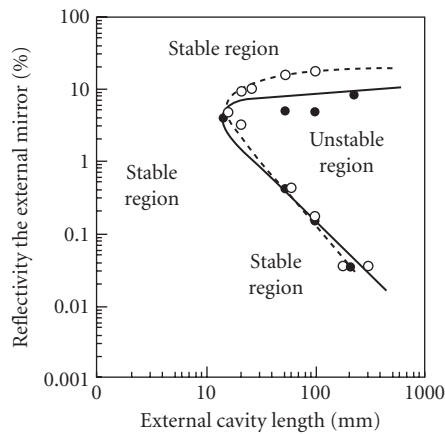


FIGURE 12 Regimes of stable and unstable operation for two laser diodes (○ and ●) when subject to external feedback at varying distances and of varying amounts.²⁴

Conclusions A laser diode subject to optical feedback exhibits a rich and complex dynamic behavior that can enhance or degrade the laser's performance significantly. Feedback can occur through unwanted back reflections—for instance, from a fiber facet—and can lead to a severe degradation of the spectral and temporal characteristics, such as in the coherence collapse or in the LFF regime. In both regimes, the laser intensity fluctuates erratically and the optical spectrum broadens, showing large sidebands. Because these unstable regimes can occur for even minute levels of feedback, optical isolators or some other means of reflection prevention are often used in systems applications.

13.6 QUANTUM WELL AND STRAINED LASERS

Introducing quantum wells and strain into the active region of diode lasers has been shown to provide higher gain, greater efficiency, and lower threshold. Essentially all high-quality lasers for optical communications use one or both of these means to improve performance over bulk heterostructure lasers.

Quantum Well Lasers

We have seen that the optimum design for low-threshold LDs uses the thinnest possible active region to confine free carriers, as long as the laser light is waveguided. When the active layer has a thickness less than a few tens of nanometers (hundreds of angstroms), it becomes a *quantum well*. That is, the layer is so thin that the confined carriers have energies that are quantized in the growth direction z , as described in Chap. 19 in Vol. II of this *Handbook*. This changes the density of states and the gain (and absorption) spectrum. While bulk semiconductors have an absorption spectrum near the band edge that increases with photon energy above the bandgap energy E_g as $(h\nu - E_g)^{1/2}$, quantum wells have an absorption spectrum that is steplike in photon energy at each of the allowed quantum states. Riding on this steplike absorption is a series of exciton resonances at the absorption steps that occur because of the Coulomb interaction between free electrons and holes, which can be seen in the spectra of Fig. 13.²⁵ These abrupt absorption features result in much higher gain for QW lasers than for bulk semiconductor lasers. The multiple spectra in Fig. 13 record the reduction in absorption as the QW states are filled with carriers. When the absorption goes to zero, transparency is reached. Figure 13 also shows that narrower wells push the bandgap to higher energies, a result of quantum confinement. The QW thickness is another design parameter in optimizing lasers for telecommunications.

Because a single quantum well (SQW) is so thin, its optical confinement factor is small. It is necessary to use either multiple QWs (separated by heterostructure barriers that contain the electron wave functions within individual wells) or to use a guided wave structure that focuses the light into a SQW. The latter is usually a GRIN structure, as shown in Fig. 2*d*. Band diagrams as a function of distance in the growth direction for typical QW separate confinement heterostructures are shown in Fig. 14. The challenge is to properly confine carriers and light using materials that can be reliably grown and processed by common crystal growth methods.

Quantum wells have provided significant improvement over bulk active regions, as originally observed in GaAs lasers. In InP lasers, Auger recombination and other losses come into play at the high carrier densities that occur in quantum-confined structures. However, it has been found that strain in the active region can improve the performance of quaternary QW lasers to a level comparable with GaAs lasers. Strained QW lasers are described in the following section.

The LD characteristics described in Secs. 13.2 to 13.5 hold for QW lasers as well as for bulk lasers. While the local gain is larger, the optical confinement factor will be much smaller. Equation (1) shows that the parameter V becomes very small when d_g is small, and Eq. (2) shows Γ_g is likewise small. With multiple quantum wells (MQWs), d_g can be the thickness of the entire region containing the MQWs and their barriers, but Γ must now be multiplied by the filling factor Γ_f of the QWs within the MQW region. If there are N_w wells, each of thickness d_w , then $\Gamma_f = N_w d_w / d_g$. With a GRINSCH structure, the optical confinement factor depends on the curvature of its refractive gradient near the center of the guide.

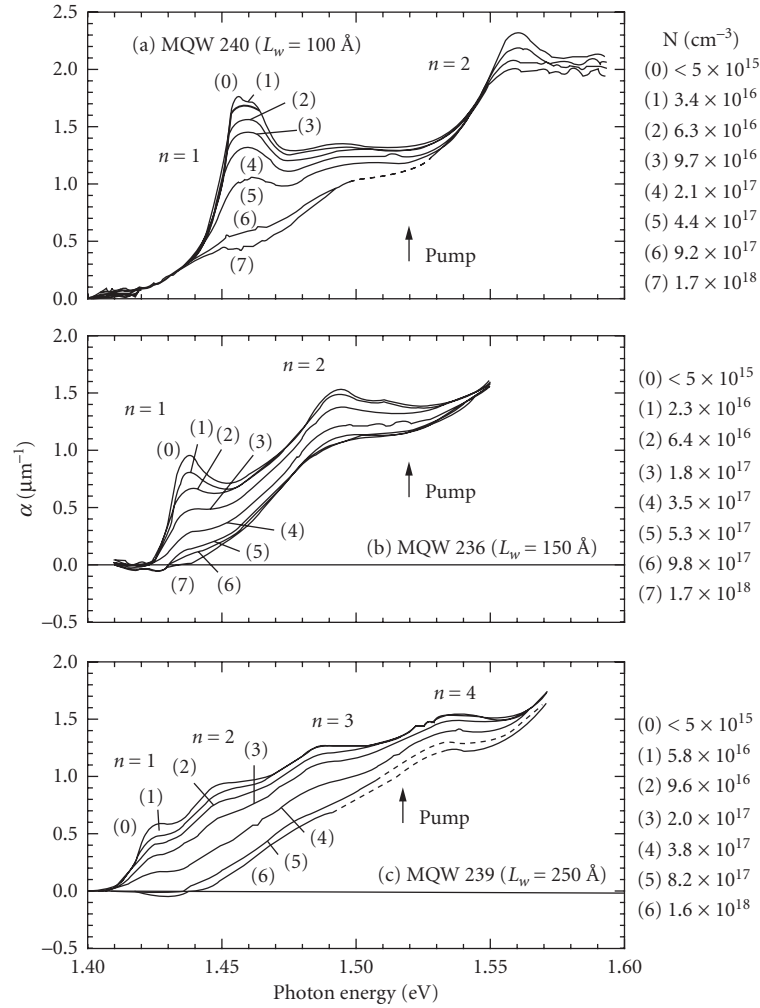


FIGURE 13 Absorption spectrum for multiple quantum wells of three different well sizes, for varying levels of optically induced carrier density, showing the decrease in absorption toward transparency. Note the stronger excitonic resonances and increased bandgap with smaller well size.²⁵

Different geometries have subtle differences in performance, depending on how many QWs are used and the extent to which a GRINSCH structure is dominant. The lowest threshold current densities have been reported for the highest Q cavities (longest lengths or highest reflectivities) using SQWs. However, for lower Q cavities the lowest threshold current densities are achieved with MQWs, even though they require higher carrier densities to achieve threshold. This is presumably because Auger recombination depends on the cube of the carrier density, so that SQW lasers will have excess losses, due to their higher carrier densities. In general, MQWs are a better choice in long-wavelength lasers, while SQWs have the advantage in GaAs lasers. However, with MQW lasers it is important to realize that the transport of carriers moving from one well to the next during high-speed modulation must be taken into account. In addition, improved characteristics of strained layer QWs make SQW devices more attractive.

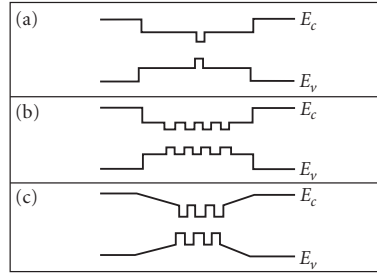


FIGURE 14 Typical band diagrams (energy of conduction band E_c and valence band E_v versus growth direction) for quantum wells in separate confinement laser heterostructures: (a) single quantum well; (b) multiple quantum wells; and (c) graded index separate confinement heterostructure (GRINSCH) and multiple quantum wells.

Strained Layer Quantum Well Lasers

Active layers containing *strained quantum wells* have proven to be an extremely valuable advance in high-performance long-wavelength InP lasers. They have lower thresholds, enhanced differential quantum efficiency η_D , larger characteristic temperature T_o , reduced linewidth enhancement factor β_c (less chirp), and enhanced high-speed characteristics (larger relaxation oscillation frequency Ω_R), compared to unstrained QW and bulk devices. This results from the effect of strain on the energy-versus-momentum band diagram. Bulk semiconductors have two valence bands that are degenerate at the potential well minimum (at momentum $k_x = 0$); as shown in Fig. 15a. They are called

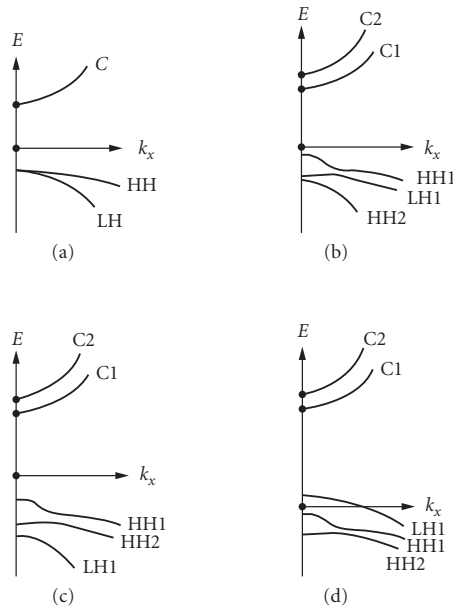


FIGURE 15 The effect of strain on the band diagram (energy E versus in-plane momentum k_x) of III-V semiconductors: (a) no strain; (b) quantum wells; (c) compressive strain; and (d) tensile strain.

heavy-hole (HH) and light-hole (LH) bands, since the smaller curvature means a heavier effective mass. Quantum wells lift this degeneracy, and interaction between the two bands near momentum $k = 0$ causes a local distortion in the formerly parabolic bands, also shown in Fig. 15b. There are now separately quantized conduction bands (C_1 and C_2) and a removal of the valence band degeneracy, with the lowest energy heavy holes HH_1 no longer having the same energy as the lowest energy light holes LH_1 at $k = 0$. The heavy hole effective mass becomes smaller, more nearly approaching that of the conduction band. This allows population inversion to become more efficient, increasing the differential gain; this is one factor in the reduced threshold of QW lasers.²⁶

Strain additionally alters this structure in a way that can improve performance even more. Compressive strain in the QW moves the heavy-hole and light-hole valence bands further apart and further reduces the hole effective mass (Fig. 15c). Strain also decreases the heavy-hole effective mass by a factor of 2 or more, further increasing the differential gain and reducing the threshold carrier density. Higher differential gain also results in a smaller linewidth enhancement factor. Tensile strain moves the heavy-hole and light-hole valence bands closer together (Fig. 15d). In fact, at one particular tensile strain value these bands become degenerate at $k = 0$. Further tensile strain results in the light hole having the lowest energy at $k = 0$. These lasers will be polarized TM, because of the angular momentum properties of the light-hole band. This polarization has a larger optical matrix element, which can enhance the gain within some wavelength regions.

In addition to the heavy- and light-hole bands, there is an additional, higher-energy valence band (called the *split-off band*, not shown in Fig. 15) which participates in Auger recombination and intervalence band absorption, both of which reduce quantum efficiency. In unstrained material there is a near-resonance between the bandgap energy and the difference in energy between the heavy-hole and split-off valence bands, which enhances these mechanisms for nonradiative recombination. Strain removes this near-resonance and reduces those losses that are caused by Auger recombination and intervalence band absorption. This means that incorporating strain is essential in long-wavelength laser diodes intended to be operated at high carrier densities. The reliability of strained layer QW lasers is excellent, when properly designed. However, strain does increase the intraband relaxation time, making the gain compression factor worse, so strained lasers tend to be more difficult to modulate at high-speed.

Specific performance parameters depend strongly on the specific material, amount of strain, size and number of QWs, and device geometry, as well as the quality of crystal growth. Calculations show that compressive strain provides the lowest transparency current density, but tensile strain provides the largest gain (at sufficiently high carrier densities), as shown in Fig. 16.²⁷ The lowest threshold lasers, then, will typically be compressively strained. Nonetheless, calculations show that, far enough above the band edge, the differential gain is 4 times higher in tensile strain compared to compressive strain. This results in a smaller linewidth enhancement factor, even if the refractive index changes per carrier density are larger. It has also been found that tensile strain in the active region reduces

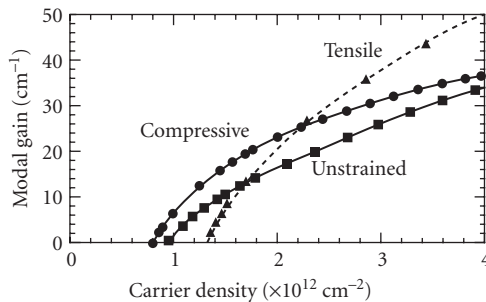


FIGURE 16 Modal gain at 1.55 μm in InGaAs QW lasers calculated as a function of the carrier density per unit area contained in the quantum well. Well widths were determined by specifying wavelength.²⁷

the Auger recombination, decreasing the losses introduced at higher temperatures. This means that T_o can increase with strain, particularly tensile strain. Strained QWs enable performance at 1.55 μm comparable with that of GaAs lasers. Deciding between compressively and tensilely strained QWs will be a matter of desired performance for specific applications.

Threshold current densities under 200 A/cm² have been reported at 1.55 μm ; T_o values on the order of 140 K have been reported—3 times better than bulk lasers. Strained QW lasers have improved modulation properties compared with bulk DH lasers. Because the gain coefficient can be almost double, the relaxation oscillation frequency is expected to be almost 50 percent higher, enhancing the modulation bandwidth and decreasing the relative intensity noise for the same output power. Even the frequency chirp under modulation will be less, because the linewidth enhancement factor is less. The typical laser geometry, operating characteristics, transient response, noise, frequency chirping, and the effects of external optical feedback are all similar in the strained QW lasers to what has been described previously for bulk lasers. Only the experimentally derived numerical parameters will be somewhat different; strained long-wavelength InP-based semiconductor lasers have performance parameters comparable to those of GaAs lasers. One difference is that the polarization of the light emitted from strained lasers may differ from that emitted from bulk lasers. As explained in Sec. 13.3, the gain in bulk semiconductors is independent of polarization, but lasers tend to be polarized in-plane because of higher facet reflectivity for that polarization. The use of QWs causes the gain for the TE polarization to be slightly (~10 percent) higher than for the TM polarization, so lattice-matched QW lasers operate with in-plane polarization. Compressive strain causes the TE polarization to have significantly more gain than the TM polarization (typically 50 to 100 percent more), so these lasers are also polarized in-plane. However, tensile strain severely depresses the TE gain, and these lasers have the potential to operate in TM polarization.

Typical 1.3- and 1.5- μm InP lasers today use from 5 to 15 wells that are grown with internal strain. By providing strain-compensating compressive barriers, there is no net buildup of strain. Typical threshold current densities today are ~1000 A/cm², threshold currents ~10 mA, T_o ~ 50 to 70 K, maximum powers ~40 mW, differential efficiencies ~0.3 W/A, and maximum operating temperatures ~70°C before the maximum power drops by 50 percent. There are trade-offs on all these parameters; some can be made better at the expense of some of the others.

13.7 DISTRIBUTED FEEDBACK AND DISTRIBUTED BRAGG REFLECTOR LASERS

Rather than cleaved facets, some lasers use distributed reflection from corrugated waveguide surfaces. Each groove provides some slight reflectivity, which adds up coherently along the waveguide at the wavelength given by the corrugation. This has two advantages. First, it defines the wavelength (by choice of grating spacing) and can be used to fabricate single-mode lasers. Second, it is an in-plane technology (no cleaves) and is therefore compatible with monolithic integration with modulators and/or other devices.

Distributed Bragg Reflector Lasers

The *distributed Bragg reflector* (DBR) laser replaces one or both laser facet reflectors with a waveguide diffraction grating located outside the active region, as shown in Fig. 17.²⁸ The reflectivity of a Bragg mirror is the square of the reflection coefficient (given here for the assumption of lossless mirrors):²⁹

$$r = \frac{\kappa}{\delta - iS \coth(SL)} \quad (34)$$

where κ is the *coupling coefficient* due to the corrugation (which is real for corrugations that modify the effective refractive index in the waveguide, but would be imaginary for periodic modulations in the gain

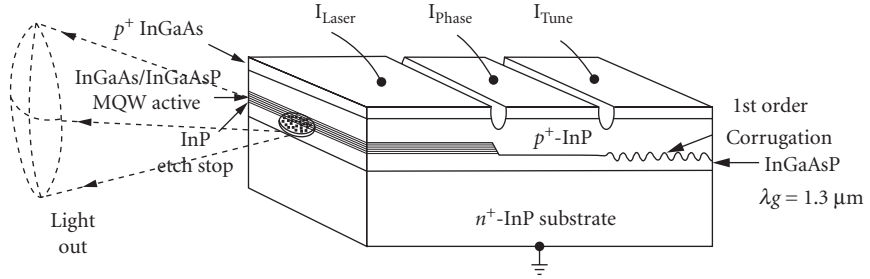


FIGURE 17 Schematic for DBR laser configuration in a geometry that includes a phase portion for phase tuning and a tunable DBR grating. Fixed-wavelength DBR lasers do not require this tuning region. Designed for 1.55- μm output, light is waveguided in the transparent layer below the MQW that has a bandgap at a wavelength of 1.3 μm . The guided wave reflects from the rear grating, sees gain in the MQW active region, and is partially emitted and partially reflected from the cleaved front facet. Fully planar integration is possible if the front cleave is replaced by another DBR grating.²⁸

and could, indeed, be complex). Also, δ is a *detuning parameter* that measures the offset of the optical wavelength λ from the grating periodicity Λ . When the grating is used in the m th order, the detuning δ relates to the optical wavelength λ by

$$\delta = \frac{2\pi n_g}{\lambda} - \frac{m\pi}{\Lambda} \quad (35)$$

where n_g is the effective group refractive index of the waveguide mode, and m is any integer. Also, S is given by $S^2 = \kappa^2 - \delta^2$. When detuning $\delta > \kappa$, Eq. (34) is still valid, and is analytically evaluated with S as imaginary.

The Bragg mirror has its maximum reflectivity on resonance when $\delta \rightarrow 0$ and the wavelength on resonance λ_m is determined by the m th order of the grating spacing through $\Lambda = m\lambda_m/2n_g$. The reflection coefficient on resonance is $r_{\max} = -i \tanh(KL)$, and the Bragg reflectivity on resonance is

$$R_{\max} = \tanh^2(KL) \quad (36)$$

where K is the coupling per unit length, $K = |\kappa|$, and is larger for deeper corrugations or when the refractive index difference between the waveguide and the cladding is larger. The reflectivity falls off as the wavelength moves away from resonance and the detuning increases. Typical resonant reflectivities are $R_{\max} = 0.93$ for $KL = 2$ and $R_{\max} = 0.9987$ for $KL = 4$.

A convenient formula for the shape of the reflectivity as a function of detuning near resonance is given by the reflection loss:

$$1 - R = \frac{1 - (\delta L)^2 / (\kappa L)^2}{\cosh^2(SL) - (\delta L)^2 / (\kappa L)^2} \quad (37)$$

The reflection loss doubles when off resonance by an amount $\delta L = 1.6$ for $\kappa L = 2$ and when $\delta L = 2$ for $\kappa L = 4$. The wavelength half-bandwidth is related to the detuning δL by $\Delta\lambda = \lambda_g^2 (\delta L / 2\pi) / n_g L$. The calculated FWHM of the resonance is 0.6 nm (when $L = 500 \mu\text{m}$, $\lambda = 1.3 \mu\text{m}$) for a 99.9 percent reflective mirror. The wavelength of this narrow resonance is fixable by choosing the grating spacing and can be modulated by varying the refractive index (with, for example, carrier injection), properties that make the DBR laser very favorable for use in optical communication systems.

The characteristics of Fabry-Perot lasers described previously still hold for DBR lasers, except that the narrow resonance can ensure that these lasers are single-mode, even at high excitation levels.

Distributed Feedback Lasers

When the corrugation is put directly on the active region or its cladding, this is called *distributed feedback* (DFB). One typical BH example is shown in Fig. 18, with a buried grating waveguide that was grown on top of a grating-etched substrate, which forms the separate confinement heterostructure laser. The cross-hatched region contains the MQW active layer. A stripe mesa was etched and regrown to bury the heterostructure. Reflection from the cleaved facets must be suppressed by means of an antireflection coating. As before, the grating spacing is chosen such that, for a desired wavelength near λ_o , $\Lambda = m\lambda_o/2n_{ge}$, where n_{ge} is the effective group refractive index of the laser mode inside its waveguiding active region, and m is any integer. A laser operating under the action of this grating has feedback that is distributed throughout the laser gain medium. In this case, Eq. (34) is generalized to allow for the gain: $\delta = \delta_o + ig_L$, where g_L is the laser gain and $\delta_o = 2\pi n_{ge}/\lambda - 2\pi n_{ge}/\lambda_o$. Equations (34) to (37) remain valid, understanding that now δ is complex.

The laser oscillation condition requires that after a round-trip inside the laser cavity, a wave must have the same phase that it started out with, so that successive reflections add in phase. Thus, the phase of the product of the complex reflection coefficients (which now include gain) must be an integral number of 2π . This forces r^2 to be a positive real number. So, laser oscillation requires $r^2 > 0$.

On resonance $\delta_o = 0$ and $S_o^2 = \kappa^2 + g_L^2$, so that S_o is pure real for simple corrugations (κ real). Since the denominator in Eq. (34) is now pure imaginary, r^2 is negative and the round-trip laser oscillation condition cannot be met. Thus, there is no on-resonance solution to a simple DFB laser with a corrugated waveguide and/or a periodic refractive index. The DFB laser oscillates slightly off-resonance.

DFB Threshold We look for an off-resonance solution to the DFB. A laser requires sufficient gain that the reflection coefficient becomes infinite. That is, $\delta_{th} = iS_{th} \coth(S_{th}L)$, where $S_{th}^2 = \kappa^2 - \delta_{th}^2$. By simple algebraic manipulation, the expression for δ_{th} can be inverted. For large gain, $\delta_{th}^2 \gg \kappa^2$, so that $S_{th} = i\delta_{th} = i\delta_o - g_L$, and the inverted equation becomes³⁰

$$\exp(2S_{th}) \frac{4(g_L - i\delta_o)^2}{K^2} = -1 \quad (38)$$

This is a complex eigenvalue equation that has both a real and an imaginary part, which give both the detuning δ_o and the required gain g_L .

The required laser gain is found from the magnitude of Eq. (38) through

$$2 \tan^{-1} \left(\frac{\delta_o}{g_L} \right) - 2\delta_o L + \delta_o L \frac{K^2}{g_L^2 + \delta_o^2} = (2m + 1)\pi \quad (39)$$

There is a series of solutions, depending on the value of m .

For the largest possible gains, $\delta_o L = -(m + 1/2)\pi$. There are two solutions, $m = -1$ and $m = 0$, giving $\delta_o L = -\pi/2$ and $\delta_o L = +\pi/2$. These are two modes equally spaced around the Bragg resonance.

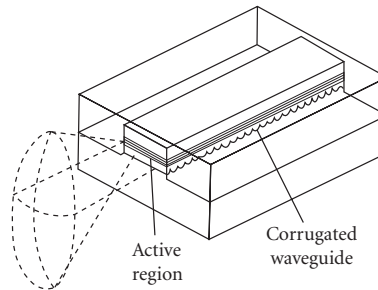


FIGURE 18 Geometry for a buried grating heterostructure DFB laser.

Converting to wavelength units, the mode detuning becomes $\delta_o L = -2\pi n_g L (\delta\lambda/\lambda^2)$, where $\delta\lambda$ is the deviation from the Bragg wavelength. Considering $\delta_o L = \pi/2$, for $L = 500 \mu\text{m}$, $n_g = 3.5$, and $\lambda = 1.55 \mu\text{m}$, this corresponds to $\delta\lambda = 0.34 \text{ nm}$. The mode spacing is twice this, or 0.7 nm .

The required laser gain is found from the magnitude of Eq. (38) through

$$\frac{K^2}{4} = (g_L^2 L^2 + \delta_o^2 L^2) \exp(-2g_L L) \quad (40)$$

For the required detuning $\delta_o L = -\pi/2$, the gain can be found by plotting Eq. (40) as a function of gain g_L , which gives $K(g_L)$, which can be inverted to give $g_L(K)$.

These results show that there is a symmetry around $\delta_o = 0$, so that there will tend to be *two* modes, equally spaced around λ_o . Such a multimode laser is not useful for communication systems so something must be done about this. The first reality is that there are usually cleaved facets, at least at the output end of the DFB laser. This changes the analysis from that given here, requiring additional Fresnel reflection to be added to the analysis. The additional reflection will usually favor one mode over the other, and the DFB will end up as a single-mode. However, there is very little control over the exact positioning of these additional cleaved facets with respect to the grating, and this has not proven to be a reliable way to achieve single-mode operation. The most common solution to this multimode problem is to use a *quarter-wavelength-shifted grating*, as shown in Fig. 19. Midway along the grating, the phase is made to change by $\pi/2$ and the two-mode degeneracy is lifted. This is the way DFB lasers are made today.

Quarter-Wavelength-Shifted Grating Introducing an additional phase shift of π to the round-trip optical wave enables an on-resonance DFB laser. This is done by interjecting an additional phase region of length $\Lambda/2$, or $\lambda/4n_g$, as shown in Fig. 19. This provides an additional $\pi/2$ phase each way, so that the high-gain oscillation condition becomes $\delta_o L = -m\pi$. Now there is a unique solution at $m = 0$, given by Eq. (40) with $\delta_o = 0$:

$$KL = g_L L \exp(-g_L L) \quad (41)$$

Given a value for the DFB coupling parameter KL , the gain can be calculated. Alternatively, the gain can be varied, and the coupling coefficient that must be used with that gain can be calculated. It can be seen that if there are internal losses α_i , the laser must have sufficient gain to overcome them as well: $g_L \rightarrow g_L + \alpha_i$.

Quarter-wavelength-shifted DFB lasers are commonly used in telecommunication applications. DFB corrugations can be placed in a variety of ways with respect to the active layer. Most common is to place the corrugations laterally on either side of the active region, where the evanescent wave of the guided mode experiences sufficient distributed feedback for threshold to be achieved. Alternative methods place the corrugations on a thin cladding above the active layer. Because the process of corrugation may introduce defects, it is traditional to avoid corrugating the active layer directly. Once a

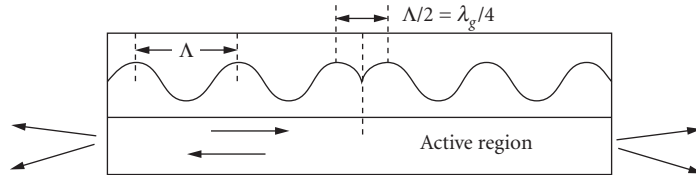


FIGURE 19 Side view of a quarter-wavelength-shifted grating, etched into a separate confinement waveguide above the active laser region. Light with wavelength in the medium λ_g sees a $\pi/4$ phase shift, resulting in a single-mode DFB laser operating on line-center.

DFB laser has been properly designed, it will be single-mode at essentially all power levels and under all modulation conditions. Then the single-mode laser characteristics described in the early part of this chapter will be well satisfied. However, it is crucial to avoid reflections from fibers back into the laser, because instabilities may arise, and the output may cease to be single-mode.

A different technique that is sometimes used is to spatially modulate the gain. This renders κ complex and enables an on-resonance solution for the DFB laser, since S will then be complex on-resonance. Corrugation directly on the active region makes this possible, but care must be taken to avoid introducing centers for nonradiative recombination.

More than 35 years of research and development have gone into semiconductor lasers for telecommunications. Today it appears that the optimal sources for these applications are strained QW distributed feedback lasers operating at 1.3 or 1.55 μm wavelength.

13.8 TUNABLE LASERS

The motivation to use tunable lasers in optical communication systems comes from wavelength division multiplexing (WDM), in which a number of independent signals are transmitted simultaneously, each at a different wavelength. The first WDM systems used wavelengths far apart (so-called *coarse* WDM) and settled on a standard of 20-nm wavelength spacing (~ 2500 GHz). But interest grew rapidly toward *dense* wavelength division multiplication (DWDM), with much closer wavelength spacing. The International Telecommunications Union (ITU) defined a standard for a grid of optical frequencies, each referring to a reference frequency which has been fixed at 193.10 THz (1552.5 nm). The grid separation can be as narrow as 12.5 GHz or as wide as 100 GHz. Tuning range can extend across the conventional erbium amplifier window C band (1530 to 1565 nm) and ideally extends to either side. Tuning to longer wavelengths will extend through the L band out to 1625 nm or even farther through the ultra-long U band to 1675 nm. On the short wavelength side, the S band goes to 1460 nm, after which the extended E band transmits only in fibers without water absorption. The original O band lies between 1260 and 1360 nm. An ideal tuning range would extend throughout all these optical fiber transmission bands.

Two kinds of tunable lasers have application to fiber optical communications. The first is a laser with a set of fixed wavelengths at the ITU frequencies that can be tuned to any wavelength on the grid and operated permanently at that frequency. This approach may be cost-effective because network operators do not have to stock-pile lasers at each of the ITU frequencies; they can purchase a few identical tunable lasers and set them to the required frequency when replacements are needed. The other kind of tunable laser is agile in frequency; it can be tuned in real time to whatever frequency is open to use within the system. These agile tunable lasers offer the greatest systems potential, perhaps someday enabling wavelength switching even at high-speed packet rates. These agile lasers also tend to be more expensive, and at the present time somewhat less reliable.

Most tunable laser diodes in fiber optics communications can be divided into three categories: an array of different frequency lasers with a moveable external mirror, a tunable external cavity laser (ECL), and a monolithic tunable laser. Each of these will be discussed in the following sections.

Array with External Mirror

Many applications do not require rapid change in frequency, such as for replacement lasers. In this case it is sufficient to have an array of DFB lasers, each operating at a different frequency on the ITU grid, and to move an external mirror to align the desired laser to the output fiber. Fujitsu, NTT, Furukawa, and Santur have all presented this approach at various conferences. A typical system might contain a collimating lens, a tilting mirror [often a MEMS (micro-electromechanical structure)] and a lens that focuses into the fiber. The challenge is to develop a miniature device that is cost-effective. A MEMS mirror may be fast enough to enable agile wavelength switching at the circuit level; speeds are typically milliseconds but may advance to a few tenths of a millisecond.

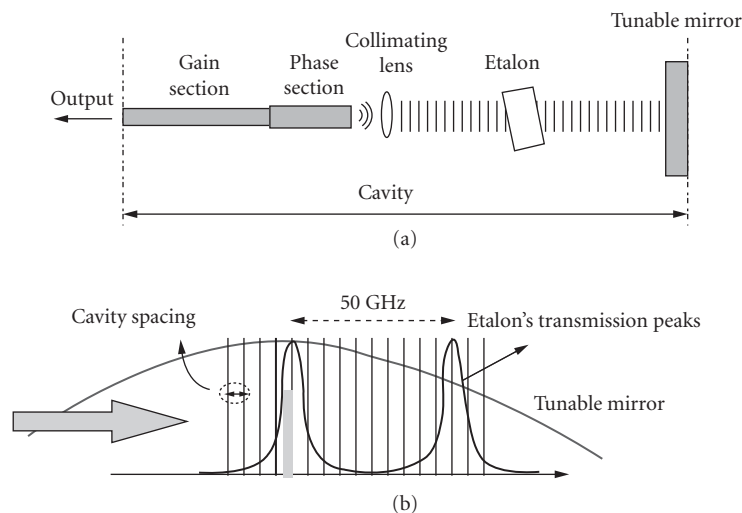


FIGURE 20 Tunable external cavity laser: (a) geometry, showing etalon set at the ITU spacing of 50 GHz and a tunable mirror and (b) spectrum of mirror reflectivity. Spectral maximum can be tuned to pick the desired etalon transmission maximum. The cavity spacing must match the two other maxima, which is done with the phase section. (Adapted from Ref. 31.)

External Cavity Laser

Tunable lasers may consist of a single laser diode in a tiny, wavelength-tunable external cavity, as shown in Fig 20a. A tunable single-mode is achieved by inserting two etalons inside the cavity. One is set at the 50-GHz spacing of the ITU standard; the other provides tuning across the etalon grid, as shown in Fig. 20b. A tunable phase section is also required to ensure that the mode selected by the overall laser cavity length adds constructively to the mode selection from the etalon and mirror. The mechanism for tuning has varied from a liquid crystal mirror,³¹ to thermal tuning of two Fabry-Perot filters within the cavity, reported by researchers at Intel. Pirelli is another company that uses an ECL; they have not reported their method of tuning, but previous work from their laboratory suggests that a polished fiber coupler with variable core separation could be used to tune a laser wavelength.

An alternative tunable filter is acousto-optic, fabricated in lithium niobate, which has been shown to have a tuning range of 132 nm, covering the entire L, C, and S bands.³² Stable oscillation was achieved for 167 channels, each separated by 100 GHz, although there is no evidence that this has become a commercial device.

The speed of tuning external cavity lasers to date is on the order of milliseconds, perhaps fast enough to enable circuit switching to different wavelengths; additional research is underway to achieve faster switching times.

Monolithic Tunable Lasers

Integration of all elements on one substrate offers the greatest potential for compact, inexpensive devices that can switch rapidly from one wavelength to another. The aim is to tune across the entire ITU frequency grid, as far as the laser gain spectrum will allow. All monolithic tunable lasers reported to date involve some sort of grating vernier so that a small amount of tuning can result in a large spectral shift. When the periods of two reflection spectra are slightly mismatched, lasing will occur at that pair of reflectivity maxima that are aligned. Inducing a small index change in one mirror

relative to the other causes adjacent reflectivity maxima to come into alignment, shifting the lasing wavelength a large amount for a small index change. However, to achieve continuous tuning between ITU grid frequencies, the phase within the two-mirror cavity must be adjusted so that its mode also matches the chosen ITU frequency.

The refractive index in the gratings is usually changed by current injection, since changing the free-carrier density in a semiconductor alters its refractive index. Free-carrier injection also introduces loss, which is made up for by the semiconductor optical amplifier (SOA). In principle this modulation speed can be as fast as carriers can be injected and removed. Thermal tuning of the refractive index is an important alternative, because of the low thermal conductivity of InP-based materials. Local resistive heating is enough to create the 0.2 percent change in refractive index needed for effective tuning. The electro-optic effect under reverse bias is not used at present, because the effect does not create large-enough index change at moderate voltages.

In order for the grating to be retroreflective, its periodicity must be half the wavelength of light in the medium (or an odd integral of that); this spacing was discussed in the section on DBR gratings (Sec. 13.7). The other requirement is that there be a periodicity Λ at a scale that provides a comb of possible frequencies at the ITU-T grid (like the etalon in the ECL). This is done in monolithic grating devices by installing an overall periodicity to the grating at the grid spacing. One way to do this is with a *sampled grating* (SG), as shown in Fig. 21a. Only samples of the grating are provided, periodically at frequency Λ ; this is usually done by removing periodic regions of a continuous grating. A laser with DBR mirrors containing sampled gratings is called a *sampled grating DBR* (SG-DBR) laser. If the sampling is abrupt, the reflectance spectrum of the overall comb of frequencies will have the conventional sinc function. The comb reflectance spectrum can be made flat by adding a semiconductor optical amplifier in-line with the DBR laser, or inserting an electroabsorption modulator (EAM) (which will be described in Sec. 13.12), or both, as shown in Fig. 21b.³³ The EAM can be used to modulate the laser output, rather than using direct modulation of the laser. JDS Uniphase tunable diode lasers apparently have this geometry.

An alternative approach to achieving a flat spectrum over the tuning range has been to divide the grating into identical elements, each Λ long, and each containing its own structure.³⁴ This concept has been titled the *superstructure grating* (SSG). The periodicity Λ provides the ITU grid frequencies. When the structure of each element is a phase grating that is chirped quadratically (Fig. 22a), the overall reflectance spectrum is roughly constant and the number of frequencies can be very large. Without the quadratic phase grating structure, the amplitude of the overall reflectance spectrum would have the typical sinc function. Figure 22b shows how the desired quadratic phase shift can be achieved with uneven spacing of the grating teeth, and Fig. 22c shows the resulting measured flat reflectance spectrum.³⁴

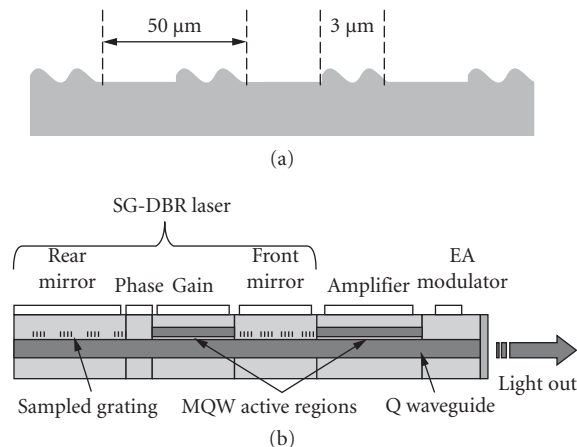


FIGURE 21 Sampled grating: (a) the geometry and (b) sampled grating DBR laser integrated with SOA and EAM.³³

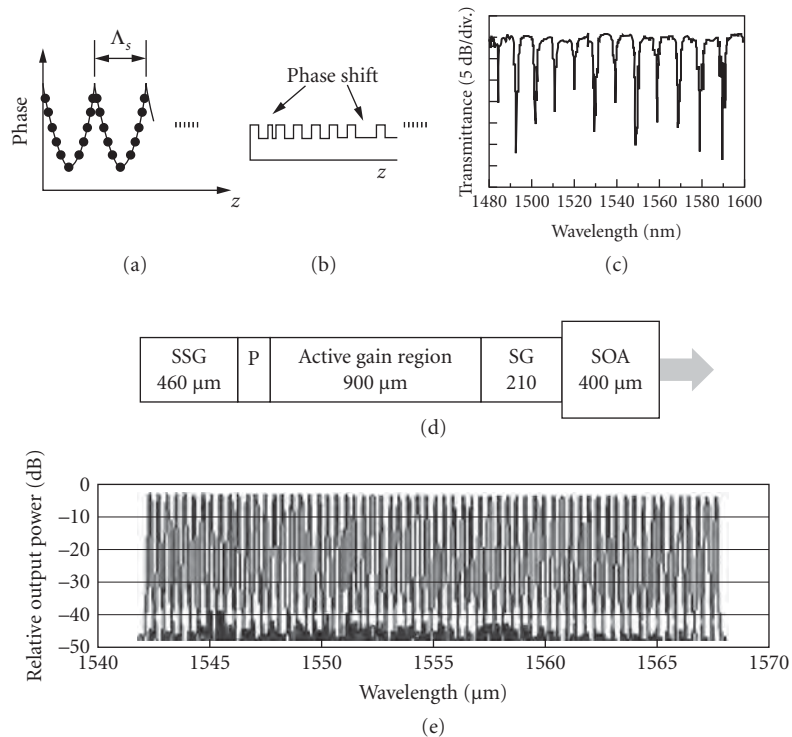


FIGURE 22 Sampled superstructure grating: One period of (a) quadratic phase shift and (b) quadratic phase superstructure; (c) SSG reflectance spectrum;³⁴ (d) geometry for wide-bandwidth integrated tunable laser; and (e) measured tuning output spectrum.³⁵

A relatively new tunable laser diode that can be tuned for DWDM throughout most of the C band contains a front SG-DBR and a rear SSG-DBR.³⁵ Figure 22d shows the design of this device, with lengths given in micrometers (P represents a phase shift region 150 μm long). At the bottom is the spectral output of such a monolithic laser, tuned successively across each wavelength of the ITU grid. A short, low-reflectivity front mirror enables high output power, while keeping the minimum reflectivity that enables wavelength selection based on the Vernier mechanism. A long SSG-DBR was adopted as the rear mirror along with phase control inside the laser cavity, to provide a uniform reflectivity spectrum envelope with a high peak reflectivity (greater than 90%). This monolithic tunable laser includes an integrated SOA for high output power.

A laser tunable for coarse WDM uses a rear reflector comprising a number of equal lengths of uniform phase grating separated by π -phase shifts and a single continuously chirped grating at the front.³⁶ By the correct choice of the number and positions of the phase shifts, the response of the grating can be tailored to produce flat comb of reflection peaks throughout the gain bandwidth. The phase grating was designed to provide seven reflection peaks, each with a 6.8-nm spacing. Over the 300- μm -long front grating is placed a series of short contacts for injecting current into different parts of the grating in a controlled way (Fig. 23). This enables the enhancement of the reflection at a desired wavelength simply by injecting current into a localized part of the chirped grating. The chirp rate was chosen to yield a total reflector bandwidth of around 70 nm. As with other devices, the monolithic chip includes a phase control region and a SOA. As seen in Fig. 23, the waveguide path through the SOA is curved to avoid retroreflection back into the laser cavity—a standard technique.

A different approach is to use a multimode interferometer (MMI) as a Y branch (which will be explained in Sec. 13.11 on modulators).³⁷ This separates backward-going light into two branches, each

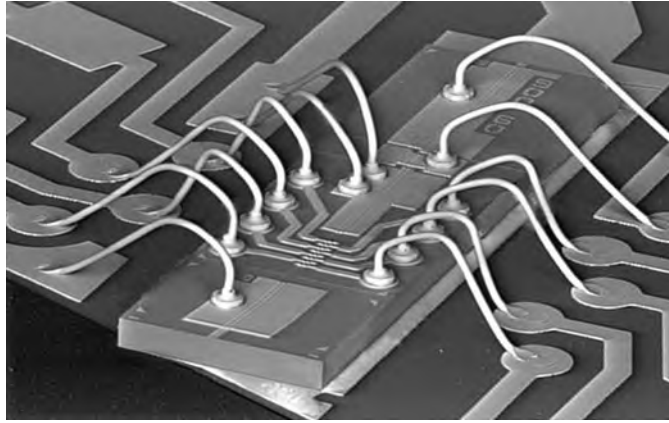


FIGURE 23 Scanning electron microscope image of a monolithically integrated tunable laser for coarse WDM.³⁶

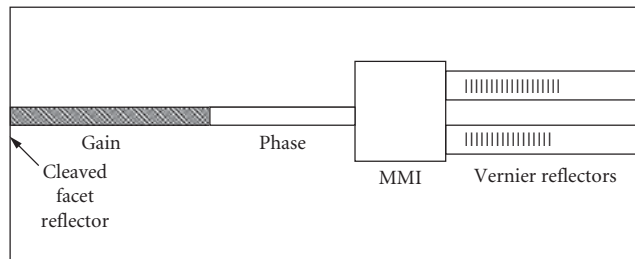


FIGURE 24 Conceptual design of tunable modulated grating DBR laser, including multimode interferometric beam splitter as a Vernier.

reflecting from a grating of a different periodicity, as shown in Fig. 24. The Vernier effect extends the tuning range through parallel coupling of these two *modulated grating* (MG) reflectors with slightly different periods; both reflections are combined at the MMI. The aggregate reflection seen from the input port of the MMI coupler gives a large reflection only when the reflectivity peaks of both gratings align. A large tuning range (40 nm) is obtained for relatively small tuning of a single reflector (by an amount equal to the difference in peak separation). A phase section aligns a longitudinal cavity mode with the overlapping reflectivity peaks. Tunable high-speed direct modulation at 10 Gb/s has been demonstrated with low injection current, with low power consumption and little heat dissipation.

Commercial tunable laser diodes for optical communications are still in their infancy; it is still unclear which of these technologies will be optimum for practical systems. The ultimate question is whether for any given application it is worth the added cost and complexity of tunability.

13.9 LIGHT-EMITTING DIODES

Sources for low-cost fiber communication systems, such as used for communicating data, have traditionally used light-emitting diodes (LEDs). These may be *edge-emitting LEDs* (E-LEDs), which resemble laser diodes, or *surface-emitting LEDs* (S-LEDs), which emit light from the surface of the diode and can be butt-coupled to multimode fibers. The S-LEDs resemble today's VCSELs, discussed

in the following section. The LED can be considered a laser diode operated below threshold, but it must be specially designed to maximize its output.

When a pn junction is forward biased, electrons are injected from the n region and holes are injected from the p region into the active region. When free electrons and free holes coexist with comparable momentum, they will combine and may emit photons of energy near that of the bandgap, resulting in an LED. The process is called *injection-* or *electroluminescence*, since injected carriers recombine and emit light by spontaneous emission. A semiconductor laser diode below threshold acts as an LED. Indeed, a laser diode without mirrors *is* an LED. Because LEDs have no threshold, they usually are not as critical to operate and are often much less expensive because they do not require the fabrication step to provide optical feedback (in the form of cleaved facets or DFB). Because the LED operates by spontaneous emission, it is an incoherent light source, typically emitted from a larger aperture (out the top surface) with a wider far-field angle and a much wider wavelength range (30 to 50 nm). In addition, LEDs are slower to modulate than laser diodes because stimulated emission does not remove carriers. Nonetheless, they can be excellent sources for inexpensive multimode fiber communication systems, as they use simpler drive circuitry. They are longer lived, exhibit more linear input-output characteristics, are less temperature sensitive, and are essentially noise-free electrical-to-optical converters. The disadvantages are lower power output, smaller modulation bandwidths, and pulse distortion in fiber systems because of the wide wavelength band emitted. Some general characteristics of LEDs are discussed in Chap. 17 in Vol. II of this *Handbook*.

In fiber communication systems, LEDs are used for low-cost, high-reliability sources typically operating with graded index multimode fibers (core diameters approximately 62 μm) at data rates up to 622 Mb/s. For short fiber lengths they may be used with step-index plastic fibers. The emission wavelength will be at the bandgap of the active region in the LED; different alloys and materials have different bandgaps. For medium-range distances up to ~ 10 km (limited by modal dispersion), LEDs of $\text{In}_x\text{Ga}_y\text{As}_{1-x-y}\text{P}_{1-y}$ grown on InP and operating at $\lambda = 1.3$ μm offer low-cost, high-reliability transmitters. For short-distance systems, up to 2 km, GaAs LEDs operating near $\lambda = 850$ nm are used, because they have the lowest cost, both to fabricate and to operate, and the least temperature dependence. The link length is limited to ~ 2 km because of chromatic dispersion in the fiber and the finite linewidth of the LED. For lower data rates (a few megabits per second) and short distances (a few tens of meters), very inexpensive systems consisting of red-emitting LEDs with $\text{Al}_x\text{Ga}_{1-x}\text{As}$ or GaInP_{1-y} active regions emitting at 650 nm can be used with plastic fibers and standard silicon detectors. The 650-nm wavelength is a window in the absorption in acrylic plastic fiber, where the loss is ~ 0.3 dB/m; a number of companies now offer 650-nm LEDs.

A typical GaAs LED heterostructure is shown in Fig. 25, with (a) showing the device geometry and (b) showing the heterostructure bandgap under forward bias. The forward-biased pn junction injects electrons and holes into the narrowband GaAs active region. The $\text{Al}_x\text{Ga}_{1-x}\text{As}$ cladding layers confine the carriers in the active region. High-speed operation requires high levels of injection (and/or doping) so that the spontaneous recombination rate of electrons and holes is very high. This means that the active region should be very thin. However, nonradiative recombination increases at high carrier concentrations, so there is a trade-off between internal quantum efficiency and speed. Under some conditions, LED performance is improved by using QWs or strained layers. The improvement is not as marked as with lasers, however.

Spontaneous emission causes light to be emitted in all directions inside the active layer, with an internal quantum efficiency that may approach 100 percent in these direct band semiconductors. However, only the light that gets out of the LED and into the fiber is useful in a communication system, as illustrated in Fig. 25a. The challenge, then, is to collect as much light as possible into the fiber end. The simplest approach is to butt-couple a multimode fiber to the S-LED surface as shown. Light emitted at too large an angle will not be guided in the fiber core, or will miss the core altogether. Light from the edge-emitting E-LED (in the geometry of Fig. 1, with antireflection-coated cleaved facets) is more directional and can be focused into a single-mode fiber. Its inexpensive fabrication and integration process makes the S-LED common for inexpensive data communication. The E-LED has a niche in its ability to couple with reasonable efficiency into single-mode fibers. Both LED types can be modulated at bit rates up to 622 Mbps, an asynchronous transfer mode (ATM) standard, but many commercial LEDs have considerably smaller modulation bandwidths.

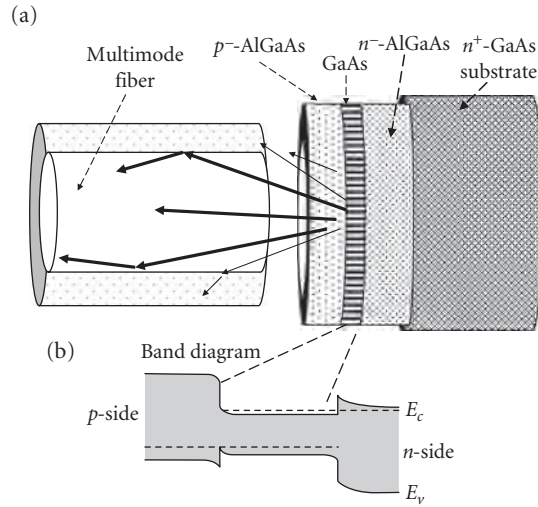


FIGURE 25 GaAs light-emitting diode (LED) structure: (a) cross-section of surface-emitting LED aligned to a multimode fiber, showing rays that are guided by the fiber core and rays that cannot be captured by the fiber and (b) conduction band E_c and valence band E_v as a function of distance through the LED.

Surface-Emitting LEDs

The coupling efficiency of an S-LED butt-coupled to a multimode fiber (shown in Fig. 26a) is typically small, unless methods are employed to optimize it. Because light is spontaneously emitted in all internal directions, only half is emitted toward the top surface. In addition, light emitted at too great an angle to the surface normal is totally internally reflected back down and is lost (although it may be reabsorbed, creating more electron-hole pairs). The critical angle for total internal reflection between the semiconductor of refractive index n_s and the output medium (air or plastic encapsulant) of refractive index n_o is given by $\sin \theta_c = n_o/n_s$. The refractive index of GaAs is $n_s \sim 3.3$, when the output medium is air, the critical angle $\theta_c \sim 18^\circ$. Because this angle is so small, less than 2 percent of the total internal spontaneous emission can come out the top surface, at any angle. A butt-coupled fiber can accept only spontaneous emission at those external angles that are smaller than its numerical aperture. For a typical fiber $NA \approx 0.25$, this corresponds to an external angle (in air) of 14° , which corresponds to only 4.4° inside the GaAs. This means

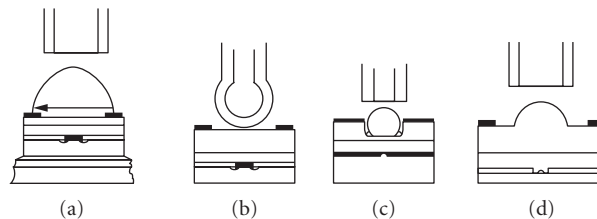


FIGURE 26 Typical geometries for coupling from LEDs into fibers: (a) hemispherical lens attached with encapsulating plastic; (b) lensed fiber tip; (c) microlens aligned through use of an etched well; and (d) spherical semiconductor surface formed on the substrate side of the LED.

that the cone of spontaneous emission that can be accepted by the fiber in this simple geometry is only ~ 0.2 percent of the entire spontaneous emission! Fresnel reflection losses make this number even smaller.

For InP-based LEDs, operating in the 1.3- or 1.55- μm wavelength region, the substrate is transparent and LED light can be emitted out the substrate. In this geometry the top contact to the p -type material no longer need be a ring; it can be solid and reflective, so light emitted backward can be reflected toward the substrate, increasing the efficiency by a factor of 2.

The coupling efficiency can be increased in a variety of other ways, as shown in Fig. 26. The LED source is incoherent, a lambertian emitter, and follows the law of imaging optics: A lens can be used to reduce the angle of divergence of LED light, but this will enlarge the apparent source. The image of the LED source must be smaller than the fiber into which it is to be coupled. Unlike a laser, the LED has no modal interference and the output of a well-designed LED has a smooth intensity distribution that lends itself to imaging.

The LED can be encapsulated in materials such as plastic or epoxy, with direct attachment to a focusing lens (Fig. 26a). The output cone angle will depend on the design of this encapsulating lens. Even with a parabolic surface, the finite size of the emitting aperture and resulting aberrations will be the limiting consideration. In general, the user must know both the area of the emitting aperture and the angular divergence in order to optimize coupling efficiency into a fiber. Typical commercially available LEDs at $\lambda = 850$ nm for fiber optic applications have external half-angles of $\sim 25^\circ$ without a lens and $\sim 10^\circ$ with a lens, suitable for butt-coupling to multimode fiber.

Improvement can also be achieved by lensing the pigtailed fiber to increase its acceptance angle (Fig. 26b); this example shows the light emitted through and out the substrate. Another alternative is to place a microlens between the LED and the fiber (Fig. 26c), possibly within an etched well. Substrate-side emission enables a very effective geometry for capturing light by means of a domed surface fabricated directly on the substrate, as shown in in Fig. 26d. Because the refractive index of encapsulating plastic is < 1.5 , compared to 3.3 of the semiconductor, only a dome etched within the semiconductor can entirely eliminate total internal reflection. Integrated semiconductor domes require advanced semiconductor fabrication technology, but have proven effective. In GaAs diodes the substrate is absorptive, but etching a well in the substrate and inserting a fiber can serve to collect backside emission. For any of these geometries, improvement in efficiency by as much as a factor of 2 can be obtained if a mirror is provided to reflect backward-emitted light. This mirror can be either metal or a dielectric stack deposited at the air-semiconductor interface, or even a DBR mirror grown within the semiconductor structure.

Current must be confined to the surface area of emission, which is typically 25 to 75 μm in diameter. This is done by constricting the flow of injection current by mesa etching or by using an oxide-defined electrode. Regrowth using n pn blocking layers or semi-insulating material in the surrounding areas (as in lasers) has the advantage of reducing thermal heating. Surface-emitting LEDs require that light be emitted out of the surface in a gaussian-like pattern; it must not be obscured by the contacting electrode. Typically, a highly conductive cap layer brings the current in from a ring electrode. In InP-based devices, when light is collected out of the substrate side, a solid top electrode can be reflective and electrical contact may be made to the substrate surrounding an optical output aperture.

A typical S-LED at 1310-nm wavelength might couple 35 μW of light into a pigtailed multimode fiber of 62.5 μm diameter, if driven with 62-mA input current at ~ 1 -V forward bias, for an efficiency of less than 0.06 percent. The spectral width is 160 nm, and the rise or fall time of 3 ns means 200-MHz modulation capability. By contrast, a similar S-LED at 850-nm wavelength under similar drive conditions has comparable power at half the modulation speed and a fifth the bandwidth.³⁸

Improved performance has been obtained by sandwiching the active layer between DBR mirrors to form a resonant cavity (RC-LED). This reduces the spontaneous emission linewidth, thereby increasing the modulation bandwidth that can be transmitted through dispersive fibers. The RC-LEDs look very much like VCSELs operated below threshold (see the following section). The resonant cavity promotes emission into resonances supported by the cavity while suppressing off-resonance emission. The cavity narrows the emission angles and increases the external quantum efficiency to greater than 20 percent. By controlling reflectivities and drive current, the output can be tailored anywhere from a wide spectrum LED to a narrow spectrum laser. Below threshold the RC-LED will have the operating characteristics of a spontaneous LED, as described in this section. Above threshold, the resonant structure becomes a VCSEL. The more exacting fabrication of these devices increases their cost, however.

A typical GaAs LED has a spectral width of 50 nm, while an RC-LED has a spectral width of 6.7 nm and 5 times the output power when coupled into a multimode fiber. The narrower spectrum also means much less pulse-broadening in fiber communication systems. The RC-LED, with a spot size of 20 μm , will have an output power typically a quarter that of a VCSEL laser. In comparison, a typical VCSEL has a spot size of 8 μm and a spectral width of 3.2 nm.

RC-LEDs have enhanced modulation bandwidth compared to conventional LEDs, due to a higher carrier density for a given current, which leads to a reduction in the spontaneous lifetime. However, the RC-LEDs bandwidth is current dependent. The fastest modulation speed is obtained by lasers, of course, but they have a highly nonlinear light-current relationship, unlike LEDs.

Edge-Emitting LEDs

Edge-emitting LEDs (E-LEDs or EELEDs) have a geometry that is similar to that of a conventional LD (Fig. 1), but without a feedback cavity. That is, light travels back and forth in the plane of the active region of an E-LED and is emitted out of one antireflection-coated cleaved end. As in a laser, the active layer is 0.1 to 0.2 μm thick. Because the light in an E-LED is waveguided in the out-of-plane dimension and is lambertian in-plane, the output radiation pattern will be elliptical, with the largest divergence in-plane with full width at half-maximum (FWHM) angle of 120°. The out-of-plane guided direction typically radiates with a 30° half-angle. An elliptical collimating lens will be needed to optimally couple light into a fiber. The efficiency can be doubled by depositing a reflector on the back facet of the E-LED, just as in the case of a laser.

Edge-emitting LEDs can be coupled into fibers with greater efficiency because their source area is smaller than that of a S-LED. However, the alignment and packaging is more cumbersome than with S-LEDs. Typically, E-LEDs can be obtained already pigtailed to fibers. Edge-emitting diodes can be coupled into *single-mode* fiber with modest efficiency. A single-mode fiber pigtailed to an E-LED can typically transmit 30 μW at 150-mA drive at 1 V, for an overall efficiency of 0.04 percent.³⁹ This efficiency is comparable to the emission of surface-emitting lasers into *multimode* fiber (with an area 50 times larger than single-mode fibers). Because of their wide emission wavelength bandwidth, E-LEDs are typically used as low-coherence sources for fiber sensor applications, rather than in communications applications.

Operating Characteristics of LEDs

In an LED, the output optical power P_{opt} is linearly proportional to the drive current I ; the relation defines the output efficiency η :

$$P_{\text{out}} = \frac{\eta h \nu I}{e} \quad (42)$$

This efficiency is affected by the geometry of the LED. The power coupled into a fiber is further reduced by the coupling efficiency between the LED emitter and the fiber, which depends on the location, size, and numerical aperture of the fiber as well as on the spatial distribution of the LED output light and the optics of any intervening lens. The *internal* quantum efficiency (ratio of emitted photons to incident electrons) is usually close to 100 percent.

Figure 27 shows a typical result for power coupled into a graded index multimode fiber as a function of current for various temperatures. The nonlinearity in the light out versus current, which is much less than in a laser diode, nevertheless causes some nonlinearity in the modulation of LEDs. This LED nonlinearity arises from both material properties and device configuration; it may be made worse by ohmic heating at high drive currents. The residual nonlinearity is an important characteristic of any LED used in communication systems. Edge emitters are typically less linear because they operate nearer the amplified spontaneous limit.

The InP-based S-LED shows approximately 10 percent reduction in output power for a 25°C increase in temperature (compared to ~50 percent reduction for a typical laser). Unlike a laser, there is no temperature-dependent threshold. Also, the geometric factors that determine the fraction of light

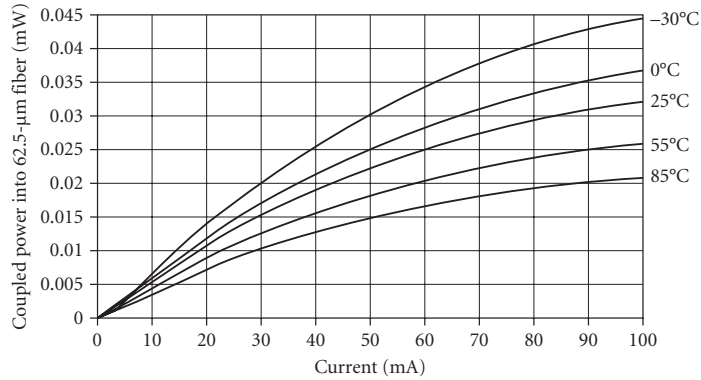


FIGURE 27 Optical power coupled from an *InGaAsP* S-LED into graded index fiber at 1.3- μm wavelength as a function of drive current, for several temperatures.⁴⁰

emitted from the LED are not temperature dependent. Nonetheless, InP-based LEDs have a stronger temperature dependence than GaAs-based LEDs, because of more nonradiative recombination, particularly at the high injection levels required by high-speed LEDs.

The spectrum of the incoherent light emitted from an LED has a roughly gaussian shape with an FWHM around 40 nm in the case of a typical GaAs/AlGaAs LED operating near $\lambda = 0.8 \mu\text{m}$. This bandwidth, along with chromatic dispersion in graded index fibers, limits the distance over which these LEDs can be used in fiber systems. InGaAsP/InP LEDs have wider linewidths (due to alloy scattering, heavy doping, and temperature fluctuations), which depend on the details of their design. As temperature increases, the peak of the spectrum shifts to longer wavelength and the spectrum widens. The variation of the central wavelength with temperature at $\lambda = 1.3 \mu\text{m}$, is approximately 5 meV/°C. However, graded index fibers have negligible chromatic dispersion at this wavelength, so this is not usually a problem; if it is, heat sinking and/or cooling can be provided. Resonant cavity LEDs can provide narrower linewidths.

LEDs do not suffer from the catastrophic optical damage that lasers do, because of their lower optical power densities. However, they do degrade with time; 10^6 to 10^9 hours can be expected. Because degradation processes have an exponential dependence on temperature, LED life can be shortened by operating at excessive temperatures. Experiments with thermally accelerated life testing suggest that the power out P varies with time t as

$$P(t) = P(0) \exp(-qt) \quad (43)$$

where $q = q_0 \exp(-W_a/k_B T)$, with W_a as the activation energy, k_B as Boltzman's constant, and T as temperature. In GaAs LEDs, W_a is 0.6 to 1 eV. Of course this assumes that the LEDs are placed in a proper electrical circuit.

LED light is typically unpolarized, since there is no preferred polarization for spontaneous emission.

Transient Response Most LEDs respond in times faster than 1 μs ; with optimization, they can reach the nanosecond response times needed for optical communication systems. To achieve the 125 Mb/s rate of the fiber distributed data interface (FDDI) standard requires maximum rise and fall times of 3.5 ns; to achieve the 622 Mb/s rate of the asynchronous transfer mode standard, the necessary times drop to 0.7 ns.

The speed of an LED is limited by the recombination time of injected carriers; it does not have the turn-on delay of lasers, nor any relaxation oscillations, but it also does not have the fast decay of stimulated emission. The LED intrinsic frequency response (defined as the ratio of the AC components of the emitted light to the DC value) is⁴¹

$$r(\omega) = (1 + \omega^2 \tau^2)^{-1/2} \quad (44)$$

where τ is the minority carrier lifetime in the injected region. This shows that high-speed LEDs require small minority carrier lifetimes. The square-root dependence comes out of the rate equation solution.

When the active region is doped higher than the injected carrier density, the lifetime τ_L decreases as the background doping density N_o increases:

$$\frac{1}{\tau_L} = BN_o \quad (45)$$

The challenge is to provide high levels of doping without increasing the nonradiative recombination. Typical high-speed response is about 1 ns, although doping with beryllium (or carbon) at levels as high as $7 \times 10^{19} \text{ cm}^{-3}$ has allowed speeds to increase to as much as 0.1 ns, resulting in a cutoff frequency of 1.7 GHz (at the sacrifice of some efficiency).⁴²

When operating in the high-injection regime, the injected carrier density N can be much larger than the doping density, and $1/\tau_H = BN$, where N is created by a current density J such that $N = J\tau/ed$. Combining these two equations

$$\frac{1}{\tau_H} = \left(\frac{BJ}{ed} \right)^{1/2} \quad (46)$$

The recombination time may be reduced by thinning the active region and by increasing the drive current. However, too much injection may lead to thermal problems, which in turn may cause modulation nonlinearity. LEDs with thin active layers operated in the high-injection regime will have the fastest response. Bandwidths in excess of 1 GHz have been achieved in practical LEDs.

Because LEDs have such wide wavelength spectra, frequency chirping is negligible. Because LEDs do not have optical cavities, as do lasers, they will not have modal interference and noise. Also, there will not be strong feedback effects coming from external fiber facets, such as the coherence collapse. Because of their inherent light-current linearity at moderate drive levels, the modulation response of LEDs should be a direct measure of their frequency response. They add no noise to the circuit, and they add distortion only at the highest drive levels.

Drive Circuitry and Packaging The LED is operated under sufficient forward bias to flatten the bands of the *pn* junction. This voltage depends on the bandgap and doping and is typically between 1 and 2 V. The current will be converted directly to light; typically a hundred milliamperes is required to produce a few milliwatts of output into a fiber, with a series resistor used to limit the current.

The LED is modulated by varying the drive current. A typical circuit might apply the signal to the base circuit of a transistor connected in series with the LED and a current-limiting resistor. The variation in current flowing through the LED (and therefore in the light out) is proportional to the input voltage in the base circuit. LEDs are typically mounted on standard headers such as TO-18 or TO-46 cans; SMA and ST connectors are also used. The header is covered by a metal cap with a clear glass top through which light can pass.

13.10 VERTICAL CAVITY SURFACE-EMITTING LASERS

The *vertical cavity surface-emitting laser* (VCSEL) has advantages for low-cost data transmission. The use of a laser means that multi-gigahertz modulation is possible, and the stimulated emission is directional, rather than the isotropic spontaneous emission of LEDs, leading to much higher efficiencies. Because the light is emitted directly from the surface, a single or multimode fiber can be directly butt-coupled with an inexpensive mounting technology, and the coupling efficiency can be very high. VCSELs can be fabricated in linear arrays that can be coupled inexpensively to linear arrays of fibers for parallel fiber interconnects with aggregate bit rates of several gigabits per second,

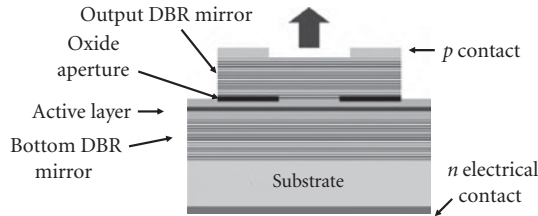


FIGURE 28 Cross-sectional view of an oxidized GaAs VCSEL. The oxidized AlAs layer aperture is shown in black, as is the active layer.⁴³

amortizing the alignment cost over the number of elements in the array. VCSELs lend themselves to two-dimensional arrays as well, which makes them attractive to use with smart modulating pixels. The planar fabrication of VCSELs allows for wafer-scale testing, another cost saving.

Advantages of VCSELs include single longitudinal-mode operation and a circular emission pattern that allows for better coupling into optical fiber without the need for beam-shaping optics. The planar structure also allows much easier fabrication and testing, resulting in higher device yield. VCSELs can also be fabricated in two-dimensional laser arrays with each laser operating at a slightly different frequency, for WDM.

The VCSEL requires mirrors on the top and bottom of the active layer, forming a vertical cavity, as shown in Fig. 28. These lasers use the fact that a DBR (or a multilayer quarter-wavelength dielectric stack) can achieve very high reflectance. Thus, the very short path length through a few QWs (at normal incidence to the plane) is sufficient to reach threshold.

The first VCSELs were based on GaAs: either GaAs active regions that emitted at $\lambda = 850$ nm, or strained InGaAs active regions that emitted at $\lambda = 980$ nm. The former are of greater interest in low-cost communication systems because they are compatible with inexpensive silicon detectors. This section explains VCSEL concepts in terms of GaAs-based devices operating in the 850-nm region. With an active region of $\text{Al}_x\text{In}_y\text{Ga}_{1-x-y}\text{P}$, GaAs-based devices can be fabricated in the 650-nm wavelength regime that is optimum for use with plastic fiber. Highly strained InGaAs QWs enable GaAs-based devices to operate near the 1.3- μm wavelength zero-dispersion regime. The addition of a small amount of nitrogen in the active region extends GaAs-based devices to even longer wavelengths. The development of GaAs-based VCSELs has been relatively straightforward because of the existence of epitaxial GaAs/AlGaAs high-reflectivity DBR mirrors.

VCSELs based on InP technology have been more challenging because the reflectivity of InP-based DBR mirrors is quite low. Also, InP-based QWs have lower gain and significant free carrier absorption, especially in the p layers. Also, the quaternaries have poor thermal conductivity, so the active layer has a relatively high temperature. Several novel technologies have been developed that look like they might overcome these challenges, which will be described in subsequent sections.

Number of Quantum Wells

A single quantum well of GaAs requires input current densities of approximately 100 A/cm^2 to achieve transparency; N wells require N times this current. To keep the threshold current less than 1 kA/cm^2 , then, means active regions with less than 10 QWs. The VCSEL requires an optical standing wave that has a period of a half-wavelength, which is approximately 120 nm in GaAs. The gain region should be confined to the quarter-wavelength region at the peak of the optical standing wave, a region of about 60 nm. Thus, a typical active region might consist of 3 QWs of 8 μm thickness, each separated by approximately 10 nm. Strained QWs have higher gain than unstrained QWs, and can be grown by adding indium or phosphorous into the composition.

InP-based QWs for 1.3 μm and 1.55 μm operation have somewhat lower gain, but typical geometries still use 3 to 5 QWs; higher reflectivity mirrors compensate for lower gain.

Mirror Reflectivity

When the mirror reflectivity R in a laser is close to 1, a simple expression for the threshold gain-length product, $G_L L$ is

$$G_L L = (1 - R_1)(1 - R_2) \quad (47)$$

Typical GaAs lasers have gains $G_L \sim 1000 \text{ cm}^{-1}$. For a quantum well thickness of 10 nm, the gain per quantum well is 10^{-3} , so that with 3 QW, reflectivities of ~ 98 percent for each mirror should be sufficient to achieve threshold. Very often, however, in order to lower the threshold much higher reflectivities are used, particularly on the back mirror.

The on-resonance Bragg mirror reflectivity is the square of the reflection coefficient r , given by

$$r = \frac{1 - (n_f/n_i)(n_l/n_h)^{2N}}{1 + (n_f/n_i)(n_l/n_h)^{2N}} \quad (48)$$

where there are N pairs of quarter-wavelength layers that alternate high-index and low-index (n_h and n_p respectively), and n_f and n_i are the refractive index of the final and initial media, respectively.⁴⁴

For high-reflectance Bragg mirrors, the second term in the numerator and denominator is small, and the reflectivity can be simplified to

$$\varepsilon = 1 - R = 1 - r^2 = 4 \left(\frac{n_f}{n_i} \right) \left(\frac{n_l}{n_h} \right)^{2N} \quad (49)$$

Higher reflectivity (smaller ε) is provided by either more layer pairs or a larger refractive index difference between the two compositions in the layer pairs. Also, Eq. (49) shows that internal mirrors ($n_f = n_i$) will have a smaller reflectivity than external mirrors ($n_f = 1$), for the same number of layer pairs. If the layer pair consists of GaAs ($n \sim 3.6$) and AlAs ($n \sim 3.0$), a mirror of 15 layer pairs will have an internal reflectivity $R = 98$ percent and an external reflectivity $R = 99.5$ percent. Thirty layer pairs are required to increase the internal reflectivity to 99.96 percent. Bragg mirrors with a smaller fraction of AlAs in the low-index layers will require even more layer pairs to achieve the same reflectivity.

In long-wavelength InP-based VCSELs, the $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}/\text{InP}$ refractive index difference is small and many layers are required to achieve sufficiently high reflectivity. Often the top DBR layer is replaced by a deposited mirror, such as alternating quarter-wave layers of ZnSe and MgF, or amorphous silicon and aluminum oxide. Hybrid mirrors use fewer dielectric layers and can terminate with a gold layer. Because the substrate is transparent, these devices are mounted upside-down, with the gold layer attaching to a thermal heat sink.

Another option has been to replace epitaxial InP-based DBR mirrors with fusion-bonded traditional GaAs-AlAs DBR mirrors that have been grown on GaAs substrates. Some commercial VCSELs have used this approach. Wafer fusion occurs when pressing the two wafers together (after removing oxide off their surfaces) at 15 atm and heating to 630°C under hydrogen for 20 min. Mirrors can be wafer-fused on both sides of the VCSEL by etching away the InP substrate and one of the GaAs substrates.

Finally, new multicomponent compositions show promise of alternative mirrors, especially by adding aluminum. Forty lattice-matched pairs of $\text{Al}_x\text{Ga}_{1-x}\text{In}_y\text{As}_{1-y}$ and InP quarter-wave layers have been shown to yield a reflectivity of 99.9 percent.⁴⁵ Some commercial VCSEL manufacturers have favored this all-epitaxial fabrication process.

An entirely different approach, which is not yet commercial, is to push GaAs-based devices out to longer wavelength. Adding dilute amounts of nitrogen yields an active layer of $\text{Ga}_x\text{In}_{1-x}\text{N}_y\text{As}_{1-y}$, which has been shown to produce VCSELs near the $1.3\text{-}\mu\text{m}$ wavelength range. It turns out that adding a small amount of antimony produces QWs of higher quality. This has suggested that the five-component material $\text{Ga}_x\text{In}_{1-x}\text{N}_y\text{As}_z\text{Sb}_{1-y-z}$ has the potential to achieve even lower bandgaps; VCSELs of this composition have been demonstrated, but not yet as far out in wavelength as $1.55 \mu\text{m}$.

Electrical Injection and Current Confinement

As shown in Fig. 28, light is emitted through a window hole in the top electrode. It is difficult to inject carriers through the Bragg reflector because the wide bandwidth layers provide potential wells that trap carriers; but this can be overcome with increased operating voltage. Sometimes graded layers are used in the DBR structure to reduce carrier trapping. Current confinement to a finite area was originally done by proton implantation or by etching mesas and then planarizing with polyimide. VCSELs of fairly large diameter ($>10\text{ }\mu\text{m}$) are straight-forward to make and are useful when low threshold and single-mode are not required. Recently a selective oxidation technique has been developed that enables a small oxide-defined current aperture. A high-aluminum fraction $\text{Al}_x\text{Ga}_{1-x}\text{As}$ layer ($\sim 98\%$) is grown above the active layer and a mesa is etched to below that layer. Then a long, soaking, wet-oxidization process selectively creates a ring of native oxide that can stop carrier transport. The chemical reaction moves in from the side of the etched pillar and is stopped when the desired diameter is achieved. Such a resistive aperture confines current only where needed, and can double the maximum conversion efficiency to almost 60 percent.

Threshold voltages less than 3 V are common in diameters $\sim 12\text{ }\mu\text{m}$. The oxide-defined current channel increases the efficiency, but tends to cause multiple transverse modes due to relatively strong oxide-induced index guiding. This may introduce modal noise into fiber communication systems. Single-mode requirements force the diameter to be very small (below 4 to $5\text{ }\mu\text{m}$) or for the design to incorporate additional features, as discussed in the following section.

Transverse injection eliminates the need for the current to travel through the DBR region, but typically requires even higher voltage. This approach has been proven useful when highly conductive layers are grown just above and below the active region. Because carriers have to travel farther with transverse injection, it is important that these layers have as high mobility as possible. This has been achieved by injecting carriers from both sides through n -type layers. Such a structure can still inject holes into the active layer if a buried tunnel junction is provided, as shown in Fig. 29. The tunnel junction consists of a single layer-pair of very thin highly doped n^{++} and p^{++} layers and must be located near the active layer so that its holes can be utilized. After the first As-based growth step, the tunnel junction is laterally structured by means of standard photolithography and chemical dry etching. It is then regrown with phosphorous-containing n -layers. The lateral areas surrounding the tunnel junction contain an npn electronic structure and do not conduct electricity. Only within the area of the tunnel junction are electrons from the n -type InP spacer converted into holes.

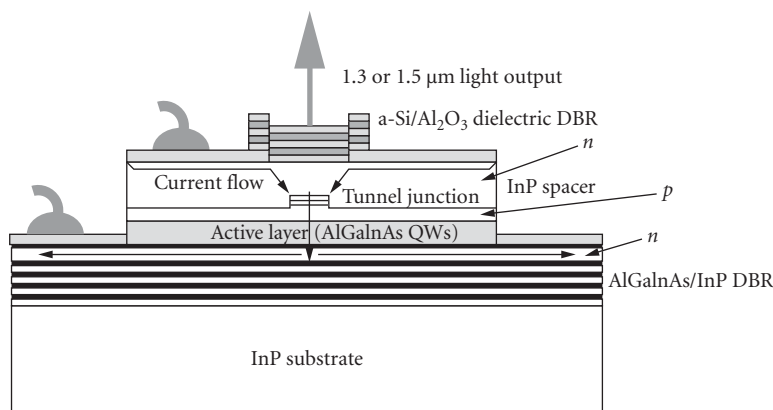


FIGURE 29 BHT geometry for single-mode VCSEL, showing flow of current. (Adapted from Ref. 45.)

Spatial Characteristics of Emitted Light

Single transverse mode operation remains a challenge for VCSELs, particularly at the larger diameters and higher current levels. When modulated, lateral spatial instabilities tend to set in and spatial hole burning causes transverse modes to jump. This can introduce considerable modal noise when coupling VCSEL light into fibers.

The two most common methods to control transverse modes are the same as used to control current spreading: ion implantation and an internal oxidized aperture. Ion implantation keeps the threshold relatively low, but thermal lensing coupled with weak index guiding is insufficient to prevent multilateral-mode operation due to spatial hole burning; also the implanted geometry does not provide inherent polarization discrimination. The current confining aluminum oxide aperture formed by selective oxidization acts as a spatial filter and encourages the laser to operate in low-order modes. Devices with small oxide apertures (2×2 to $4 \times 4 \mu\text{m}^2$) can operate in a single-mode. Devices with $3.5\text{-}\mu\text{m}$ diameters have achieved single-mode output powers on the order of 5 mW, but devices with larger apertures will rapidly operate in multiple transverse modes as the current is raised.⁴⁶

VCSELs with small diameter are limited in the amount of power they can emit while remaining single-mode, and their efficiency falls off as the diameter becomes smaller. A variety of designs have been reported for larger-aperture VCSELs that emit single-mode. This section lists several approaches; some have become commercially available and others are presently at the research stage: (1) Ion implantation and oxide-defined spatial filters have been *combined* with some success at achieving single-mode. (2) Etched pillar mesas favor single-mode operation because they have sidewall scattering losses that are higher for higher-order modes. The requirement is that the mode selective losses must be large enough to overcome the effects of spatial hole burning.⁴⁷ (3) As with traditional lasers, an etched pillar mesa can be overgrown to create a buried heterostructure (BH), providing a real index guide that can be structured to be single-mode.⁴⁸ (4) A BH design can be combined with ion implantation and/or selectively oxidized apertures, for the greatest design flexibility. (5) Surface relief has been integrated on top of the cladding layer (before depositing a top dielectric mirror), physically structuring it so as to eliminate higher-order modes; the surface relief incorporates a quarter-wave ring structure that decreases the reflectivity for higher-order modes.⁴⁹ (6) The VCSEL can be surrounded with a second growth of *higher* refractive semiconductor material, which causes an *antiguide* that preferentially confines the lowest-order mode.⁵⁰ (7) A photonic crystal has been incorporated under the top dielectric mirror, which provides an effective graded index structure that favors maintaining a single-mode.⁵¹

All of these approaches can be used with current injection through the DBR mirrors, or with lateral injection, usually through an oxide aperture. A number of single-mode geometries use buried tunnel junctions (BTJ), which may be made with small enough area to create single-mode lasers without incorporating any other features.⁴⁵ Often higher powers are achieved by combining wider area BTJ along with some of the other approaches outlined above.

Light Out versus Current In

The VCSEL will, in general, have similar L - I performance to edge-emitting laser diodes, with some small differences. Because the acceptance angle for the mode is higher than in edge-emitting diodes, there will be more spontaneous emission, which will show up as a more graceful turn-on of light out versus voltage in. As previously mentioned, the operating voltage is 2 to 3 times that of edge-emitting lasers. Thus, Eq. (5) must be modified to take into account the operating voltage drop across the resistance R of the device. The operating power efficiency is

$$\eta_{\text{eff}} = \eta_D \frac{I_{\text{op}} - I_{\text{th}}}{I_{\text{th}}} \frac{V_g}{V_g + I_{\text{op}} R} \quad (50)$$

Small diameter single-mode VCSELs would typically have a $5\text{-}\mu\text{m}$ radius, a carrier injection efficiency of 80 to 90 percent, an internal optical absorption loss $\alpha_i L$ of 0.003, an optical scattering loss

of 0.001, and a net transmission through the front mirror of 0.005 to 0.0095. Carrier losses reducing the quantum efficiency are typically due to spontaneous emission in the wells, spontaneous emission in the barriers, Auger recombination, and carrier leakage.

Typical commercial VCSELs designed for compatibility with single-mode fiber incorporate an 8- μm proton implantation window and 10- μm -diameter window in the top contact. Such diodes may have threshold voltages of ~ 3 V and threshold currents of a few milliamperes. These lasers may emit up to ~ 2 mW maximum output power. Devices will operate in zero-order transverse spatial mode with gaussian near-field profile when operated with DC drive current less than about twice threshold. When there is more than one spatial mode, or both polarizations, there will usually be kinks in the L - I curve, as with multimode edge-emitting lasers.

Spectral Characteristics

Since the laser cavity is short, the longitudinal modes are much farther apart in wavelength than in a cleaved cavity laser, typically separated by 50 nm, so only one longitudinal mode will appear, and there is longitudinal mode purity. The problem is with *lateral* spatial modes, since at higher power levels the laser does not operate in a single spatial mode. Each spatial mode will have a slightly different wavelength, perhaps separated by 0.01 to 0.02 nm. Lasers that start out as single-mode and single frequency at threshold will often demonstrate frequency broadening at higher currents due to multiple modes, as shown in Fig. 30a.⁵²

Even when the laser operates in a single spatial mode, it may have two orthogonal directions of polarization (discussed next), that will exhibit different frequencies, as shown in Fig. 30b.⁵³ Thus both single-mode and polarization stability are required to obtain a true single-mode.

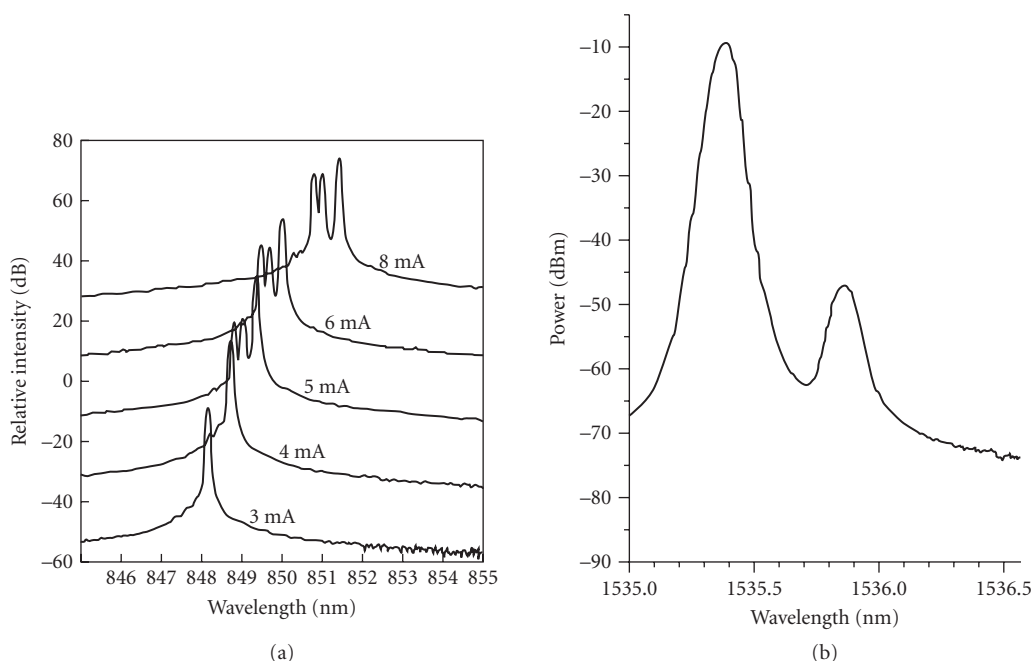


FIGURE 30 VCSEL spectra: (a) emission spectra recorded at different injection currents for BH-VCSELs of 10 μm diameter⁵² and (b) different emission spectra due to different polarizations of a BJT single-mode VCSEL.⁵³

When a VCSEL is modulated, lateral spatial instabilities may set in and spatial hole burning may cause transverse modes to jump. This can broaden the spectrum. In addition, external reflections can cause instabilities and increased relative intensity noise, just as in edge-emitting lasers.⁵⁴ For very short cavities, such as between the VCSEL and a butt-coupled fiber (with ~ 4 percent reflectivity), instabilities do not set in, but the output power can be affected by the additional mirror, which forms a Fabry-Perot cavity with the output mirror and can reduce or increase its effective reflectivity, depending on the round-trip phase difference. When the external reflection comes from ~ 1 cm away, bifurcations and chaos can be introduced with a feedback parameter $F > 10^{-4}$, where $F = C_e \sqrt{f_{\text{ext}}}$, with C_e and f_{ext} as defined in the discussion surrounding Eq. (31). For $R_o = 0.995$, $R_{\text{ext}} = 0.04$, the feedback parameter $F \sim 10^{-3}$, instabilities can be observed if reflections get back into the VCSEL.

Polarization

A VCSEL with a circular aperture has no preferred polarization state. The output tends to oscillate in linear but random polarization states, which may wander with time (and temperature) and may have slightly different emission wavelengths (Fig. 30*b*). Polarization-preserving VCSELs require breaking the symmetry by introducing anisotropy in the optical gain or loss. Some polarization selection may arise from an elliptical current aperture. The strongest polarization selectivity has come from growth on (311) GaAs substrates, which causes anisotropic gain.

Commercial VCSELs

The most readily available VCSELs are GaAs-based, emitting at 850-nm wavelength. Commercial specifications for these devices list typical multimode output powers from 1 to 2.4 mW and single-mode output powers from 0.5 to 2 mW, depending on design. Drive voltages vary from 1.8 to 3 V, with series resistance typically about 100 Ω . Spectral width for multimode lasers is about 0.1 nm, while for single-mode lasers it can be as narrow as 100 MHz. Beam divergence FWHM is typically 18° to 25° for multimode lasers and 8° to 12° for single-mode VCSELs, with between 20 to 30 dB sidemode suppression.⁵⁵

Red VCSELs, emitting at 665 nm, are available from fewer suppliers, and have output powers of 1 mW, threshold currents between 0.6 and 2.5 mA, and operating voltages of 2.8 to 3.5 V. Their divergence angle is 14° to 20° and slope efficiency is 0.9 mW/mA, with sidemode suppression between 14 and 50 dB. Reported bandwidths are 3 to 3.5 GHz.⁵⁶

Long-wavelength VCSELs have output powers between 0.7 and 1 mW, with threshold currents between 1.1 and 2.5 mA. Series resistance is 100 Ω , with operating voltage between 2 and 3 V. Single-mode spectral width is 30 MHz and modulation bandwidth is 3 GHz. Sidemode suppression is between 30 and 40 dB, with slope efficiency of 0.2 mW/mA and an angular divergence between 9° and 20° .⁵⁷

13.11 LITHIUM NIOBATE MODULATORS

The most direct way to create a modulated optical signal for communication application is to directly modulate the current driving the laser diode. However, as discussed in Secs. 13.4 and 13.5, this may cause turn-on delay, relaxation oscillation, mode-hopping, and/or chirping of the optical wavelength. Therefore, an alternative often used is to operate the laser in a continuous manner and to place a modulator after the laser. This modulator turns the laser light off and on without impacting the laser itself. The modulator can be butt-coupled directly to the laser, located in the laser chip package and optically coupled by a microlens, or remotely attached by means of a fiber pigtail between the laser and modulator.

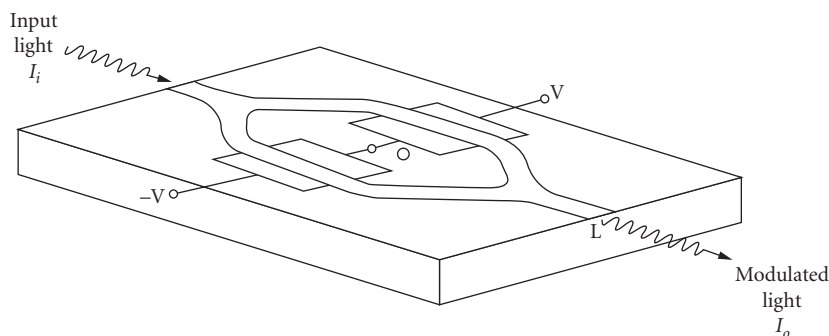


FIGURE 31 Y-branch interferometric modulator in the “push-pull” configuration. Center electrodes are grounded. Opposite polarity electrodes are placed on the outsides of the waveguides. Light is modulated by applying positive or negative voltage to the outer electrodes.

Lithium niobate modulators have become one of the main technologies used for high-speed modulation of continuous-wave (CW) diode lasers, particularly in applications (such as cable television) where extremely linear modulation is required, or where chirp is to be avoided at all costs. These modulators operate by the electro-optic effect, in which an applied electric field changes the refractive index. Integrated optic waveguide modulators are fabricated by diffusion into lithium niobate substrates. The end faces are polished and butt-coupled (or lens-coupled) to a single-mode fiber pigtail (or to the laser driver itself). This section describes the electro-optic effect in lithium niobate, its use as a phase modulator and an intensity modulator, considerations for high-speed operation, and the difficulties in achieving polarization independence.⁵⁸

Most common is the Y-branch interferometric modulator shown in Fig. 31, discussed in a following subsection. The waveguides that are used for these modulators are fabricated in lithium niobate either by diffusing titanium into the substrate from a metallic titanium strip or by means of ion exchange. The waveguide pattern is obtained by photolithography. The standard thermal indiffusion process takes place in air at 1050°C over 10 hours. An 8- μm -wide, 50-nm thick strip of titanium creates a fiber-compatible single-mode at $\lambda = 1.3 \mu\text{m}$. The process introduces ~ 1.5 percent titanium at the surface, with a diffusion profile depth of $\sim 4 \mu\text{m}$. The result is a waveguide with increased extraordinary refractive index of 0.009 at the surface and an ordinary refractive index change of ~ 0.006 . A typical modulator will incorporate aluminum electrodes 2 cm long, deposited on either side of the waveguides, with a gap of 10 μm .

In the case of ion exchange, the lithium niobate sample is immersed in a melt containing a large proton concentration (typically benzoic acid or pyrophosphoric acid at $>170^\circ\text{C}$), with nonwaveguide areas protected from diffusion by masking; the lithium near the surface of the substrate is replaced by protons, which increases the refractive index. Ion-exchange alters only the extraordinary polarization; that is, only light polarized parallel to the z axis is waveguided. Thus, it is possible in lithium niobate to construct a polarization-independent modulator with titanium indiffusion, but not with proton-exchange. Nonetheless, ion exchange creates a much larger refractive index change (~ 0.12), which provides more flexibility in modulator design. Annealing after diffusion can reduce insertion loss and restore the degraded electro-optic effect. Interferometric modulators with moderate index changes ($\Delta n < 0.02$) are insensitive to aging at temperatures of 95°C or below. Using higher index change devices, or higher temperatures, may lead to some degradation with time. Tapered waveguides can be fabricated easily by ion exchange for high coupling efficiency.⁵⁹

Electro-Optic Effect

The *electro-optic effect* is the change in refractive index that occurs in a noncentrosymmetric crystal in the presence of an applied electric field. The linear electro-optic effect is represented by a

third-rank tensor for the refractive index. However, using symmetry rules it is sufficient to define a reduced tensor r_{ij} , where $i = 1, \dots, 6$ and $j = x, y, z$, denoted as 1, 2, 3. Then, the linear electro-optic effect is traditionally expressed as a linear change in the inverse refractive index tensor squared (see Chap. 7 in this volume):

$$\Delta\left(\frac{1}{n^2}\right)_i = \sum_j r_{ij} E_j \quad j = x, y, z \quad (51)$$

where E_j is the component of the applied electric field in the j th direction. In isotropic materials, r_{ij} is a diagonal tensor. An applied electric field can introduce off-diagonal terms in r_{ij} , as well as change the lengths of the principle dielectric axes. The general case is treated in Chap. 13, Vol. II. In lithium niobate (LiNbO_3), the material of choice for electro-optic modulators, the equations are simplified because the only nonzero components and their magnitudes are⁶⁰

$$\begin{aligned} r_{33} &= 31 \times 10^{-12} \text{ m/V} & r_{13} &= r_{23} = 8.6 \times 10^{-12} \text{ m/V} \\ r_{51} &= r_{42} = 28 \times 10^{-12} \text{ m/V} & r_{22} &= -r_{12} = -r_{61} = 3.4 \times 10^{-12} \text{ m/V} \end{aligned}$$

The crystal orientation is usually chosen so as to obtain the largest electro-optic effect. This means that if the applied electric field is along z , then light polarized along z sees the largest field-induced change in refractive index. Since $\Delta(1/n^2)_3 = \Delta(1/n_z^2) = r_{33}E_z$, performing the difference gives

$$\Delta n_z = -\frac{n_z^3}{2} r_{33} E_z \Gamma \quad (52)$$

A *filling factor* Γ (also called an *optical-electrical field overlap parameter*) has been included due to the fact that the applied field may not be uniform as it overlaps the waveguide, resulting in an effective field that is somewhat less than 100 percent of the maximum field.

In the general case for the applied electric field along z , the tensor remains diagonal and $\Delta(1/n^2)_1 = r_{13}E_z = \Delta(1/n^2)_2 = r_{23}E_z$, and $\Delta(1/n^2)_3 = r_{33}E_z$. This means that the index ellipsoid has not rotated, its axes have merely changed in length. Light polarized along any of these axes will see a pure phase modulation. Because r_{33} is largest, polarizing the light along z and providing the applied field along z will provide the largest phase modulation for a given field. Light polarized along either x or y will have the same index change, which might be a better direction if polarization-independent modulation is desired. However, this would require light to enter along z , which is the direction in which the field is applied, so it is not practical.

As another example, consider the applied electric field along y . In this case the nonzero terms are

$$\Delta\left(\frac{1}{n^2}\right)_1 = r_{12}E_y \quad \Delta\left(\frac{1}{n^2}\right)_2 = r_{22}E_y = -r_{12}E_y \quad \Delta\left(\frac{1}{n^2}\right)_4 = r_{42}E_y \quad (53)$$

There is now a yz cross-term, coming from r_{42} . Diagonalization of the perturbed tensor finds new principal axes, only slightly rotated about the z axis. Therefore, the principal refractive index changes are essentially along the x and y axes, with the same values as $\Delta(1/n^2)_1$ and $\Delta(1/n^2)_2$ in Eq. (53). If light enters along the z axis without a field applied, both polarizations (x and y) see an ordinary refractive index. With a field applied, both polarizations experience the same phase change (but opposite sign). In a later section titled "Polarization Independence," we describe an interferometric modulator that does not depend on the sign of the phase change. This modulator is polarization independent, using this crystal and applied-field orientation, at the expense of operating at somewhat higher voltages, because $r_{22} < r_{33}$.

Since lithium niobate is an insulator, the direction of the applied field in the material depends on how the electrodes are applied. Figure 32 shows a simple phase modulator. Electrodes that straddle the modulator provide an in-plane field as the field lines intersect the waveguide, as shown in Fig. 32b.

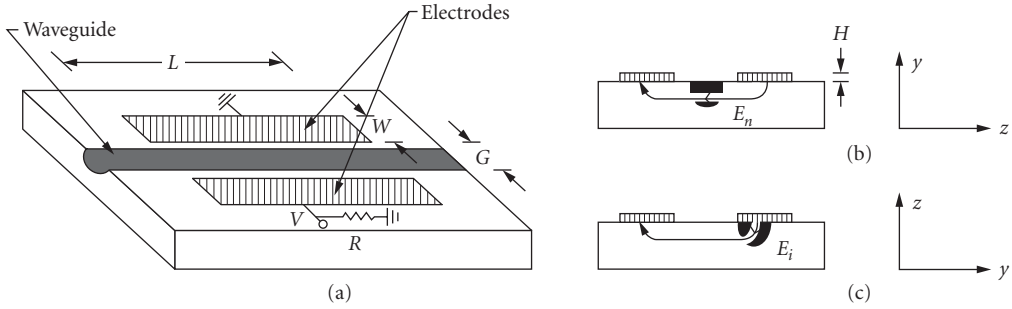


FIGURE 32 (a) Geometry for phase modulation in lithium niobate with electrodes straddling the channel waveguide. (b) End view of (a), showing how the field in the channel is parallel to the surface. (c) End view of a geometry placing one electrode over the channel, showing how the field in the channel is essentially normal to the surface.

This requires the modulator to be *y-cut* LiNbO_3 (the y axis is normal to the wafer plane), with the field lines along the z direction; *x-cut* LiNbO_3 will perform similarly. Figure 32c shows a modulator in *z-cut* LiNbO_3 . In this case, the electrode is placed over the waveguide, with the electric field extending downward through the waveguide (along the z direction). The field lines will come up at a second, more distant electrode. In either case, the field may be fringing and nonuniform, which is why the filling factor Γ has been introduced.

Phase Modulation

Applying a field to one of the geometries shown in Fig. 32 results in pure *phase modulation*. The field is roughly V/G , where G is the gap between the two electrodes. For an electrode length L , the phase shift is

$$\Delta\phi = \Delta n_z kL = -\frac{n_o^3}{2} r_{33} \left(\frac{V}{G} \right) \Gamma kL \quad (54)$$

The refractive index for bulk LiNbO_3 is given by⁶¹

$$n_o = 2.195 + \frac{0.037}{[\lambda(\mu\text{m})]^2} \quad \text{and} \quad n_e = 2.122 + \frac{0.031}{[\lambda(\mu\text{m})]^2}$$

Inserting numbers for $\lambda = 1.55 \mu\text{m}$ gives $n_o = 2.21$. When $G = 10 \mu\text{m}$ and $V = 5 \text{ V}$, a π phase shift is expected in a length $L \sim 1 \text{ cm}$.

It can be seen from Eq. (54) that the electro-optic phase shift depends on the product of the length and voltage. Longer modulators can use smaller voltages to achieve a π phase shift. Shorter modulators require higher voltages. Thus, the figure of merit for phase modulators is typically the product of the voltage required to reach π times the length. The modulator just discussed has a $5\text{-V}\cdot\text{cm}$ figure of merit.

The electro-optic phase shift has a few direct uses, such as providing a *frequency shifter* (since $\partial\phi/\partial t \propto \Delta v$). However, in communication systems this phase shift is generally used in an interferometric configuration to provide intensity modulation, discussed in the following section.

Y-Branch Interferometric (Mach-Zehnder) Modulator

The *interferometric modulator* is shown schematically in Fig. 31. This geometry allows waveguided light from the two branches to interfere, forming the basis of an intensity modulator. The amount of

interference is tunable by providing a relative phase shift on one arm with respect to the other. Light entering a single-mode waveguide is equally divided into the two branches at the Y junction, initially with zero relative phase difference. The guided light then enters the two arms of the waveguide interferometer, which are sufficiently separated that there is no coupling between them. If no voltage is applied to the electrodes, and the arms are exactly the same length, the two guided beams arrive at the second Y junction in phase and enter the output single-mode waveguide in phase. Except for small radiation losses, the output is equal in intensity to the input. However, if a π phase difference is introduced between the two beams via the electro-optic effect, the combined beam has a lateral amplitude profile of odd spatial symmetry. This is a second-order mode and is not supported in a single-mode waveguide. The light is thus forced to radiate into the substrate and is lost. In this way, the device operates as an electrically driven optical intensity on-off modulator. Assuming perfectly equal splitting and combining, the fraction of light transmitted is

$$\eta = \left[\cos\left(\frac{\Delta\phi}{2}\right) \right]^2 \quad (55)$$

where $\Delta\phi$ is the difference in phase experienced by the light in the different arms of the interferometer: $\Delta\phi = \Delta n k L$, where $k = 2\pi/\lambda$, Δn is the difference in refractive index between the two arms, and L is the path length of the field-induced refractive index difference. The voltage at which the transmission goes to zero ($\Delta\phi = \pi$) is usually called V_π . By operating in a push-pull manner, with the index change increasing in one arm and decreasing in the other, the index difference Δn is twice the index change in either arm. This halves the required voltage.

Note that the transmitted light is periodic in phase difference (and therefore voltage). The response depends only on the integrated phase shift and not on the details of its spatial evolution. Therefore, nonuniformities in the electro-optically induced index change that may occur along the interferometer arms do not affect the extinction ratio. This property has made the Mach Zehnder (MZ) modulator the device of choice in communication applications.

For analog applications, where linear modulation is required, the modulator is prebiased to the quarter-wave point (at voltage $V_b = \pi/2$), and the transmission efficiency becomes linear in $V - V_b$ (for moderate excursions):

$$\eta = \frac{1}{2} \left[1 + \sin \frac{\pi(V - V_b)}{2V_\pi} \right] \approx \frac{1}{2} + \frac{\pi}{2} \frac{(V - V_b)}{V_\pi} \quad (56)$$

The electro-optic effect depends on the polarization. For the electrode configuration shown here, the applied field is in the plane of the lithium niobate wafer, and the polarization of the light to be modulated must also be in that plane. This will be the case if a TE-polarized laser diode is butt-coupled (or lens-coupled) with the plane of its active region parallel to the plane of the lithium niobate wafer, and if the wafer is Y -cut. Polarization-independent modulation requires a different orientation, to be described later. First, however, we discuss the electrode requirements for high-speed modulation.

High-Speed Operation

The optimal electrode design depends on how the modulator is to be driven. Because the electrode is on the order of 1 cm long, the fastest devices require traveling-wave electrodes rather than lumped electrodes. Lower-speed modulators can use lumped electrodes, in which the modulator is driven as a capacitor terminated in a parallel resistor matched to the impedance of the source line. The modulation speed depends primarily on the RC time constant determined by the electrode capacitance and the terminating resistance. To a smaller extent, the speed also depends on the resistivity of the electrode itself. The capacitance per unit length is a critical design parameter. This depends on the material dielectric constant, the electrode gap G and the electrode width W . With

increasing G , the capacitance per unit length decreases and the bandwidth-length product increases essentially logarithmically. In LiNbO_3 , when the electrode widths and gap are equal, the capacitance per unit length is 2.3 pF/cm and the bandwidth-length product is $\Delta f_{\text{RC}} L = 2.5 \text{ GHz} \cdot \text{cm}$. The trade-off is between large G/W to reduce capacitance and small G/W to reduce drive voltage and electrode resistance. The ultimate speed of lumped electrode devices is limited by the electric signal transit time, with a bandwidth-length product of 2.2 GHz · cm. The way to achieve higher speed modulation is to use traveling-wave electrodes.

The traveling-wave electrode is a miniature transmission line. Ideally, the impedance of this coplanar line is matched to the electrical drive line and is terminated in its characteristic impedance. In this case, the modulator bandwidth is determined by the difference in velocity between the optical and electrical signals (velocity mismatch or walk-off), and any electrical propagation loss. Because of competing requirements between a small gap to reduce drive voltage and a wide electrode width to reduce RF losses, as well as reflections at any impedance transition, subtle trade-offs must be considered in designing traveling-wave devices.

Lithium niobate MZ modulators operating out to 35 GHz at $\lambda = 1.55 \mu\text{m}$ are commercially available, with $V_\pi = 10 \text{ V}$, with <5 dB insertion loss and >20 dB extinction ratio.⁶² To operate near quadrature, which is the linear modulation point, a bias voltage of ~4 V is required. Direct coupling from a laser or polarization-maintaining fiber is required, since these modulators operate on only one polarization.

Insertion Loss

Modulator insertion loss can be due to Fresnel reflection at the lithium niobate-air interfaces, which can be reduced by antireflection coatings or index matching (which only helps, but does not eliminate this loss, because of the very high refractive index of lithium niobate). The other cause of insertion loss is mode mismatch. To match the spatial profile of the fiber mode, a deep and buried waveguide must be diffused. Typically, the waveguide will be 9 μm wide and 5 μm deep. While the in-plane mode can be gaussian and can match well to the fiber mode, the out-of-plane mode is asymmetric, and its depth must be carefully optimized. In an optimized modulator, the coupling loss per face is about 0.35 dB and the propagation loss is about 0.3 dB/cm. This result includes a residual index-matched Fresnel loss of 0.12 dB.

Misalignment can also cause insertion loss. An offset of 2 μm typically increases the coupling loss by 0.25 dB. The angular misalignment must be maintained below 0.5° in order to keep the excess loss below 0.25 dB.⁶³

Propagation loss comes about from absorption, metallic overlay, scattering from the volume or surface, bend loss, and excess loss in the Y-branches. Absorption loss at 1.3 and 1.55 μm wavelengths appears to be <0.1 dB/cm. Bend loss can be large, unless any curvature of guides is small. The attenuation coefficient in a bend has the form $\alpha = C_1 \exp(-C_2 R)$, where $C_1 = 15 \text{ mm}^{-1}$ and $C_2 = 0.4 \text{ mm}^{-1}$ in titanium indiffused lithium niobate, at wavelengths around 1.3 to 1.5 μm . This means that a 5-mm-long section of constant radius 20 mm will introduce only 0.1 dB of excess loss.⁶³

A final source of loss in Y-branches is excess radiation introduced by sharp transitions at the Y junction. These junctions must be fabricated carefully to avoid such losses, since tolerances on waveguide roughness are critically small.

Polarization Independence

As previously shown, if the light is incident along the z axis and the field is along the y axis, then light polarized along x and y experience the same phase shift, but with opposite signs. This requires an x -cut crystal, with an in-plane field along y , which provides polarization-independent interferometric modulation at the sacrifice of somewhat higher half-wave voltage (e.g., 17 V).⁶⁴ Because of the difficulty of achieving exactly reproducible lengths in the two arms of the Y-branch interferometer, it was found useful to do a postfabrication phase correction using laser ablation.

Photorefractivity and Optical Damage

Lithium niobate exhibits *photorefractivity*, also called *optical damage*, when it is a nuisance. This phenomenon is a change in refractive index as a result of photoconduction originating in weak absorption by deep traps and a subsequent redistribution of charges within the lithium niobate. Because the photoconductive crystal is electro-optic, the change in electric field resulting from charge motion shows up as a change in refractive index, altering the phase shift as well as the waveguiding properties. While photorefractivity seriously limits the performance of lithium niobate modulators at shorter wavelengths (even at 850 nm),⁶⁵ it is not a serious concern at 1.3 μm and 1.55 μm .

However, partial screening by photocarriers may cause a drift in the required bias voltage of modulators, and systems designers may need to be sensitive to this.

Multimode Interferometric Mach-Zehnder Modulator

An alternative geometry has been developed for the waveguide Mach-Zehnder interferometer that replaces the Y branch by a multimode interferometric (MMI) power splitter, which is shown in Fig. 33. The MMI modulator works on the principle that there are distances at which light traveling down a step-index waveguide self-images. These self-imaging planes are analogous to Talbot planes for plane waves. The process of self-imaging is indicated in Fig. 33a.⁶⁶ It can be seen that if light is incident in a spatial distribution localized to only one-half of the guide, at a distance of $1/2 (3L_\pi)$ the power will be split equally into two spatial distributions. This enables a simple power splitter, as shown in Fig 33b.⁶⁶

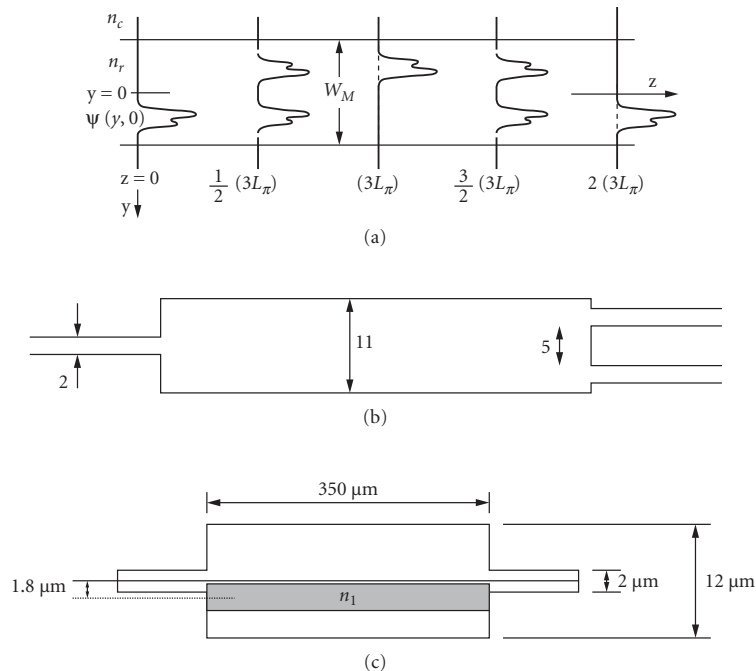


FIGURE 33 Multimode interferometric (MMI) devices as observed from above: (a) multimode waveguide showing the input field $\Psi(y, 0)$, a mirrored single image at $(3L_\pi)$, a direct single image at $2(3L_\pi)$, and two-fold images at $1/2 (3L_\pi)$ and $3/2 (3L_\pi)$,⁶⁶ (b) power splitter, achieved at $1/2 (3L_\pi)$ with typical dimensions as shown in micrometers,⁶⁶ and (c) interference modulator, achieved at $(3L_\pi)$, with voltage applied by the darkened electrode; ground plane is under the substrate.⁶⁷

In a step-index multimode guide, because the modes are equally spaced, the beat length L_π is given by

$$L_\pi = \pi/(\beta_o - \beta_i) \approx 4n_{co} W_e^2/3\lambda_o \quad (57)$$

where W_e is the effective width of the guide, including Goos-Hanschen shifts. For $W_e = 11 \mu\text{m}$, this gives $L_\pi \sim 0.5 \text{ mm}$ in semiconductors, so a power splitter could be less than a millimeter long.

Ultracompact modulators can be constructed by the method shown in Fig. 33c,⁶⁷ with an electrode applied to only part of the structure, which will nonuniformly change the phases and destroy the coherent mixing that self-images the light to the output waveguide. Different electrode structures and electro-optic media have been demonstrated, including prism electrodes in electro-optic polymers.⁶⁸

Electro-Optic Polymer Modulators

For a number of years researchers have been working to make polymer modulators a reality. Electro-optic polymers contain nonlinear chromophores capable of providing large electro-optic effects at transmission wavelengths of 1.3 and 1.55 μm , with electro-optic coefficients of $\sim 70 \text{ pm/V}$, with the potential of $\sim 100 \text{ pm/V}$, which is 3 to 4 times greater than of lithium niobate. The technology typically involves spin coating of the polymer onto a silicon substrate, poling it, and applying electrodes. Electro-optic polymers are compatible with semiconductor fabrication methods and have low dielectric constants, so they can be easily velocity matched when applying traveling-wave electrodes for high-speed switching.

However, electro-optic polymers have had serious reliability issues. Gradual relaxation of the electro-optic coefficient is due to the slow misalignment of the chromophores, which form antiparallel pairs and effectively cancel out the electro-optic effect. This problem has been largely resolved using chemical anchoring schemes and/or poling in inert environments. There continues to be a problem with gradual changes in refractive index; however, some device designs are relatively robust to refractive index changes. The MMI is proposed to be such a structure, with simple design tolerance toward fluctuation of refractive indices.

For these modulators, highly active chromophores are dispersed at high concentrations in host polymers. A traditional commercially available chromophore is Disperse Red 1 (DR1), while much higher activity chromophores have been researched and are rapidly becoming commercially available. The polymer host is amorphous polycarbonate or polysulfone. The chromophore-doped polymer is dissolved in cyclohexanone (CHN) and tetrahydrofuran and then spin coated on the substrate. Cladding is typically spun on using commercially available acrylate. Relief structures, such as ridge waveguides, are plasma-etched into the films. After electrodes are deposited, the chromophore-doped polymer is made electro-optic by poling. In this process the sample is heated (e.g., at 130°C for 15 min) with a constant voltage applied (e.g., 600 V) between the bottom and the top electrodes. The temperature is dropped back to room temperature with the electric field still applied to ensure that the chromophores are oriented to achieve noncentrosymmetric alignment.⁶⁹

13.12 ELECTROABSORPTION MODULATORS

Semiconductors exhibit field-dependent absorption and refractive index. Such modulators can be integrated directly on the same chip as the laser, or placed external to the laser chip. External modulators may be butt-coupled to the laser, coupled by means of a microlens or by means of a fiber pigtail.

Electroabsorption

The electric field dependence of the absorption near the band edge of a semiconductor is called *electroabsorption*, and is particularly strong in quantum wells, where it is often called the *quantum-confined*

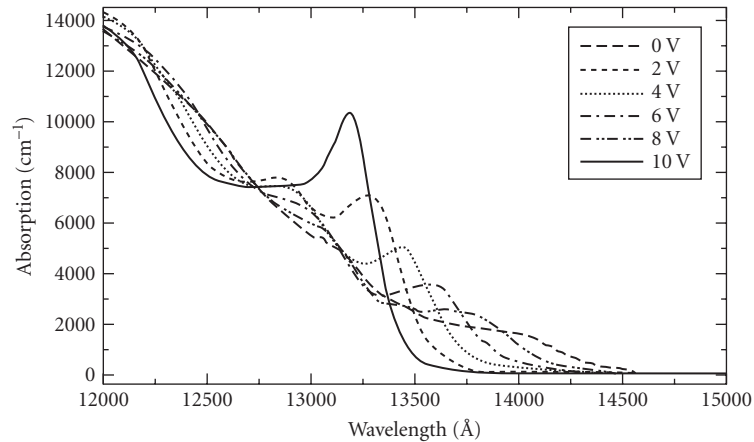


FIGURE 34 Spectrum of quantum-confined Stark effect (QCSE) in InAsP/InP-strained MQWs. The absorption changes with applied field.⁷⁰

Stark Effect (QCSE) An example of the wavelength dependence of the QCSE is shown in Fig. 34. The absorption spectrum of QWs exhibits a peak at the *exciton resonance*: when a field is applied, the exciton resonance moves to longer wavelengths, becomes weaker, and broadens. This means that on the long-wavelength side, the absorption increases with field, as the exciton resonance moves to longer wavelengths. At wavelengths closer to the exciton resonance, the absorption will first increase with field, then plateau, and finally decrease, as the field continues to grow. At wavelengths shorter than the zero-field exciton resonance, the absorption will decrease with increasing field, as the resonance moves to longer wavelengths.

While electroabsorption in QWs is much larger than in bulk, due to the sharpness of the excitonic-enhanced absorption edge, the useful absorption change must be multiplied by the filling factor of the QW in the waveguide, which reduces its effective magnitude. Under some conditions, electroabsorption near the band edge in bulk semiconductors (typically called the *Franz-Keldysh effect*) may also be useful in electroabsorption modulators.

Waveguide Modulators

When light traverses a length of QW material, the transmission will be a function of applied voltage. An electroabsorption modulator consists of a length of waveguide containing QWs. The waveguide is necessary to confine the light to the QW region so that it does not diffract away. Thus, low-refractive-index layers must surround the layer containing the QWs. Discrete electroabsorption modulators are typically made by using geometries similar to those of edge-emitting lasers (Fig. 1), but without mirrors. They are cleaved, antireflection coated, and then butt-coupled to the laser chip. They are operated by a reverse bias, rather than the forward bias of a laser. Or the modulator can be integrated on the laser chip, with the electroabsorption modulation region following a DFB or DBR laser in the optical train, as shown in Fig. 35. This figure shows the simplest electroabsorption modulator, with the same MQW composition as the DFB laser. This ridge waveguide device has been demonstrated with a 3-dB bandwidth of 30 GHz. The on-off contrast ratio is 12.5 dB for a 3-V drive voltage in a 90- μm -long modulator.⁷¹ The use of the same QWs is possible by setting the grating that determines the laser wavelength to well below the exciton resonance. Because of the inherently wide-gain spectrum exhibited by strained layer MQWs, this detuning is possible for the laser and still allows it to operate in the optimal wavelength region for the electroabsorption modulator.

Other integrated electroabsorption modulators use a QW composition in the electroabsorption region that is different from that of the laser medium. Techniques for integration are discussed later.

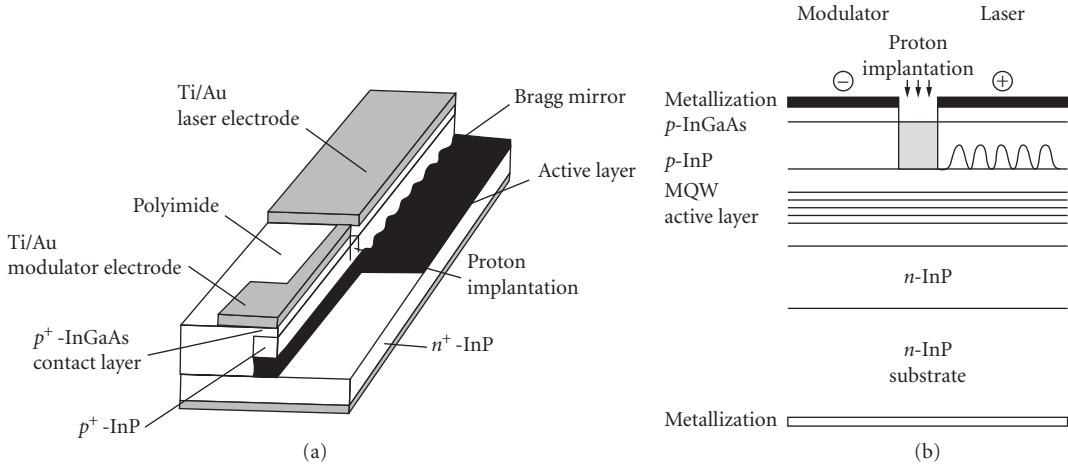


FIGURE 35 Electroabsorption modulator: (a) geometry for a channel electroabsorption modulator (foreground) integrated on the same chip with a DFB laser (background, under the Bragg mirror) and (b) side view, showing how the same MQW active layer can be used under forward bias with a grating to provide a DFB laser, and in a separate region under reverse bias for modulation, with the two regions electrically separated by proton implantation.⁵⁰

Intensity Modulation by Electroabsorption

In an electroabsorption waveguide modulator of length L , where the absorption is a function of applied field E , the transmission is a function of field: $T(E) = \exp[-\alpha(E)L]$, where α is the absorption per unit length, averaging the QW absorption over the entire waveguide. (That is, α is the QW absorption multiplied by the filling factor of the QW in the waveguide.) Performance is usually characterized by two quantities: *insertion loss* (throughput at high transmission) and *contrast ratio* (ratio of high transmission to low transmission). Assume that the loss in the QW, initially at low value α_- , increases by $\delta\alpha$. The *contrast ratio* is given by: $CR \equiv T_{\text{high}}/T_{\text{low}} = \exp(\delta\alpha L)$. The *insertion loss* is given by $A \equiv 1 - T_{\text{high}} = 1 - \exp(-\alpha_- L) \approx \alpha_- L$. A long path length L means a high contrast ratio but also a large insertion loss and large capacitance, which results in a slower speed. Choosing the most practical length for any given application requires trading off the contrast ratio against insertion loss and speed.

To keep a moderate insertion loss, waveguide lengths should be chosen so that $L \approx 1/\alpha_-$, which sets the contrast ratio as $CR = \exp(\delta\alpha/\alpha_-)$. The contrast ratio depends on the ratio of the change in absorption to the absorption in the low-loss state; this fact is used to design the QW composition and dimensions relative to the wavelength of operation. In general, the contrast ratio improves farther from the band edge, but the maximum absorption is smaller there, so the modulator must be longer, which increases its capacitance, decreases its speed, and increases its loss. Contrast ratios may reach 10/1 or more with <2 V applied for optimized electroabsorption modulators. In a waveguide, the contrast ratio does not depend on the filling factor of the QW in the waveguide, but the required length L does. Since high-speed modulators require small capacitance and small length, the filling factor should be as high as possible.

Waveguide modulators are used at wavelengths where the absorption is not too large, well below the band edge. In this wavelength region, electroabsorption at a fixed wavelength can be modeled by a pure quadratic dependence on field. Thus $\alpha(E) \approx \alpha_0 + \alpha_2 E^2$, where α_2 will typically depend on the wavelength, the QW and barrier dimensions and composition, and the waveguide filling factor. Intimately connected with this change in absorption is a change in refractive index with a similar field dependence: $\delta n(E) \approx n_2 E^2$, where n_2 is also strongly dependent on wavelength. Both electroabsorption and electrorefraction are about an order of magnitude larger in QWs than in bulk material. Specific

numerical values depend on the detailed design, but typical values are on the order of $\alpha_0 \sim 100 \text{ cm}^{-1}$, $\delta\alpha \sim 1000 \text{ cm}^{-1}$, $L \sim 200 \text{ }\mu\text{m}$ for 2 V applied across an i region $2.5\text{-}\mu\text{m}$ thick, for a field of $\sim 10 \text{ kV/cm}$. This means $\alpha_2 \sim 2 \times 10^{-5} \text{ cm/V}^2$. Also, $n_2 \sim 2 \times 10^{-11} \text{ V}^{-2}$.

Applying a Field in a Semiconductor

The electric field is usually applied by reverse biasing a *pin* junction. The electric field is supported by the semiconductor depletion region that exists within a *pin* junction, or at a metal-semiconductor junction (Schottky barrier). Charge carrier depletion in the n and p regions may play a role in determining the electric field across thin intrinsic regions. Taking this into account while assuming an undoped i region, the electric field across the i region of an ideally abrupt *pin* junction, when the undoped layer d_i is sufficiently large, can be expressed as⁷²

$$E = \frac{V_{\text{tot}}}{d_i} \left[1 - \frac{\epsilon V_{\text{tot}} (1 + N_d/N_a)}{2eN_d d_i^2} \right] \quad (58)$$

where N_d is the (donor) doping density in the n region, N_a is the (acceptor) doping density in the p region, e is the elementary charge, ϵ is the dielectric constant, d_i is the thickness of the intrinsic region, and V_{tot} is the sum of the applied and built-in field (defined positive for reverse bias). This assumes the n and p regions are highly doped and the i region is undoped, so that most of the voltage is dropped across the i region. To lowest order, this is just the field across a capacitor of thickness d_i . For a typical applied voltage of 5 V and $d_i = 0.25 \text{ }\mu\text{m}$, with $N_a = 10^{18} \text{ cm}^{-3}$ and $N_d = 10^{18} \text{ cm}^{-3}$, $E = 2 \times 10^5 \text{ V/cm}$. How much this will change the absorption and refractive index depends on wavelength and, of course, material design.

Operating Characteristics

In addition to contrast ratio, insertion loss, and required voltage, performance of electroabsorption modulators depends on speed, chirp, polarization dependence, optical power-handling capabilities, and linearity. These factors all depend on the wavelength of operation, the materials, the presence of strain, the QW and waveguide geometry, and the device design. Extensive trade-offs must be made to achieve the best possible operation for a given application. Modulators will differ, depending on the laser and the proposed applications.

Chirp Because a change in refractive index usually occurs during any absorption change $\delta\alpha$, electroabsorption modulators, in general, exhibit *chirp* (frequency broadening due to the time-varying refractive index, also observed in modulated lasers), which can seriously limit their usefulness. As with semiconductor lasers, the figure of merit is $\beta = k_a \delta n / \delta\alpha$. Unlike lasers, however, there are particular wavelengths of sizable absorption change at which $\delta n = 0$. Studies have shown that these nulls in index change can be positioned where $\delta\alpha$ is large by using coupled quantum wells (CQWs).⁷⁴ These structures provide two, three, or more wells so closely spaced that the electron wave functions overlap between them. If desired, several sets of these CQWs may be used in a single waveguide, if they are separated by large enough barriers that they do not interact. Chirp-free design is an important aspect of electroabsorption modulators.

On the other hand, since the chirp can be controlled in electroabsorption modulators, there are conditions under which it is advantageous to provide a negative chirp to cancel out the positive chirp introduced by fibers. This allows $1.55\text{-}\mu\text{m}$ laser pulses to travel down normally dispersive fiber (with zero material dispersion at $1.3 \text{ }\mu\text{m}$) without the pulses unduly spreading.

Polarization Dependence In general, the QCSE is strongly polarization dependent, although there may be specific wavelengths at which TE and TM polarized light experience the same values of

electroabsorption (and/or electrorefraction). It turns out that polarization-independent modulation is more readily achieved by using strained QWs. In addition, the contrast ratio of electroabsorption change at long wavelengths can be improved by using strained QWs.

Optical Power Dependence During the process of electroabsorption, the modulators can absorb some of the incident light, creating electron-hole pairs. If these pairs remain in the QWs, at high optical powers they will introduce a free carrier plasma field that can screen the exciton resonance. This broadens the absorption spectrum and reduces the contrast ratio. In some cases, electroabsorption modulators operating at the band edge of bulk semiconductors (the *Franz-Keldysh effect*) may be able to operate with higher laser power. A common approach is to use shallow QWs, so that the electrons and holes may escape easily.

Even when the electron-hole pairs created by absorption escape the QWs, they will move across the junction to screen the applied field. This will tend to reduce the applied field, so the performance will depend on the magnitude of absorbed light. Photogenerated carriers must also be removed, or they will slow down the modulator's response time. Carriers may be removed by leakage currents in the electrodes or by recombination.

Built-in Bias Because *pin* junctions have built-in fields, even at zero applied voltage, electroabsorption modulators have a prebias. Some applications use a small forward bias to achieve even larger modulation depths. However, the large forward current resulting from the forward bias limits the usefulness of this approach. There are, at present, some research approaches to remove the internal fields using an internal strain-induced piezoelectric effect to offset the *pn* junction intrinsic field.

Commercial Discrete Electroabsorption Modulators Discrete electroabsorption modulators can be fast: 30 GHz or more. They can be designed either for low chirp (α can be < 0.5), or low polarization-dependent loss (< 0.5), but not both. Typically they may operate with voltages from -4 to $+1$ V with 20-dB extinction ratio.⁷⁵ They are limited in optical power to 20 mW and require temperature-stabilization with a thermoelectric cooler. Their greatest disadvantage, however, is their large insertion loss (typically 7.5 to 10 dB), because mode profiles from single-mode fibers do not match the electroabsorptive semiconductor waveguides.

One approach to reduce the insertion loss is to use the electroabsorption modulator in reflection (R-EAM). Devices have been reported with 4.5 dB typical insertion loss and 0.5 dB polarization dependent loss. These devices have 11 dB modulation depth with a voltage range from 0 to -3 V, and 14 GHz bandwidth. Other devices operate at 35 GHz with a 1 dB polarization dependent loss and 10 dB contrast ratio.⁷⁶

Integrating the Modulator

Stripe-geometry modulators can be cleaved from a wafer, antireflection coated, and butt-coupled to either a laser or a fiber pigtail. Typical insertion losses may be ~ 10 dB. Or, the modulator may be monolithically integrated with the laser. A portion of the same epitaxial layer grown for the laser's active region can be used as an electroabsorption modulator by providing a separate contact and applying a reverse bias. When such a modulator is placed inside the laser cavity, a multielement laser results that can have interesting switching properties, including wavelength tunability. When the electroabsorption modulator is placed outside the laser cavity, it is necessary to operate an electroabsorption modulator with a higher energy bandgap than the laser medium. Otherwise, the incident light will be absorbed, creating free electron-hole pairs that will move to screen the applied field and ruin the modulator.

Etch and Regrowth Typically, a first set of epitaxial layers is grown everywhere, which includes the laser structure up through the QW layer. Then the QW layer is etched away from the regions where it is not needed. The structure is then overgrown everywhere with the same upper cladding layers.

This typically results in a bulk electroabsorption modulator, consisting of laser-cladding material. A more complex fabrication process might mask the laser region during the regrowth process and grow a different QW composition that would provide an integrated butt-coupled modulator for the DFB (or DBR) QW laser.

Vertical Coupling between Layers This approach makes it possible to use a QW modulator as well as a QW laser, with a different QW composition in each. Two sets of QWs can be grown one on top of the other and the structures can be designed so that light couples vertically from one layer to the other, using, for example, grating-assisted coupling. This may involve photolithographically defining a grating followed by a regrowth of cladding layers, depending on the design.

Selective Area Epitaxy Growth on a patterned substrate allows the width of the QWs to be varied across the wafer during a single growth. The substrate is usually coated with a SiO_2 mask in which slots are opened. Under a precise set of growth conditions no growth takes place on top of the dielectric, but surface migration of the group III species (indium) can take place for some distance across the mask to the nearest opening. The growth rate in the opened area depends on the width of the opening and the patterning on the mask. Another approach is epitaxial growth on faceted mesas, making use of the different surface diffusion lengths of deposited atomic species on different crystal facets.

Well and Barrier Intermixing The bandgap of a QW structure can be modified after growth by intermixing the well and barrier materials to form an alloy. This causes a rounding of the initially square QW bandgap profile and, in general, results in an increase of the bandgap energy. This provides a way to fabricate lasers and bandgap-shifted QCSE modulators using only one epitaxial step. Intermixing is greatly enhanced by the presence of impurities or defects in the vicinity of the QW interfaces. Then the bandgap is modified using impurity-induced disordering, laser beam-induced disordering, impurity-free vacancy diffusion, or ion implantation-enhanced interdiffusion. The challenge is to ensure that the electrical quality of the *pin* junction remains after interdiffusion; sometimes regrowth of a top p layer helps.⁷³

Electroabsorption Modulators Integrated with Lasers

An electroabsorption modulator (EAM) integrated with a DFB or DBR laser is a good choice as a transmitter device in systems because of its compactness and ease of operation. Selective epitaxy enables separate optimization of the laser and the modulator. The laser may have either a DFB or DBR geometry and the EAM is a reverse biased *pin* structure. The two devices may be fabricated in a butt-coupled geometry, or separated and a waveguide grown to connect them.

Electroabsorption-modulated lasers are now commercially available from such companies as Oki, Cyoptics, Mitsubishi, JDS Uniphase, SVEDICE in Sweden, among others. They typically provide modulation to 40 Gb/s, with parameters such as 20-dB extinction ratio, 6-mW optical power, and 10-MHz spectral width with 35-dB sidemode suppression. These devices require thermoelectric coolers that typically dissipate 2 W. The specific designs are usually company-proprietary, but enough operational data is provided that they can be operated in a relatively straightforward manner. The great advantage of the electroabsorption modulator is the severe reduction in chirp, compared to modulating the laser current.

An alternative to selective epitaxy is the use of a localized quantum well intermixing process that has been shown to increase the bandgap. Intermixing can occur by impurity-induced disordering, ion-implantation enhanced intermixing, and impurity-free vacancy diffusion. The latter method involves deposition of a dielectric capping material and subsequent thermal annealing. In GaAs–AlGaAs QW materials, SiO_x induces outdiffusion of Ga during annealing, causing vacancies that enhance the intermixing of Ga and Al in the QWs. The bandgap becomes larger in the QW because of partial disordering of the two materials. In the InGaAs(P)–InP QW materials, SiN is used as a cap.

13.13 ELECTRO-OPTIC AND ELECTROREFRACTIVE MODULATORS

Some semiconductor modulators are based on phase modulation that is converted to amplitude modulation by using a Mach-Zehnder interferometer, in the same manner as discussed in Sec. 13.11. Such modulators can be integrated on the same substrate as the laser, but do not have the chirp issues that electroabsorption modulators exhibit.

Electro-Optic Effect in Semiconductors

The III-V semiconductors are electro-optic. Although not initially anisotropic, they become so when an electric field is applied. Referring to the discussion of the electro-optic effect in Sec. 13.11 for definitions, the GaAs electro-optic coefficients have only one nonzero term: $r_{41} = r_{52} = r_{63} = 1.4 \times 10^{-12}$ m/V. Crystals are typically grown on the (001) face, with the z axis normal to the surface, and the field is usually applied along z . The only electro-optically induced index change will be $\Delta(1/n^2)_4 = r_{41}E_z$, which introduces nondiagonal terms in the (x, y, z) coordinate system. Diagonalizing the matrix results in new axes (x', y', z) that are at 45° to the (x, y, z) crystal axes and new values of the inverse squared refractive index along these axes: $1/n_{x'}^2 = 1/n_o^2 + r_{41}E_z$ and $1/n_{y'}^2 = 1/n_o^2 - r_{41}E_z$.

Differentiation for small refractive change provides refractive index changes for light polarized at 45° to crystal axes:

$$n_{x'} = n_o + \frac{n_o^3}{2} r_{41} E_z \quad \text{and} \quad n_{y'} = n_o - \frac{n_o^3}{2} r_{41} E_z \quad (59)$$

The direction of these new optic axes (45° to the crystal axes) turns out to be in the direction that the zincblende material cleaves. Thus, TE-polarized light traveling down a waveguide normal to a cleave experiences the index change shown in Eq. (59). Depending on whether light goes along x' or y' , the index will increase or decrease.⁷⁷ Light polarized along z will not see any index change.

With an electro-optic coefficient of $r_{41} = 1.4 \times 10^{-10}$ cm/V, in a field of 10 kV/cm (2 V across 2 μ m), and since $n_o = 3.3$, the index change for the TE polarization in GaAs will be 2.5×10^{-5} . The index change in InP-based materials is comparable. The phase shift in a sample of length L is $\Delta n k L$. At 1- μ m wavelength, this will require a sample of length 1 cm to achieve a π phase shift, so that the voltage-length product for electro-optic GaAs (or other semiconductor) will be ~ 20 V-mm. Practical devices require larger refractive index changes, which can be achieved by using the electrorefractive effect in QWs and choosing the exciton resonance at a shorter wavelength than that of the light to be modulated.

Electrorefraction in Semiconductors

Near the band edge in semiconductors, the change in refractive index with applied field can be particularly large, especially in QWs, and is termed *electrorefraction*, or the *electrorefractive effect*. Electrorefraction is calculated from the spectrum of electroabsorption using the Kramers-Kronig relations. The existence of electroabsorption means there will inevitably be electrorefraction at wavelengths below the band edge. Electrorefraction may be larger than the electro-optic effect and may significantly reduce the length and drive voltages required for phase modulation in semiconductor waveguides. The voltage-length product depends on how close to the absorption resonance the modulator is operating. It also depends on device design. As with electroabsorption modulators, the field is usually applied across a *pin* junction. Some reported π voltage-length products are 2.3 V-mm in GaAs/AlGaAs QWs (at 25-V bias), 1.8 V-mm in InGaAs/InAlAs QWs, and the same in GaAs/AlGaAs double heterostructures.⁷⁸ These voltage-length products depend on wavelength detuning from the exciton resonance and therefore on insertion and electroabsorption losses. The larger the voltage-length product, the greater the loss.

Typical Performance Electrorefraction is polarization dependent, because the QCSE is polarization dependent. In addition, the TE polarization experiences the electro-optic effect, which may add to or subtract from the electrorefractive effect, depending on the crystal orientation. Typically, Δn for TE polarization (in an orientation that sums these effects) will be 8×10^{-4} at 82 kV/cm (7 V across a waveguide with an i layer 0.85- μm thick). Of this, the contribution from the electro-optic effect is 2×10^{-4} . Thus, electrorefraction is about 4 times larger than the electro-optic effect and the voltage-length product will be reduced by a factor of 5. Of course, this ratio depends on the field, since the electro-optic effect is linear in field and electrorefraction is quadratic in the field. This ratio also depends on wavelength; electrorefraction can be larger at wavelengths closer to the exciton resonance, but the residual losses go up. The TE polarization of electrorefractive modulators integrates well with a TE-polarized laser, and they can be grown on the same substrate.

The TM polarization, which experiences electrorefraction alone, will be 5×10^{-4} at the same field, slightly smaller than the TE electrorefraction, because QCSE is smaller for TM than TE polarization.⁷⁹

Polarization-independent electrorefraction modulators have been demonstrated using suitably strained quantum wells.

Advanced QW Concepts Compressive strain increases electrorefraction, as it does QCSE. Measurements at the same 82 kV/cm show an increase from 2.5×10^{-4} to 7.5×10^{-4} by increasing compressive strain.⁸⁰ Strained QWs also make it possible to achieve polarization-independent electrorefractive modulators (although when integrated with a semiconductor laser, which typically has a well-defined polarization, this should not be a necessity).

Advanced QW designs have the potential to increase the refractive index change below the exciton resonance. One example analyzes asymmetric coupled QWs and finds more than 10 times enhancement in Δn below the band edge, at least at small biases. However, when fabricated and incorporated into Mach-Zehnder modulators, the complex three-well structure lowered V_π by only a factor of 3, attributed to the growth challenges of these structures.⁸¹

Nipi Modulators One way to obtain a particularly low voltage-length product is to place MQWs in a hetero-*nipi* waveguide. These structures incorporate multiple *pin* junctions (alternating *n-i-p-i-n-i-p*) and include QWs in each i layer. Selective contacts to each electrode are required, which limits how fast the modulator can be switched. A voltage-length product of 0.8 V·mm was observed at a wavelength 115 meV below the exciton resonance. The lowest voltage InGaAs modulator had $V_\pi = 0.5$ V, at speeds up to 110 MHz. Faster speeds require shorter devices and higher voltages.⁷⁸

Band-Filling Modulators When one operates sufficiently far from the band edge so that the absorption is not large, then the electrorefractive effect is only 2 to 3 times larger than the bulk electro-optic effect. This is because oscillations in the change in absorption with wavelength tend to cancel out their contributions to the change in refractive index at long wavelengths. By contrast, during band filling, the long-wavelength refractive index change is much larger, because band filling decreases the absorption at all wavelengths. However, because band filling relies on real carriers, it lasts as long as the carriers do, and it is important to find ways to remove these carriers to achieve high-speed operation.

Voltage-controlled transfer of electrons into and out of QWs (BRAQWET modulator) can yield large electrorefraction by band filling. The refractive index change at 1.55 μm can be as large as $\Delta n = 0.02$ for 6 V. One structure consists of 12 repeating elements,⁸² with the single QW replaced by three closely spaced strongly coupled QWs, demonstrating $V_\pi L = 3.2$ V·mm with negligible loss. Researchers have been looking at these QW devices, and others driven by heterostructure buried transistors (HBTs) that drive current in and out of the QWs. None seems to have reached practicality by this time, however.

An important direction for bandfilling nonlinearities is the holy grail of all-silicon photonics. In silicon the refractive index changes because of electron-hole plasma injection. One approach is to integrate a SiGe HBT together with silicon waveguide. Including a distributed Bragg reflector can convert the index change into a transmission change. The use of SiGe quantum wells enables the quantum-confined Stark effect to be used, leading to larger electrorefraction than from band-filling alone.

Semiconductor Interferometric Modulators

The issues for Mach-Zehnder modulators fabricated in semiconductors are similar to those for modulators in lithium niobate, but the design and fabrication processes in semiconductors are by no means as well developed. Fabrication tolerances, polarization dependence, interaction with lasers, and operation at high optical input powers are just some of the issues that need to be addressed. While the power splitters in the interferometer can be Y branches, fabricated by etching to form ridge waveguides, they are more usually formed by *multimode interference* (MMI) power splitters, discussed previously.

One example reports a Mach-Zehnder interferometer at 1.55 μm in InGaAlAs QWs with InAlAs barriers.⁸³ A polarization-independent extinction ratio of 30 dB was reported, over a 20-nm wavelength range without degradation at input powers of 18 dBm (63 mW). The interferometer phase-shifting region was 1000- μm long, and each MMI was 200- μm long. The insertion loss of 13 dB was due to the mismatch between the mode of the single-mode optical fiber and of the semiconductor waveguide, which was 2- μm wide and 3.5- μm high. Various semiconductor structures to convert spot size should bring this coupling loss down.

13.14 PIN DIODES

The detectors used for fiber optic communication systems are usually *pin photodiodes*. In high-sensitivity applications, such as long-distance systems operating at 1.55- μm wavelength, *avalanche photodiodes* are sometimes used because they have internal gain. Occasionally, *metal-semiconductor-metal* (MSM) *photoconductive detectors* with interdigitated electrode geometry are used because of ease of fabrication and integration. For the highest-speed applications, *Schottky photodiodes* may be chosen. This section reviews properties of *pin* photodiodes. The next section outlines the other photodetectors.

The material of choice for these photodiodes depends on the wavelength at which they will be operated. The short-wavelength *pin* silicon photodiode is perfectly suited for GaAs wavelengths (850 nm); these inexpensive detectors are paired with GaAs/AlGaAs LEDs for low-cost data communication applications. They are also used in the shorter-wavelength plastic fiber applications at 650 nm. The longer-wavelength telecommunication systems at 1.3 and 1.55 μm require longer-wavelength detectors. These are typically *pin* diodes composed of lattice-matched ternary $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$ grown on InP. Silicon is an indirect bandgap semiconductor while InGaAs is a direct-band material; this affects each material's absorption properties and therefore its photodiode design. In particular, silicon photodiodes tend to be slower than those made of GaAs or InGaAs, because silicon intrinsic regions must be thicker. Carriers absorbed in the *n* region that diffuse into the *i* layer also slow silicon *pin* detectors down because diffusion lengths in silicon are much longer than in GaAs. Speeds are also determined by carrier mobilities, which are lower in silicon than in the III-V materials.

Previous volumes in this *Handbook* have outlined the concepts behind the photodetectors discussed here. Chapter 24 in Vol. II places *pin* photodetectors in context with other detectors, and gives specific characteristics of some commercially available detectors, allowing direct comparison of silicon, InGaAsP, and germanium detectors. Chapter 25 in Vol. II describes the principles by which the *pin* and the avalanche photodiodes operate. The properties of greatest interest to fiber communications will be repeated here. Chapter 26 in Vol. II concentrates on high-speed photodetectors and provides particularly useful information on their design for high-speed applications.

The *pin* junction consists of a thin, heavily doped *n* region (called n^+), a near-intrinsic *n* region (the *i* region), and a heavily doped *p* region (called p^+). An incident photon with energy greater than or equal to the bandgap of the semiconductor generates electron-hole pairs. In a well-designed photodiode, this generation takes place in the space-charge region of the *pn* junction. As a result of the electric field in this region, the electrons and holes separate and drift in opposite directions, causing current to flow in the external circuit. This current is monitored as a change in voltage across a load resistor. The *pin* photodiode is the workhorse of fiber communication systems.

Typical Geometry

Typically, the electric field is applied across the pn junction and photocarriers are collected across the diode. A typical geometry for a silicon photodiode is shown in Fig. 36a. A pn junction is formed by a thin p^+ diffusion into a lightly doped n^- or i layer (also called the i layer since it is almost *intrinsic*) through a window in a protective SiO_2 film. Then super minus or i layer i layer between the p^+ and n^+ regions supports a *space-charge region*, which, in the dark, is depleted of free carriers (this is sometimes called the *depletion region*) and supports the voltage drop that results from the pn junction. When light is absorbed in this space-charge region, the absorption process creates electron-hole pairs that separate in the electric field (field lines are shown in Fig. 36a), the electrons falling down the potential hill to the n region and the holes moving to the p region. This separation of charge produces a current in the external circuit, which is read out as a measure of the light level. Free carriers generated within a diffusion length of the junction may diffuse into the space charge region, adding to the measured current.

Long-wavelength detectors utilize n^- or i layers that are grown with a composition that will absorb efficiently in the wavelength region of interest. The ternary $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$ can be grown lattice-matched to InP and has a spectral response that is suitable for both the 1.3- and 1.55- μm wavelength regions. Thus, this ternary is usually the material of choice, rather than the more difficult to grow quaternary InGaAsP, although the latter provides more opportunity to tune the wavelength response. Figure 36b shows a typical geometry. Epitaxial growth provides lightly doped material on a heavily doped substrate. The InP buffer layer is grown to keep the dopants from diffusing into the lightly doped absorbing InGaAs layer. The required thin p region is formed by diffusion through a silicon nitride insulating window. Because InP is transparent to 1.3- and 1.55- μm light, the photodiode can be back-illuminated, which makes electrical contacting convenient. In some embodiments, a well is etched in the substrate and the fiber is glued in place just below the photosensitive region.

Carriers generated outside the space-charge region may enter into the junction by diffusion, and can introduce considerable delay time. In silicon, the diffusion length is as long as 1 cm, so photocarriers generated anywhere within the silicon photodiode can contribute to the photocurrent. Because the diffusion velocity is much slower than the transit time across the space-charge region, diffusion currents slow down silicon photodiodes. This is particularly true in pn diodes. Thus, high-speed applications typically use pin diodes with a narrower bandgap in the i layer, limiting absorption to this region.

To minimize diffusion from the p^+ entrance region, the junction should be formed very close to the surface. The space-charge region should be sufficiently thick so that most of the light will be absorbed

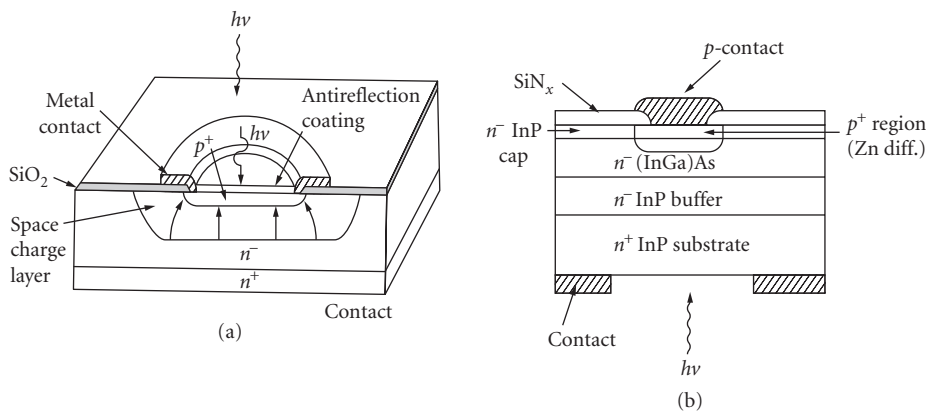


FIGURE 36 Geometry for pin photodiodes: (a) cut-away of silicon, illuminated from the top, showing the ring electrode and static electric field lines in the space-charge region and (b) cross-section of InGaAs/InP, illuminated from the bottom. The p^+ region is formed by diffusion. The low-doped n -layer is the i or nearly intrinsic layer.

there (thickness $\approx 1/\alpha$). With sufficient reverse bias, carriers will drift at their scattering-limited velocity. The space-charge layer must not be too thick, however, or transit-time will limit the frequency response. Neither should it be too thin, or excessive capacitance will result in a large RC time constant. The optimum compromise occurs when the modulation period is on the order of twice the transit time. For example, for a modulation frequency of 10 GHz, the optimum i layer thickness in silicon is about 5 μm . However, this is not enough thickness to absorb more than ~ 50 percent of the light at 850-nm wavelength. Thus, there is a trade-off between sensitivity and speed. If the wavelength drops to 980 nm, only 10 percent of the light is absorbed in a 10- μm thick i layer of silicon.

The doping must be sufficiently small so that the i region can support the voltage drop of the built-in voltage V_{bi} plus the applied voltage. When the doping density of the p^+ region is much higher than the doping density of the i layer (actually lightly doped with N_D donor density), the thickness of the space-charge layer is

$$W_s = \sqrt{\frac{2\epsilon_s(V_{bi} - V)}{e N_D}} \quad (60)$$

To achieve space-charge layer thickness $W_s = 10 \mu\text{m}$ in a silicon photodiode with 10 V applied requires an extremely pure i layer with doping density $N_D \approx 10^{14} \text{ cm}^{-3}$. If the doping is not this low, the voltage drops more rapidly, and the field will not extend fully across the low-doped region i region.

GaAs photodiodes are faster than silicon, because their diffusion length in the highly doped n region is only $\sim 100 \mu\text{m}$. Also, their transit time across the i layer is faster than the transit time in silicon. The fastest diodes have transparent n^+ and p^+ regions, such as InGaAs/GaAs or InGaAs/InP photodiodes (also GaAs/AlGaAs photodiodes). These diodes have highly doped n and p regions with wider bandgap material than the i layer, so they contribute no photocurrent. The thickness of the i layer is chosen thin enough to achieve the desired speed (trading off transit time and capacitance), with a possible sacrifice of sensitivity.

Typically, light makes a single pass through the active layer. In silicon photodiodes, the light usually enters through the p contact at the surface of the diode (Fig. 36a). The top metal contact must have a window for light to enter (or be a transparent contact, such as indium tin oxide). The InGaAs photodiodes may receive light from the p side or the n side, because neither is absorbing. In addition, some back-illuminated devices use a double pass, reflecting off a mirrored top surface, to double the absorbing length. Some more advanced detectors, *resonant photodiodes*, use integrally grown Fabry-Perot cavities (using DBR mirrors, as in VCSELs) that resonantly reflect the light back and forth across the i region, enhancing the quantum efficiency. These are typically used only at the highest bandwidths (> 20 GHz) or for wavelength division multiplexing (WDM) applications, where wavelength-selective photodetection is required. In addition, photodiodes designed for integration with other components are illuminated through a waveguide in the plane of the pn junction. The reader is directed to Chap. 26, Vol. II to obtain more information on these advanced geometries.

Sensitivity (Responsivity)

To operate a pin photodiode, it is sufficient to place a load resistor between ground and the n side and apply reverse voltage to the p side ($V < 0$). The photocurrent is monitored as a voltage drop across this load resistor. The photodiode current in the presence of an optical signal of power P_s is negative, with a magnitude given by

$$I = \eta_D \left(\frac{e}{h\nu} \right) P_s + I_D \quad (61)$$

where I_D is the magnitude of the (negative) current measured in the dark. The detector *quantum efficiency* η_D (electron-hole pairs detected per photon) is determined by the optical transmission of

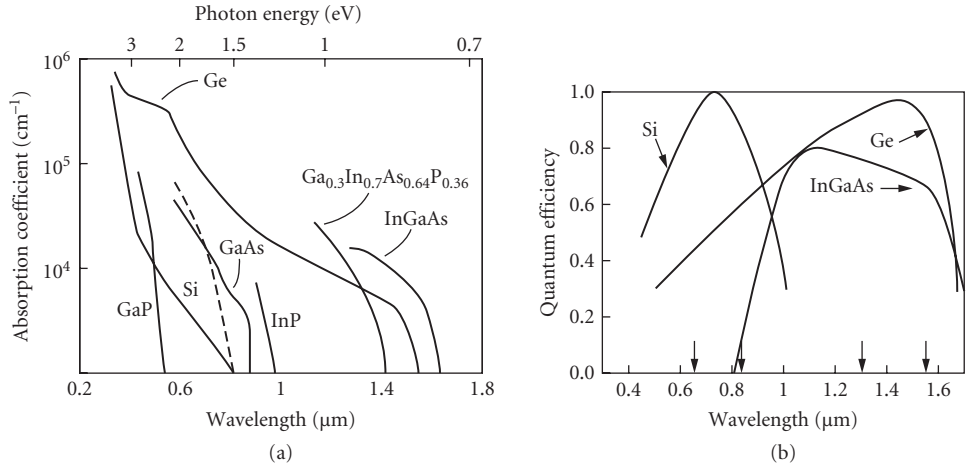


FIGURE 37 (a) Absorption coefficient as a function of wavelength for several semiconductors used in *pin* diode detectors and (b) spectral response of typical photodetectors.

the top electrode T , by the reflection R of light from the top surface of the photodiode (which can be reduced by adding an antireflective coating), and by how much light is absorbed in the region of thickness W , that is, at or within a diffusion length of the i layer (which depends on the absorption coefficient α). The detector quantum efficiency can be expressed as $\eta_D = (1 - R)T [1 - \exp(-\alpha W)]$.

The *sensitivity* (or *responsivity* $\mathcal{R}$) of a detector is the ratio of milliamperes of current out per milliwatt of light in. Thus, the responsivity is

$$\mathcal{R} = \frac{I_{PD}}{P_s} = \eta_D \frac{e}{h\nu} \quad (62)$$

For detection of a given wavelength, the photodiode material must be chosen with a bandgap sufficient to provide suitable sensitivity. The absorption spectra and quantum efficiency of candidate detector materials are shown in Fig. 37. Silicon photodiodes provide low-cost detectors for most data communications applications, with acceptable sensitivity at 850 nm (absorption coefficient $\sim 500 \text{ cm}^{-1}$). These detectors work well with the GaAs lasers and LEDs that are used in the inexpensive datacom systems and for short-distance or low-bandwidth local area network (LAN) applications. GaAs detectors are faster, both because their absorption can be larger and because their electron mobility is higher, but they are more expensive. Systems that require longer-wavelength InGaAsP/InP lasers typically use InGaAs photodiodes. Germanium has a larger dark current, so it is not usually employed for optical communications applications. Essentially all commercial photodetectors use bulk material, not quantum wells, as these are simpler, are less wavelength sensitive, and have comparable performance. Table 1 gives the sensitivity of typical detectors of interest in fiber communications, measured in units of amperes per watt, along with speed and relative dark current.

TABLE 1 Characteristics of Typical Photodiodes

	Wavelength (nm)	Sensitivity $\mathcal{R}$ (A/W)	Speed t (ns)	Dark Current (Normalized Units)
Silicon	850	0.55	3	1
	650	0.4	3	
GaInAs	1300–1600	0.95	0.2	3
Ge (<i>pn</i>)	1550	0.9	3	66

Speed

Contributions to the speed of a *pin* diode come from the transit time across the space-charge region and from the RC time constant of the diode circuit in the presence of a load resistor R_L . Finally, in silicon there may be a contribution from the diffusion of carriers generated in undepleted regions.

The transit time across the space-charge region depends on its thickness. When the photodiode is properly operated, the space-charge region should extend fully across the i layer, which has a thickness W_i . Equation (59) gives the thickness of the space-charge region W_s , as long as it is less than the thickness of the i layer W_i . Define V_i as that voltage at which $W_s = W_i$; then $-V_i = W_i^2 e N_D / 2 \epsilon_s - V_{bi}$. For any voltage with magnitude larger than this, the space-charge width is essentially W_i (since the space charge extends a negligible distance into highly doped regions).

If the electric field across the space-charge region is high enough for the carriers to reach their saturation velocity v_s and high enough to fully deplete the i region, then the carrier transit time will be $\tau_i = W_i / v_s$. For $v_s = 10^7$ cm/s and $W_i = 4$ μ m, the transit time $\tau_i = 40$ ps. A finite transit time τ_i reduces the response at modulation frequency ω , such that:⁸⁴

$$\Re(\omega) = \Re_o \frac{\sin(\omega \tau_i / 2)}{\omega \tau_i / 2} \quad (63)$$

Defining the 3-dB bandwidth as that modulation frequency at which the electrical power decreases by 50 percent, it can be shown that the transit-limited 3-dB bandwidth is $\delta\omega_i = 2.8 / \tau_i = 2.8 v_s / W_i$. (Because electrical power is proportional to I^2 and $\Re^2$, the half-power point is achieved when the current is reduced by $1/\sqrt{2}$.) There is a trade-off between sensitivity and transit time, since, for thin layers the quantum efficiency is $\eta_D \approx (1 - R) T \alpha W_i$. Thus, the quantum efficiency–bandwidth product is: $\eta_D \delta\omega_i \approx 2.8 \alpha v_s (1 - R) T$. When the top electrode has no loss or reflection, this product depends only on the semiconductor's absorption loss and saturation velocity, not the geometry.

The speed of a *pin* photodiode is also limited by its capacitance, through the RC of the load resistor. Sandwiching an intrinsic layer, which is depleted of carriers, between conductive n and p regions causes a diode capacitance proportional to the detector area A : $C_D = \epsilon_s A / W_i$.

For a given load resistance, the smaller the area, the smaller the RC time constant, and the higher the speed. We will see also that the dark current I_i decreases as the detector area decreases. The detector should be as small as possible, as long as all the light from the fiber can be collected onto the detector. Multimode fibers easily butt-couple to detectors whose area matches the fiber core size. High-speed detectors compatible with single-mode fibers can be extremely small, which increases the alignment difficulty, so typically high-speed photodetectors are already pigtailed to single-mode fiber. A low load resistance may be needed to keep the RC time constant small, which may result in a small signal that needs amplification. Speeds in excess of 1 GHz are straightforward to achieve, and speeds of 50 GHz are not uncommon.

Thicker space-charge regions provide smaller capacitance, but too thick a space-charge region causes the speed to be limited by the carrier transit time. The bandwidth with a load resistor R_L is given by

$$\omega_{3\text{ dB}} = \frac{2.8}{\tau_i} + \frac{1}{R_L C} = \frac{2.8 v_s}{W_i} + \frac{W_i}{\omega_s A R_L} \quad (64)$$

There is an optimum thickness W_i for high-speed operation. Any additional series resistance R_s or parasitic capacitance C_p must be added by using $R \rightarrow R_L + R_s$ and $C \rightarrow C + C_p$. The external connections to the photodetector can also limit speed. The gold-bonding wire may provide additional series inductance. It is important to realize that the photodiode is a high impedance load, with very high electrical reflection, so that an appropriate load resistor must be used. As pointed out in Chap. 26 in Vol. II, it is possible to integrate a matching load resistor inside the photodiode device, with a reduction in sensitivity of a factor of 2 (since half the photocurrent goes through the load resistor), but double the speed (since the RC time constant is halved). A second challenge is to build external bias circuits without high-frequency electrical resonances. Innovative design of the photodetector may integrate the necessary bias capacitor and load resistor, ensuring smooth electrical response.

Silicon photodetectors are inherently not as fast because of diffusion into the i layer of carriers absorbed in the highly doped p and n regions. Their photoresponse has a component with a slower response time governed by the carrier diffusion time: $T_D = W_D^2/2D$, where W_D is the width of the absorbing doped region, and D is the diffusion constant for whichever carrier is dominant (usually holes in the n region). For silicon, $D = 12 \text{ cm}^2/\text{s}$, so that when $W_D = 100 \text{ }\mu\text{m}$, $\tau_D = 400 \text{ ns}$.

Unitraveling-Carrier Photodiode

In conventional high-speed pin photodiodes, high optical power can saturate the current output. The cause is the space charge due to photogenerated holes, which introduces a variation of the electric field in the depletion region. This space charge effect can be relieved by using a unitraveling-carrier (UTC) design, shown in Fig. 38.

This $p^+n^-n^+$ structure utilizes only electrons as the active carriers, offering both high-speed and high-output simultaneously. The UTC structure separates the p^+ absorption layer from the depleted n^- collector layer, which is transparent because it has a wider bandgap. In the narrow bandgap p^+ photoabsorption layer, electrons and holes are generated by photoexcitation. Minority carrier electrons diffuse toward the collector, where the depletion field causes them to drift across the collector at high-speed. Because the electrical field vanishes in the p^+ absorbing layer, intense illumination has, in principle, no effect on the responsivity and dynamics of the photodetector. The speed of the UTC structure is determined only by the electron transit times in the photoabsorption and collector layers; the maximum available output voltage is much higher than that of the conventional structure. The price to pay is a low responsivity for top-illuminated devices, since the p^+ absorbing layer must be thin for fast collection of electrons.

These high-speed photodetectors have accelerated the progress of optical communication systems and measurement systems, reaching data rates as much as 40 Gb/s in long-haul transmission systems. These devices simplify and improve the performance of high-speed receivers, offering reliability and stability. The 3-dB bandwidth has been reported as high as 310 GHz; photoreceivers of up to 160 Gb/s have been reported.⁸⁵

When the photodetection mechanism takes place along the length of a waveguide, the field screening due to intense optical fields can be much less. An optical waveguide can also increase the absorbance of UTC detectors. Waveguide devices use either *edge-coupling* or *evanescent-wave coupling*. The challenge is for the optical waveguide to match the light's spatial profile from the fiber and

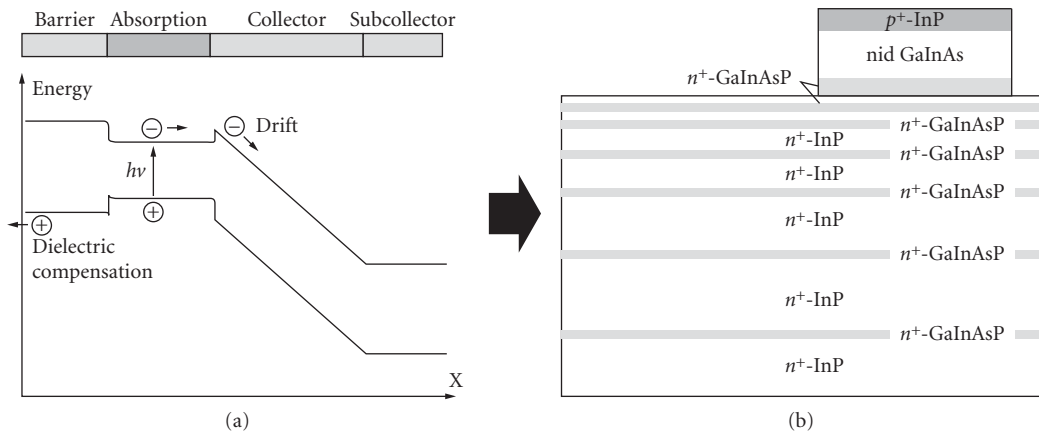


FIGURE 38 (a) Band structure of an UTC photodiode. Only electrons pass through the collector. (b) Multimode evanescent coupled PIN waveguide.

for this light to be coupled effectively into the absorption region of the detector. A variety of methods have been explored, but it is unclear that any design has found universal acceptance. One interesting approach developed at Alcatel is shown in Fig. 38b. The graded layer structure causes incident light to undertake an oscillatory path and be “focused” to the UTC $p^+n^-n^+$ structure, where it is evanescently coupled through the transparent n^+ GaInAsP and GaInAs collecting layers into the absorbing InP layer.⁸⁶

Dark Current

Semiconductor diodes can pass current even in the dark, giving rise to *dark current* that provides a background present in any measurement. This current comes primarily from the thermally generated diffusion of minority carriers out of the n and p regions into the depleted junction region, where they recombine. The current-voltage equation for a pn diode (in the dark) is

$$I = I_s \left[\exp\left(\frac{eV}{\beta kT}\right) - 1 \right] \quad (65)$$

where I_s is the *saturation current* that flows at large back bias (V large and negative). This equation represents the current that passes through any biased pn junction. Photodiodes use pn junctions reverse biased ($V < 0$) to avoid large leakage current.

Here β is the *ideality factor*, which varies from 1 to 2, depending on the diode structure. In a metal-semiconductor junction (Schottky barrier) or an ideal pn junction in which the only current in the dark is due to minority carriers that diffuse from the p and n regions, then $\beta = 1$. However, if there is thermal generation and recombination of carriers in the space-charge region, then β tends toward the value 2. This is more likely to occur in long-wavelength detectors.

The saturation current I_s is proportional to the area A of the diode in an ideal junction: $I_s = e(D_p p_{n0}/L_p + D_n n_{p0}/L_n)A$, where D_n , D_p are diffusion constants, L_n , L_p are diffusion lengths, and n_{p0} , p_{n0} are equilibrium minority carrier densities, all of electrons and holes, respectively. The saturation current I_s can be related to the diode resistance at $V = 0$, measured in the dark through $1/R_0 = -(dI/dV)|_{V=0}$ to give $R_0 = \beta kT/eI_s$. This implies that the dark resistance is inversely proportional to the diode area.

The diffusion current through the diode in Eq. (64) has two components that are of opposite sign in a forward-biased diode: a forward current $I_s \exp(eV/\beta kT)$ and a backward current $-I_s$. Each of these components is statistically independent, coming from diffusive contributions to the forward current and backward current, respectively. This fact is important in understanding the noise properties of photodiodes.

In photodiodes under reverse bias, $V \leq 0$, so $\exp(eV/\beta kT) < 1$; both currents are negative and add. Negative dark current has a direction that is opposite to the current flow in a forward-biased diode. Holes move toward the p region and electrons move toward the n region, the same as photogenerated carriers. These dark currents are thermally generated. Assuming $e|V| \gg kT$, the dark current becomes $I_D = I_s \approx \beta kT/eR_0$. The dark current increases linearly with temperature and is independent of (large enough) reverse bias. *Trap-assisted* thermal generation current increases β ; in this process, carriers trapped in impurity levels can be thermally elevated to the conduction band. The temperature of photodiodes should be kept moderate in order to avoid excess dark current.

When light is present in a reverse-biased photodiode with $V \equiv -V'$, the photocurrent is negative, moving in the direction of the applied voltage, and adding to the negative dark current. The net effect of carrier motion will be to tend to screen the internal field. Defining the magnitude of the photocurrent as $I_{PC} = \eta_D(e/h\nu)P_s$, then the total current is negative:

$$I = -[I_D + I_{PC}] = -I_s \left[1 - \exp\left(\frac{-eV'}{\beta kT}\right) \right] - I_{PC} \quad (66)$$

Noise in Photodiodes

Successful fiber optic communication systems depend on a large signal-to-noise ratio. This requires photodiodes with high sensitivity and low noise. Background noise comes from shot noise due to the discrete process of photon detection, from thermal processes in the load resistor (Johnson noise), and from generation-recombination noise due to carriers within the semiconductor. When used with a field-effect transistor (FET) amplifier, there will also be shot noise from the amplifier and $1/f$ noise in the drain current.

Shot Noise *Shot noise* is fundamental to all photodiodes and is due to the discrete nature of the conversion of photons to free carriers. The shot noise current is a statistical process. If N photons are detected in a time interval Δt , Poisson noise statistics cause the uncertainty in N to be $\sqrt{N}$. Using the fact that N electron-hole pairs create a current I through $I = eN/\Delta t$, then the signal-to-noise ratio is $N/\sqrt{N} = \sqrt{N} = \sqrt{(I\Delta t/e)}$. Writing the frequency bandwidth Δf in terms of the time interval through $\Delta f = 1/(2\Delta t)$, the signal-to-noise ratio is: $\text{SNR} = (I/2e\Delta f)^{1/2}$. The root mean square (rms) photon noise, given by $\sqrt{N}$, creates an rms shot noise current of $i_{\text{SH}} = e(N/\Delta t)^{1/2} = (eI/\Delta t)^{1/2} = (2eI\Delta f)^{1/2}$.

Shot noise depends on the average current I ; therefore, for a given photodiode, it depends on the details of the current-voltage characteristic. Expressed in terms of the photocurrent I_{PC} or the optical signal power P_s (when the dark current is small enough to be neglected) and the responsivity (or sensitivity) $\mathcal{R}$, the rms shot noise current is

$$i_{\text{SH}} = \sqrt{2eI_{\text{PC}}\Delta f} = \sqrt{2e\mathcal{R}P_s\Delta f} \quad (67)$$

The shot noise can be expressed directly in terms of the properties of the diode when all sources of noise are included. Since they are statistically independent, the contributions to the noise currents will be additive. Noise currents can exist in both the forward and backward directions, and these contributions must add, along with the photocurrent contribution. The entire noise current squared becomes

$$i_N^2 = 2e \left\{ I_{\text{PC}} + \left(\frac{\beta kT}{eR_0} \right) \left[1 + \exp\left(\frac{-eV'}{\beta kT} \right) \right] \right\} \Delta f \quad (68)$$

Clearly, noise is reduced by increasing the reverse bias. When the voltage is large, the shot noise current squared becomes $i_N^2 = 2e[I_{\text{PC}} + I_D]\Delta f$. The dark current adds linearly to the photocurrent in calculating the shot noise.

Thermal (Johnson) Noise In addition to shot noise due to the random variations in the detection process, the random thermal motion of charge carriers contributes to a *thermal noise* current, often called *Johnson* or *Nyquist noise*. It can be calculated by assuming thermal equilibrium with $V = 0$, $\beta = 1$, so that Eq. (67) becomes

$$i_{\text{th}}^2 = 4 \left(\frac{kT}{R_0} \right) \Delta f \quad (69)$$

This is just *thermal* or *Johnson noise* in the resistance of the diode. The noise appears as a fluctuating voltage, independent of bias level.

Johnson Noise from External Circuit An additional noise component will be from the load resistor R_L and resistance from the input to the preamplifier, R_i :

$$i_{\text{NJ}}^2 = 4kT \left(\frac{1}{R_L} + \frac{1}{R_i} \right) \Delta f \quad (70)$$

Note that the resistances add in parallel as they contribute to noise current.

Noise Equivalent Power The ability to detect a signal requires having a photocurrent equal to or higher than the noise current. The amount of noise that detectors produce is often characterized by the *noise equivalent power* (NEP), which is the amount of optical power required to produce a photocurrent just equal to the noise current. Define the noise equivalent photocurrent I_{NE} , which is set equal to the noise current i_{SH} . When the dark current is negligible, the *noise equivalent photocurrent* is $i_{SH} = \sqrt{2eI_{NE}\Delta f} = I_{NE}$.

Thus, the noise equivalent current is $I_{NE} = 2e\Delta f$, and depends only on the bandwidth Δf . The noise equivalent power can now be expressed in terms of the noise equivalent photo current:

$$NEP = \frac{I_{NE}}{\eta} \frac{h\nu}{e} = 2 \frac{h\nu}{\eta} \Delta f \quad (71)$$

The second equality assumes the absence of dark current. In this case, the NEP can be decreased only by increasing the quantum efficiency (for a fixed bandwidth). In terms of sensitivity (amperes per watt): $NEP = 2(e/\mathcal{R})\Delta f = I_{NE}\Delta f$. This expression is usually valid for photodetectors used in optical communication systems, which have small dark currents.

If dark current dominates, $i_N = \sqrt{2eI_D\Delta f}$, and

$$NEP = \sqrt{\frac{2I_D\Delta f}{e}} \frac{h\nu}{\eta} \quad (72)$$

This is often the case in infrared detectors such as germanium. Note that the dark-current-limited noise equivalent power is proportional to the square root of the area of the detector, because the dark current is proportional to the detector area. The NEP is also proportional to the square root of the bandwidth Δf . Thus, in photodetectors whose noise is dominated by dark current, NEP divided by the square root of area times bandwidth should be a constant. The inverse of this quantity has been called the *detectivity* D^* and is often used to describe infrared detectors. In photodiodes used for communications, dark current usually does not dominate and it is better to use Eq. (70), an expression which is independent of area, but depends linearly on bandwidth.

13.15 AVALANCHE PHOTODIODES, MSM DETECTORS, AND SCHOTTKY DIODES

The majority of optical communication systems use photodiodes, sometimes integrated with a preamplifier. Avalanche photodiodes offer an alternative way to create gain. Other detectors sometimes used are low-cost MSM detectors or ultrahigh-speed Schottky diodes. Systems decisions, such as signal-to-noise, cost, and reliability will dictate the choice.

Avalanche Detectors

When large voltages are applied to photodiodes, the avalanche process produces gain, but at the cost of excess noise and slower speed. In fiber telecommunication applications, where speed and signal-to-noise are of the essence, avalanche photodiodes (APDs) are frequently at a disadvantage. Nonetheless, in long-haul systems at 2488 Mb/s, APDs may provide up to 10 dB greater sensitivity in receivers limited by amplifier noise. While APDs are inherently complex and costly to manufacture, they are less expensive than optical amplifiers and may be used when signals are weak.

Gain (Multiplication) When a diode is subject to a high reverse-bias field, the process of impact ionization makes it possible for a single electron to gain sufficient kinetic energy to knock another electron from the valence to the conduction band, creating another electron-hole pair. This enables the quantum efficiency to be >1 . This internal multiplication of photocurrent could be compared to the gain in photomultiplier tubes. The *gain* (or *multiplication*) M of an APD is the ratio of the

photocurrent divided by that which would give unity quantum efficiency. Multiplication comes with a penalty of an excess noise factor, which multiplies shot noise. This excess noise is function of both the gain and the ratio of impact ionization rates between electrons and holes.

Phenomenologically, the low-frequency multiplication factor is

$$M_{\text{DC}} = \frac{1}{1 - (V/V_B)^n} \quad (73)$$

where the parameter n varies between 3 and 6, depending on the semiconductor, and V_B is the breakdown voltage. Gains of $M > 100$ can be achieved in silicon APDs, while they are more typically 10 to 20 for longer-wavelength detectors, before multiplied noise begins to exceed multiplied signal. A typical voltage will be 75 V in InGaAs APDs, while in silicon it can be 400 V.

The avalanche process involves using an electric field high enough to cause carriers to gain enough energy to accelerate them into ionizing collisions with the lattice, producing electron-hole pairs. Then, both the original carriers and the newly generated carriers can be accelerated to produce further ionizing collisions. The result is an avalanche process.

In an intrinsic i layer (where the electric field is uniform) of width W_i , the gain relates to the fundamental avalanche process through $M = 1/(1 - aW_i)$, where a is the *impact ionization coefficient*, which is the number of ionizing collisions per unit length. When $aW_i \rightarrow 1$, the gain becomes infinite and the diode breaks down. This means that avalanche multiplication appears in the regime before the probability of an ionizing collision is 100 percent. The gain is a strong function of voltage, and these diodes must be used very carefully. The total current will be the sum of avalanching electron current and avalanching hole current.

In most *pin* diodes the i region is really low n -doped. This means that the field is not exactly constant, and an integration of the avalanche process across the layer must be performed to determine a . The result depends on the relative ionization coefficients; in III-V materials they are approximately equal. In this case, aW_i is just the integral of the ionizing coefficient that varies rapidly with electric field.

Separate Absorber and Multiplication APDs In this design the long-wavelength infrared light is absorbed in an intrinsic narrow-bandgap InGaAs layer, and photocarriers move to a separate, more highly n -doped InP layer that supports a much higher field. This layer is designed to provide avalanche gain in a separate region without excessive dark currents from tunneling processes. This layer typically contains the *pn* junction, which traditionally has been diffused. Fabrication procedures such as etching a mesa, burying it, and introducing a guard ring electrode are all required to reduce noise and dark current. All-epitaxial structures provide low-cost batch-processed devices with high performance characteristics.⁸⁷

Speed When the gain is low, the speed is limited by the RC time constant. As the gain increases, the avalanche buildup time limits the speed, and for modulated signals the multiplication factor decreases. The multiplication factor as a function of modulation frequency is

$$M(\omega) = \frac{M_{\text{DC}}}{\sqrt{1 + M_{\text{DC}}^2 \omega^2 \tau_1^2}} \quad (74)$$

where $\tau_1 = p\tau$, with τ as the multiplication-region transit time and p as a number that changes from 2 to 1/3 as the gain changes from 1 to 1000. The gain decreases from its low-frequency value when $M_{\text{DC}}\omega = 1/\tau_1$. The gain-bandwidth product describes the characteristics of an avalanche photodiode in a communication system.

Noise The shot noise in an APD is that of a *pin* diode multiplied by M^2 times an excess noise factor F_e :

$$i_S^2 = 2eI_{\text{PC}} \Delta f M^2 F_e \quad (75)$$

where

$$F_e(M) = \beta M + (1 - \beta) \left(2 - \frac{1}{M} \right)$$

In this expression, β is the ratio of the ionization coefficient of the opposite type divided by the ionization coefficient of the carrier type that initiates multiplication. In the limit of equal ionization coefficients of electrons and holes (usually the case in III-V semiconductors), $F_e = M$ and $F_h = 1$. Typical numerical values for enhanced APD sensitivity are given in Chap. 26 in Vol. II, Fig. 15.

Dark Current and Shot Noise In an APD, dark current is the sum of the unmultiplied current I_{du} , mainly due to surface leakage, and the bulk dark current experiencing multiplication I_{dm} , multiplied by the gain: $I_d = I_{du} + MI_{dm}$. The shot noise from dark (leakage) current i_d is $i_d^2 = 2e[I_{du} + I_{dm}M^2F_e(M)]\Delta f$.

The proper use of APDs requires choosing the proper design, carefully controlling the voltage, and using the APD in a suitably designed system, since the noise is so large.

MSM Detectors

Volume II, Chap. 26, Fig. 1 of this *Handbook* shows that interdigitated electrodes on top of a semiconductor can provide a planar configuration for electrical contacts. Either a *pn* junction or bulk semiconductor material can reside under the interdigitated fingers. The MSM geometry has the advantage of lower capacitance for a given cross-sectional area, but the transit times may be longer, limited by the lithographic ability to produce very fine lines. Typically, MSM detectors are photoconductive. Volume II, Chap. 26, Fig. 17 shows the geometry of high-speed interdigitated photoconductors. These are simple to fabricate and can be integrated in a straightforward way onto MESFET preamplifiers.

Consider parallel electrodes deposited on the surface of a photoconductive semiconductor with a distance L between them. Under illumination, the photocarriers will travel laterally to the electrodes. The photocurrent in the presence of P_i input optical flux at photon energy $h\nu$ is: $I_{ph} = q\eta GP h\nu$. The photoconductive gain G is the ratio of the carrier lifetime τ to the carrier transit time τ_{tr} : $G = \tau/\tau_{tr}$. Decreasing the carrier lifetime increases the speed but decreases the sensitivity.

The output signal is due to the time-varying resistance that results from the time-varying photo-induced carrier density $N(t)$:

$$R_s(t) = \frac{L}{eN(t)\mu w d_e} \quad (76)$$

where μ is the sum of the electron and hole mobilities, w is the length along the electrodes excited by light, and d_e is the effective absorption depth into the semiconductor.

Usually, MSM detectors are not the design of choice for high-quality communication systems. Nonetheless, their ease of fabrication and integration with other components makes them desirable for some low-cost applications—for example, when there are a number of parallel channels and dense integration is required.

Schottky Photodiodes

A Schottky photodiode uses a metal-semiconductor junction rather than a *pin* junction. An abrupt contact between metal and semiconductor can produce a space-charge region. Absorption of light in this region causes photocurrent that can be detected in an external circuit. Because metal-semiconductor diodes are majority carrier devices they may be faster than *pin* diodes (they rely on drift currents only; there is no minority carrier diffusion). Modulation speeds up to 100 GHz have been reported in a $5 \times 5\text{-}\mu\text{m}$ area detector with a $0.3\text{-}\mu\text{m}$ thin drift region using a semitransparent platinum film 10 nm thick to provide the abrupt Schottky contact. Resonant reflective enhancement of the light has been used to improve sensitivity.

13.16 REFERENCES

1. D. Botez, *IEEE J. Quant. Electr.* **17**:178 (1981).
2. See, for example, E. Garmire and M. Tavis, *IEEE J. Quant. Electr.* **20**:1277 (1984).
3. B. B. Elenkrig, S. Smetona, J. G. Simmons, T. Making, and J. D. Evans, *J. Appl. Phys.* **85**:2367 (1999).
4. M. Yamada, T. Anan, K. Tokutome, and S. Sugou, *IEEE Photon. Technol. Lett.* **11**:164 (1999).
5. T. C. Hasenberg and E. Garmire, *IEEE J. Quant. Electr.* **23**:948 (1987).
6. D. Botez and M. Ettenberg, *IEEE J. Quant. Electr.* **14**:827 (1978).
7. G. P. Agrawal and N. K. Dutta, *Semiconductor Lasers*, 2d ed., Van Nostrand Reinhold, New York, 1993, Sec. 6.4.
8. G. H. B. Thompson, *Physics of Semiconductor Laser Devices*, John Wiley & Sons, New York, 1980, Fig. 7.8.
9. K. Tatah and E. Garmire, *IEEE J. Quant. Electr.* **25**:1800 (1989).
10. G. P. Agrawal and N. K. Dutta, Sec. 6.4.3.
11. W. H. Cheng, A. Mar, J. E. Bowers, R. T. Huang, and C. B. Su, *IEEE J. Quant. Electr.* **29**:1650 (1993).
12. J. T. Verdeyen, *Laser Electronics*, 3d ed., Prentice Hall, Englewood Cliffs, N.J., 1995, p. 490.
13. G. P. Agrawal and N. K. Dutta, Eq. 6.6.32.
14. N. K. Dutta, N. A. Olsson, L. A. Koszi, P. Besomi, and R. B. Wilson, *J. Appl. Phys.* **56**:2167 (1984).
15. G. P. Agrawal and N. K. Dutta, Sec. 6.6.2.
16. G. P. Agrawal and N. K. Dutta, Sec. 6.5.2.
17. L. A. Coldren and S. W. Corizine, *Diode Lasers and Photonic Integrated Circuits*, John Wiley & Sons, New York, 1995, Sec. 5.5.
18. M. C. Tatham, I. F. Lealman, C. P. Seltzer, L. D. Westbrook, and D. M. Cooper, *IEEE J. Quant. Electr.* **28**:408 (1992).
19. H. Jackel and G. Guekos, *Opt. Quant. Electr.* **9**:223 (1977).
20. M. K. Aoki, K. Uomi, T. Tsuchiya, S. Sasaki, M. Okai, and N. Chinone, *IEEE J. Quant. Electr.* **27**:1782 (1991).
21. K. Petermann, *IEEE J. Sel. Top. Quant. Electr.* **1**:480 (1995).
22. R. W. Tkach and A. R. Chaplyvy, *J. Lightwave Technol.* **LT-4**:1655 (1986).
23. T. Hirono, T. Kurosaki, and M. Fukuda, *IEEE J. Quant. Electr.* **32**:829 (1996).
24. Y. Kitaoka, *IEEE J. Quant. Electr.* **32**:822 (1996) Fig. 2.
25. M. Kawase, E. Garmire, H. C. Lee, and P. D. Dapkus, *IEEE J. Quant. Electr.* **30**:981 (1994).
26. L. A. Coldren and S. W. Corizine, Sec. 4.3.
27. S. L. Chuang, *Physics of Optoelectronic Devices*, John Wiley & Sons, New York, 1995, Fig. 10.33.
28. T. L. Koch and U. Koren, *IEEE J. Quant. Electr.* **27**:641 (1991).
29. B. G. Kim and E. Garmire, *J. Opt. Soc. Am.* **A9**:132 (1992).
30. A. Yariv, *Optical Electronics*, 4th ed., Saunders, Philadelphia, Pa., 1991, Eq. 13.6–19.
31. J. De Merlier, K. Mizutani, S. Sudo, K. Naniwae, Y. Furushima, S. Sato, K. Sato, and K. Kudo, *IEEE Photon. Technol. Lett.* **17**:681 (2005).
32. K. Takabayashi, K. Takada, N. Hashimoto, M. Doi, S. Tomabechi, G. Nakagawa, H. Miyata, T. Nakazawa, and K. Morito, *Electron. Lett.* **40**:1187 (2004).
33. Y. A. Akulova, G. A. Fish, P.-C. Koh, C. L. Schow, P. Kozodoy, A. P. Dahl, S. Nakagawa, et al., *IEEE J. Sel. Top. Quant. Electr.* **8**:1349 (2002).
34. H. Ishii, Y. Tohmori, Y. Yoshikuni, T. Tamamura, and Y. Kondo, *IEEE Photon. Technol. Lett.* **5**:613 (1993).
35. M. Gotoda, T. Nishimura, and Y. Tokuda, *J. Lightwave Technol.* **23**:2331 (2005).
36. Information from <http://www.bookham.com/pr/20070104.cfm>, accessed July, 2008.
37. L. B. Soldano and E. C. M. Pennings, *IEEE J. Lightwave Technol.* **13**:615 (1995).
38. Information from <http://www.pd-ld.com/pdf/PSLEDSeries.pdf>, accessed July, 2008.
39. Information from <http://www.qphotonics.com/catalog/SINGLE-MODE-FIBER-COUPLED-LASER-DIODE-30mW--1550-nm-p-204.html>, accessed July 2008.

40. C. L. Jiang and B. H. Reysen, *Proc. SPIE* **3002**:168 (1997) Fig. 7.
41. See, for example, P. Bhattacharya, *Semiconductor Optoelectronic Devices*, Prentice-Hall, New York, 1998, Chap. 6.
42. C. H. Chen, M. Hargis, J. M. Woodall, M. R. Melloch, J. S. Reynolds, E. Yablonovitch, and W. Wang, *Appl. Phys. Lett.* **74**:3140 (1999).
43. Y. A. Wu, G. S. Li, W. Yuen, C. Caneau, and C. J. Chang-Hasnain, *IEEE J. Sel. Topics Quant. Electr.* **3**:429 (1997).
44. See, for example, F. L. Pedrotti and L. S. Pedrotti, *Introduction to Optics*, Prentice-Hall, Englewood Cliffs, N.J., 1987.
45. N. Nishiyama, C. Caneau, B. Hall, G. Guryanov, M. H. Hu, X. S. Liu, M.-J. Li, R. Bhat, and C. E. Zah, *IEEE J. Sel. Top. Quant. Electr.* **11**:990 (2005).
46. M. Orenstein, A. Von Lehmen, C. J. Chang-Hasnain, N. G. Stoffel, J. P. Harbison, L. T. Florez, E. Clausen, and J. E. Jewell, *Appl. Phys. Lett.* **56**:2384 (1990).
47. Y. A. Wu, G. S. Li, R. F. Nabiev, K. D. Choquette, C. Caneau, and C. J. Chang-Hasnain, *IEEE J. Sel. Top. Quant. Electr.* **1**:629 (1995).
48. K. Mori, T. Asaka, H. Iwano, M. Ogura, S. Fujii, T. Okada, and S. Mukai, *Appl. Phys. Lett.* **60**:21 (1992).
49. E. Söderberg, J. S. Gustavsson, P. Modh, A. Larsson, Z. Zhang, J. Berggren, and M. Hammar, *IEEE Photon. Technol. Lett.* **19**:327 (2007).
50. C. J. Chang-Hasnain, M. Orenstein, A. Von Lehmen, L. T. Florez, J. P. Harbison, and N. G. Stoffel, *Appl. Phys. Lett.* **57**:218 (1990).
51. F. Romstad, S. Bischoff, M. Juhl, S. Jacobsen and D. Birkedal, *Proc. SPIE*, **C-1-14**: 69080 (2008).
52. C. Carlsson, C. Angulo Barrios, E. R. Messmer, A. Löfqvist, J. Halonen, J. Vukusic, M. Ghisoni, S. Lourdudoss, and A. Larsson, *IEEE J. Quant. Electr.* **37**:945 (2001).
53. A. Valle, M. Gómez-Molina, and L. Pesquera, *IEEE J. Sel. Top. Quant. Electr.* **14**:895 (2008).
54. J. W. Law and G. P. Agrawal, *IEEE J. Sel. Top. Quant. Electr.* **3**:353 (1997).
55. Data from the Web sites of Lasermate, Finisar, ULM, Raycan, JDS Uniphase and LuxNet, accessed on July 24, 2008.
56. Data from the Web sites of Firecomms and Vixar, accessed on July, 2008.
57. Data from the Web sites of Vertilas and Raycan, accessed on July, 2008.
58. S. K. Korotky and R. C. Alferness, "Ti:LiNbO₃ Integrated Optic Technology," in L. D. Hutcheson (ed.), *Integrated Optical Circuits and Components*, Dekker, New York, 1987.
59. G. Y. Wang and E. Garmire, *Opt. Lett.* **21**:42 (1996).
60. A. Yariv, Table 9.2.
61. G. D. Boyd, R. C. Miller, K. Nassau, W. L. Bond, and A. Savage, *Appl. Phys. Lett.* **5**:234 (1964).
62. Information from www.covega.com, accessed on June, 2008.
63. F. P. Leonberger and J. P. Donnelly, "Semiconductor Integrated Optic Devices," in T. Tamir (ed.), *Guided Wave Optoelectronics*, Springer-Verlag, 1990, p. 340.
64. C. C. Chen, H. Porte, A. Carencu, J. P. Goedgebuer, and V. Armbruster, *IEEE Photon. Technol. Lett.* **9**:1361 (1997).
65. C. T. Mueller and E. Garmire, *Appl. Opt.* **23**:4348 (1984).
66. L. B. Soldano and E. C. M. Pennings, *IEEE J. Lightwave Technol.* **13**:615 (1995).
67. D. A. May-Arrijo, P. LiKamWa, R. J. Selvas-Aguilar, and J. J. Sánchez-Mondragón, *Opt. Quant. Electr.* **36**:1275 (2004).
68. R. Thapliya, T. Kikuchi, and S. Nakamura, *Appl. Opt.* **46**:4155 (2007).
69. R. Thapliya, S. Nakamura, and T. Kikuchi, *Appl. Opt.* **45**:5404 (2006).
70. H. Q. Hou, A. N. Cheng, H. H. Wieder, W. S. C. Chang, and C. W. Tu, *Appl. Phys. Lett.* **63**:1833 (1993).
71. A. Ramdane, F. Devauz, N. Souli, D. Dalprat, and A. Ougazzaden, *IEEE J. Sel. Top. Quant. Electr.* **2**:326 (1996).
72. S. D. Koehler and E. M. Garmire, in T. Tamir, H. Bertoni, and G. Griffel (eds.), *Guided-Wave Optoelectronics: Device Characterization, Analysis and Design*, Plenum Press, New York, 1995.
73. See, for example, S. Carbonneau, E. S. Koteles, P. J. Poole, J. J. He, G. C. Aers, J. Haysom, M. Buchanan, et al., *IEEE J. Sel. Top. Quantum Electron.* **4**:772 (1998).

- 74. J. A. Trezza, J. S. Powell, and J. S. Harris, *IEEE Photon. Technol. Lett.* **9**:330 (1997).
- 75. Information from www.okioptical.com, accessed on June 30, 2008.
- 76. Information from www.ciphotonics.com, accessed on June 30, 2008.
- 77. M. Jupina, E. Garmire, M. Zembutsu, and N. Shibata, *IEEE J. Quant. Electr.* **28**:663 (1992).
- 78. S. D. Koehler, E. M. Garmire, A. R. Kost, D. Yap, D. P. Doctor, and T. C. Hasenberg, *IEEE Photon. Technol. Lett.* **7**:878 (1995).
- 79. A. Sneh, J. E. Zucker, B. I. Miller, and L. W. Stultz, *IEEE Photon. Technol. Lett.* **9**:1589 (1997).
- 80. J. Pamulapati, J. P. Loehr, J. Singh, and P. K Bhattacharya, *J. Appl. Phys.* **69**:4071 (1991).
- 81. H. Feng, J. P. Pang, M. Sugiyama, K. Tada, and Y. Nakano, *IEEE J. Quant. Electr.* **34**:1197 (1998).
- 82. J. Wang, J. E. Zucker, J. P. Leburton, T. Y. Chang, and N. J. Sauer, *Appl. Phys. Lett.* **65**:2196 (1994).
- 83. N. Yoshimoto, Y. Shibata, S. Oku, S. Kondo, and Y. Noguchi, *IEEE Photon. Technol. Lett.* **10**:531 (1998).
- 84. A. Yariv, Sec. 11.7.
- 85. Y. Muramoto, K. Kato, M. Mitsuhara, O. Nakajima, Y. Matsuoka, N. Shimizu and T. Ishibashi, *Electro. Lett.* **34**(1):122 (1998).
- 86. M. Achouche, V. Magnin, J. Harari, F. Lelarge, E. Derouin, C. Jany, D. Carpentier, F. Blache, and D. Decoster. *IEEE Photon. Technol. Lett.* **16**:584, 2004.
- 87. E. Hasnain et al., *IEEE J. Quant. Electr.* **34**:2321 (1998).

OPTICAL FIBER AMPLIFIERS

John A. Buck

*Georgia Institute of Technology
School of Electrical and Computer Engineering
Atlanta, Georgia*

14.1 INTRODUCTION

The development of optical fiber amplifiers has led to dramatic increases in the transport capacities of fiber communication systems. At the present time, fiber amplifiers are used in practically every long-haul optical fiber link and advanced large-scale network. Additional applications of fiber amplifiers include their use as gain media in fiber lasers, as wavelength converters, and as stand-alone high-intensity light sources.

The original intent in fiber amplifier development was to provide a simpler alternative to the electronic repeater, chiefly by allowing the signal to remain in optical form throughout a link or network. Fiber amplifiers as repeaters offer additional advantages, which include the ability to change system data rates as needed, or to simultaneously transmit multiple rates—all without the need to modify the transmission link. A further advantage is that signal power at multiple wavelengths can be simultaneously boosted by a single amplifier—a task that would otherwise require a separate electronic repeater for each wavelength. This latter feature contributed to the realization of dense wavelength division multiplexed (DWDM) systems that provide terabit per second data rates.¹ As an illustration, the useful gain in a normally configured erbium-doped fiber amplifier (EDFA) occupies a wavelength range spanning 1.53 to 1.56 μm , which defines the C band. In DWDM systems this allows, for example, the use of some 40 channels having 100 GHz spacing.

A fundamental disadvantage of the fiber amplifier as a repeater is that dispersion is not reset. This requires additional efforts in dispersion management, which may include optical or electronic equalization methods.^{2,3} More recently, special fiber amplifiers have been developed that provide compensation for dispersion.^{4,5} The development⁶ and deployment⁷ of long-range systems that employ optical solitons have also occurred. The use of solitons (pulses that maintain their shape by balancing linear group velocity dispersion with nonlinear self-phase modulation) requires fiber links in which optical power levels can be adequately sustained over long distances. The use of fiber amplifiers allows this possibility. The deployment of fiber amplifiers in commercial networks demonstrates the move toward *transparent* fiber systems, in which signals are maintained in optical form, and in which multiple wavelengths, data rates, and modulation formats are supported.

Successful amplifiers can be grouped into three main categories. These include (1) rare-earth-doped fibers (including EDFAs), in which dopant ions in the fiber core provide gain through stimulated emission, (2) Raman amplifiers, in which gain for almost any optical wavelength can be formed

TABLE 1 Fiber Transmission Bands Showing Fiber Amplifier Coverage

Designation	Meaning	Wavelength Range (μm)	Amplifier (Pump Wavelength) [Ref]
O	Original	1.26–1.36	PDFA (1.02) ⁸
E	Extended	1.36–1.46	Raman (1.28–1.37)
S	Short	1.46–1.53	TDFA (0.8, 1.06, or 1.56 with 1.41) ¹¹
C	Conventional	1.53–1.56	EDFA (0.98, 1.48), EYDFA (1.06) ¹⁰
L	Long	1.56–1.63	Reconfigured EDFA (0.98, 1.48) ¹³
U (XL)	Ultralong	1.63–1.68	Raman (1.52–1.56)

in conventional or specialty fiber through stimulated Raman scattering, and (3) parametric amplification, in which signals are amplified through nonlinear four-wave mixing in fiber.

Within the first category, the most widely used are the erbium-doped fiber amplifiers, in which gain occurs at wavelengths in the vicinity of 1.53 μm . The amplifiers are optically pumped using light at either 1.48 μm or (more commonly) 0.98 μm wavelengths. Other devices include praseodymium-doped fiber amplifiers (PDFAs), which provide gain at 1.3 μm and which are pumped at 1.02 μm .⁸ Ytterbium-doped fibers (YDFAs)⁹ amplify from 0.98 to 1.15 μm , using pump wavelengths between 0.91 and 1.06 μm ; erbium-ytterbium codoped fibers (EYDFAs) enable use of pump light at 1.06 μm while providing gain at 1.55 μm .¹⁰ Additionally, thulium and thulium/terbium-doped fluoride fibers have been constructed for amplification at 0.8, 1.4, and 1.65 μm .¹¹

In the second category, gain from stimulated Raman scattering (SRS) develops as power couples from an optical pump wave to a longer-wavelength signal (Stokes) wave, which is to be amplified. This process occurs as both waves, which either copropagate or counterpropagate, interact with vibrational resonances in the glass material. Raman fiber amplifiers have the advantage of enabling useful gain to occur at any wavelength (or multiple wavelengths) at which the fiber exhibits low loss, and for which the required pump wavelength is available. Long spans (typically several kilometers) are usually necessary for Raman amplifiers, whereas only a few meters are needed for a rare-earth-doped amplifier. Incorporating Raman amplifiers in systems is therefore often done by introducing the required pump power into a section of the existing transmission fiber.

Finally, in nonlinear parametric amplification (arising from four-wave mixing) signals are again amplified in the presence of a strong pump wave at a different frequency. The process differs in that the electronic (catalytic) nonlinearity in fiber is responsible for wave coupling, as opposed to vibrational resonances (optical phonons) that mediate wave coupling in SRS. Parametric amplifiers can exhibit exponential gain coefficients that are twice the value of those possible for SRS in fiber.¹² Achieving high parametric gain is a more complicated problem in practice, however, as will be discussed. A useful by-product of the parametric process is a wavelength-shifted replica of the signal wave (the idler) which is proportional to the phase conjugate of the signal.

Table 1 summarizes doped fiber amplifier usage in the optical fiber transmission bands, in which it is understood that Raman or parametric amplification can in principle be performed in any of the indicated bands. In cases where doped fibers are unavailable, Raman amplification is indicated, along with the corresponding pump wavelengths.

14.2 RARE-EARTH-DOPED AMPLIFIER CONFIGURATION AND OPERATION

Pump Configuration and Optimum Fiber Length

A typical rare-earth-doped fiber amplifier configuration consists of the doped fiber positioned between polarization-independent optical isolators. Pump light is input by way of a wavelength-selective coupler (WSC) which can be configured for forward, backward, or bidirectional pumping (Fig. 1). Pump absorption throughout the amplifier length results in a population inversion that

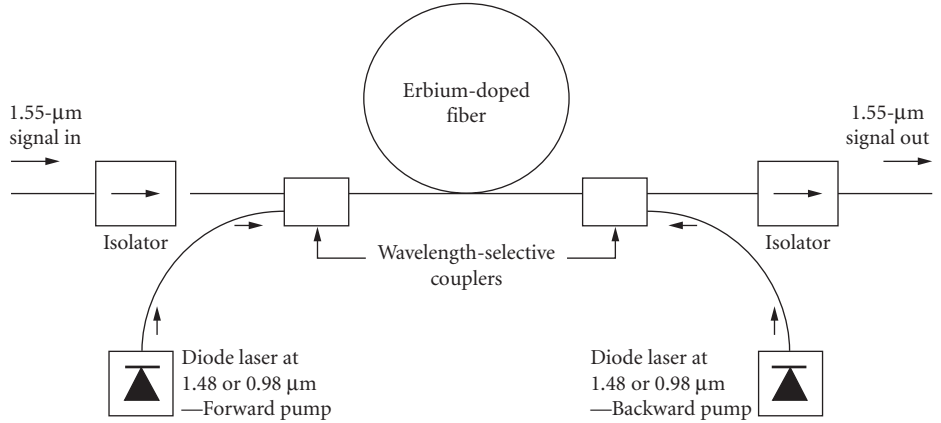


FIGURE 1 General erbium-doped fiber amplifier configuration showing bidirectional pumping.

varies with position along the fiber; this reaches a minimum at the fiber end opposite the pump laser for unidirectional pumping, or minimizes at midlength for bidirectional pumping, using equal pump powers. To achieve the highest overall gain for unidirectional pumping, the fiber length is chosen so that at the output (the point of minimum pump power), the exponential gain coefficient is zero—and no less. If the amplifier is too long, some reabsorption of the signal will occur beyond the transparency point, as the gain goes negative. With lengths shorter than the optimum, full use is not made of the available pump energy, and the overall gain factor is reduced. Other factors may modify the optimum length, particularly if substantial gain saturation occurs, or if amplified spontaneous emission (ASE), which can result in additional gain saturation and noise, is present.¹⁴

Isolators maintain unidirectional light propagation so that, for example, no Rayleigh backscattered or reflected light from further down the link can re-enter the amplifier and cause gain quenching, noise enhancement, or possibly lasing. Double-pass and segmented configurations are also used; in the latter, isolators are positioned between two or more lengths of amplifying fiber which are separately pumped. The result is that gain quenching and noise arising from backscattered light or from amplified spontaneous emission (ASE) are reduced over that of a single fiber amplifier of the combined lengths.

Regimes of Operation

There are roughly three operating regimes, the choice between which is determined by the use intended for the amplifier.^{15,16} These are (1) small-signal or linear regime, (2) saturation, and (3) deep saturation.

In the linear regime, low input signal levels ($< 1 \mu\text{W}$) are amplified with negligible gain saturation, using the optimum amplifier length as discussed in the last section. EDFA gains that range between 25 and 35 dB are possible in this regime.¹⁶ Amplifier gain in decibels is defined in terms of the input and output signal powers as $G \text{ (dB)} = 10 \log_{10}(P_s^{\text{out}}/P_s^{\text{in}})$.

In the saturation regime, the input signal level is high enough to cause a measurable reduction (compression) in the net gain. A useful figure of merit is the *input saturation power*, $P_{\text{sat}}^{\text{in}}$, defined as the input signal power required to compress the net amplifier gain by 3 dB. Specifically, the gain in this case is $G = G_{\text{max}} - 3 \text{ dB}$, where G_{max} is the small-signal gain. A related parameter is the *saturation output power*, $P_{\text{sat}}^{\text{out}}$, defined as the amplifier output that is achieved when the overall gain is compressed by 3 dB. The two quantities are thus related through $G_{\text{max}} - 3 \text{ dB} = 10 \log_{10}(P_{\text{sat}}^{\text{out}}/P_{\text{sat}}^{\text{in}})$. Again, it is assumed that the amplifier length is optimized when defining these parameters.

The *dynamic range* of the amplifier is defined through $P_s^{\text{in}} \leq P_{\text{sat}}^{\text{in}}$, or equivalently $P_s^{\text{out}} \leq P_{\text{sat}}^{\text{out}}$. For an N -channel wavelength-division multiplexed signal, the dynamic range is reduced accordingly by a factor of $1/N$, assuming a flat gain spectrum.¹⁶

With the amplifier operating in deep saturation, gain compressions on the order of 20 to 40 dB occur.¹⁵ This is typical of *power amplifier* applications, in which input signal levels are high, and where the maximum output signal power is desired. In this application, the concept of *power conversion efficiency* (PCE) between pump and signal becomes important. It is defined as $PCE = (P_s^{\text{out}} - P_s^{\text{in}})/P_p^{\text{in}}$, where P_p^{in} is the input pump power.

Another important quantity that is pertinent to the deep saturation regime is the *saturated output power*, $P_s^{\text{out}}(\text{max})$ (not to be confused with the saturation output power described above). $P_s^{\text{out}}(\text{max})$ is the maximum output signal power that can be achieved for a given input signal level and available pump power. This quantity would maximize when the amplifier, having previously been fully inverted, is then completely saturated by the signal. Maximum saturation, however, requires the input signal power to be extremely high, such that ultimately, $P_s^{\text{out}}(\text{max}) \approx P_p^{\text{in}}$, representing a net gain of nearly 0 dB. Clearly the more important situations are those in which moderate signal powers are to be amplified; in these cases the choice of pump power and pumping configuration can substantially influence $P_s^{\text{out}}(\text{max})$.

14.3 EDFA PHYSICAL STRUCTURE AND LIGHT INTERACTIONS

Energy Levels in the EDFA

Gain in the erbium-doped fiber system occurs when an inverted population exists between parts of the $^4I_{13/2}$ and $^4I_{15/2}$ states, as shown in Fig. 2a.¹⁷ This notation uses the standard form, $^{(2S+1)}L_J$, where L , S , and J are the orbital, spin, and total angular momenta, respectively. EDFAs are manufactured by incorporating erbium ions into the glass matrix that forms the fiber core. Interactions between the ions and the host matrix induces Stark splitting of the ion energy levels, as shown in Fig. 2a. This produces an average spacing between adjacent Stark levels of 50 cm^{-1} , and an overall spread of 300 to 400 cm^{-1} within each state. A broader emission spectrum results, since more de-excitation pathways are produced, which occur at different transition wavelengths.

Other mechanisms further broaden the emission spectrum. First, the extent to which ions interact with the glass varies from site to site, as a result of the nonuniform structure of the amorphous glass matrix. This produces some degree of inhomogeneous broadening in the emission spectrum, the extent of which varies with the type of glass host used.¹⁸ Second, thermal fluctuations in the material lead to homogeneous broadening of the individual Stark transitions. The magnitudes of the two broadening mechanisms are 27 to 60 cm^{-1} for inhomogeneous, and 8 to 49 cm^{-1} for homogeneous.¹⁸

The choice of host material strongly affects the shape of the emission spectrum, owing to the character of the ion-host interactions. For example, in pure silica (SiO_2), the spectrum of the Er-doped system is narrowest and has the least smoothness. Use of an aluminosilicate host ($\text{SiO}_2\text{-Al}_2\text{O}_3$), produces slight broadening and smoothing.¹⁹ The broadest spectra, however, occur when using fluoride-based glass, such as ZBLAN ($\text{ZrF}_4\text{-BaF}_2\text{-LaF}_3\text{-AlF}_3\text{-NaF}$).²⁰

Gain Formation

Figure 2b shows how the net emission spectrum is constructed from the superposition of the individual Stark spectra; the latter are associated with the transitions shown in Fig. 2a. Similar diagrams can be constructed for the upward (absorptive) transitions, from which the absorption spectrum can be developed.²⁰ The shapes of both spectra are further influenced by the populations within the Stark-split levels, which assume a Maxwell-Boltzman distribution. The sequence of events in the population dynamics is (1) pump light boosts population from the ground state, $^4I_{15/2}$, to the upper Stark levels in the first excited state, $^4I_{13/2}$; (2) the upper state Stark level populations thermalize; and (3) de-excitation from $^4I_{13/2}$ to $^4I_{15/2}$ occurs through either spontaneous or stimulated emission.

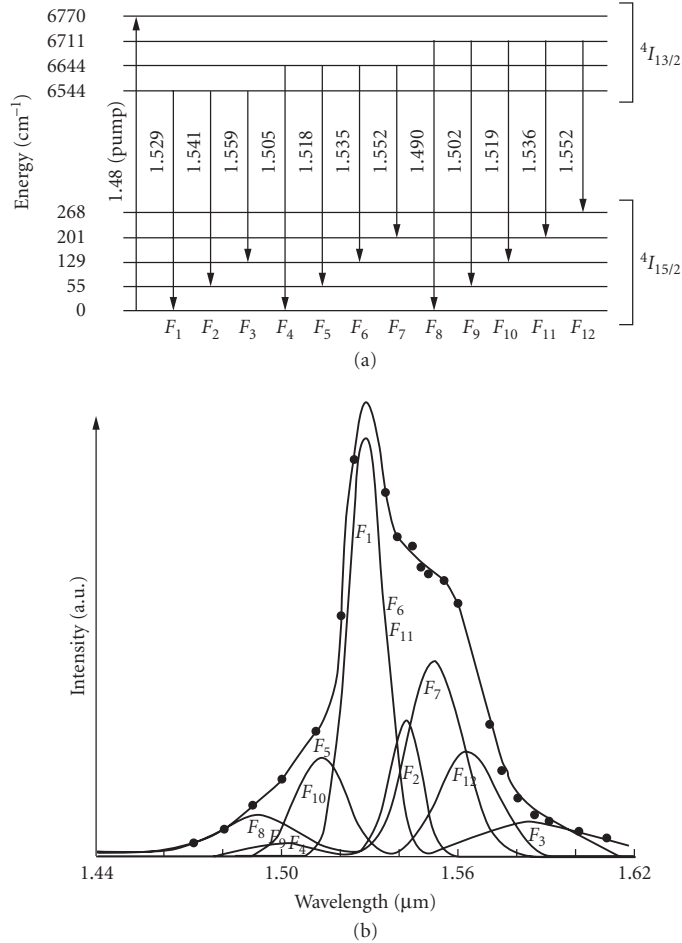


FIGURE 2 (a) Emissive transitions between Stark-split levels of erbium in an aluminosilicate glass host. Values on transition arrows indicate wavelengths in micrometers. (Adapted from Ref. 17.) (b) EDFA fluorescence spectrum arising from the transitions in Fig. 2a. (Reprinted with permission from Ref. 18.)

The system can be treated using a simple two-level (1.48 μm pump) or three-level model (0.98 μm pump), from which rate equations can be constructed that incorporate the actual wavelength- and temperature-dependent absorption and emission crosssections. These models have been formulated with and without inhomogeneous broadening. In most cases, results that are in excellent agreement with experiment have been achieved by assuming only homogeneous broadening.^{21–23}

Pump Wavelength Options in EDFAs

The 1.48 μm pump wavelength corresponds to the energy difference between the two most widely spaced Stark levels, as shown in Fig. 2a. A better alternative is to pump with light at 0.98 μm , which boosts the ground state population to the second excited state, $^4I_{11/2}$, which lies above $^4I_{13/2}$.

This is followed by rapid nonradiative decay into $^4I_{13/2}$ and gain is formed as before. The pumping efficiency suffers slightly at 0.98 μm , owing to some excited state absorption (ESA) between $^4I_{11/2}$ and the higher-lying $^4F_{7/2}$ state at this wavelength.²⁴ Use of 0.98 μm pump light as opposed to 1.48 μm , will nevertheless yield a more efficient system, since the 0.98 μm pump will not contribute to the de-excitation process, as occurs when 1.48 μm is used.

The *gain efficiency* of a rare-earth-doped fiber is defined as the ratio of the maximum small signal gain to the input pump power, using the optimized fiber length. EDFA efficiencies are typically on the order of 10 dB/mW for pumping at 0.98 μm . For pumping at 1.48 μm , efficiencies are about half the values obtainable at 0.98 μm , and require about twice the fiber length. Other pump wavelengths can be used,²⁴ but with some penalty to be paid in the form of excited state absorption from the $^4I_{13/2}$ state into various upper levels, thus depleting the gain that would otherwise be available. This problem is minimized when using either 0.98 or 1.48 μm , and so these two wavelengths are almost exclusively used in erbium-doped fibers.

Noise

Performance is degraded by the presence of noise from two fundamental sources. These are (1) amplified spontaneous emission (ASE), and (2) Rayleigh scattering. Both processes lead to additional light that propagates in the forward and backward directions, and which encounters considerable gain over long amplifier lengths. The more serious of the two noise sources is ASE. In severe cases, involving high-gain amplifiers of long lengths, ASE can be of high enough intensity to partially saturate the gain, thus reducing the available gain for signal amplification. This *self-saturation* effect has been reduced by using the backward pumping geometry.²⁵

In general, ASE can be reduced by (1) assuring that the population inversion is as high as possible (ideally, completely inverted), (2) operating the amplifier in the deep saturation regime, or (3) using two or more amplifier stages rather than one continuous length of fiber, and positioning bandpass filters and isolators between stages. Rayleigh scattering noise can be minimized by using multistage configurations, in addition to placing adequate controls on dopant concentration and confinement during the manufacturing stage.²⁶

The *noise figure* of a rare-earth-doped fiber amplifier is stated in a manner consistent with the IEEE standard definition for a general amplifier (Friis definition). This is the signal-to-noise ratio of the fiber amplifier input divided by the signal-to-noise ratio of the output, expressed in decibels, where the input signal is shot noise limited. Although this definition is widely used, it has become the subject of a debate, arising from the physical nature of ASE noise, and the resulting awkwardness in applying the definition to cascaded amplifier systems.²⁷ An in-depth review of this subject is in Ref. 28.

The best noise figures for EDFAs are achieved by using pump configurations that yield the highest population inversions. Again, the use of 0.98 μm is preferred, yielding noise figures that approach the theoretical limit of 3 dB. Pumping at 1.48 μm gives best results of about 4 dB.¹⁵

Gain Flattening

Use of multiple wavelength channels in WDM systems produces a strong motivation to construct a fiber amplifier in which the gain is uniform for all wavelengths. Thus some means needs to be employed which will effectively flatten the emission spectrum as depicted in Fig. 2b. Flattening techniques can be classified into roughly three categories. First, intrinsic methods can be used; these involve choices of fiber host materials such as fluoride glass²⁹ that yield smoother and broader gain spectra. In addition, by carefully choosing pump power levels, a degree of population inversion can be obtained which will allow some cancellation to occur between the slopes of the absorption and emission spectra,³⁰ thus producing a flatter gain spectrum. Second, spectral filtering at the output of a single amplifier or between cascaded amplifiers can be employed; this effectively produces higher loss for wavelengths that have achieved higher gain. Examples of successful filtering devices include long-period fiber gratings³¹ and Mach-Zehnder filters.³² Third, hybrid amplifiers that use cascaded

configurations of different gain media can be used to produce an overall gain spectrum that is reasonably flat. Flattened gain spectra have been obtained having approximate widths that range from 12 to 85 nm. Reference 33 is recommended for an excellent discussion and comparison of the methods.

14.4 OTHER RARE-EARTH SYSTEMS

Praseodymium-Doped Fiber Amplifiers (PDFAs)

In the praseodymium-doped fluoride system, the strongest gain occurs in the vicinity of 1.3 μm , with the pump wavelength at 1.02 μm . Gain formation is described by a basic three-level model, in which pump light excites the system from the ground state, $^3H_4$, to the metastable excited state, $^1G_4$. Gain for 1.3 μm light is associated with the downward $^1G_4 - ^3H_5$ transition, which peaks in the vicinity of 1.32 to 1.34 μm . Gain diminishes at longer wavelengths, principally as a result of ground state absorption from $^3H_4$ to $^3F_3$.¹⁸

The main problem with the PDFa system has been the reduction of the available gain through the competing $^1G_4 - ^3F_4$ transition (2900 cm^{-1} spacing), occurring through multiphonon relaxation. The result is that the *radiative quantum efficiency* (defined as the ratio of the desired transition rate to itself plus all competing transition rates) can be low enough in conventional glass host materials to make the system impractical. The multiphonon relaxation rate is reduced when using hosts having low phonon energies, such as fluoride or chalcogenide glasses. Use of the latter material has essentially solved the problem, by yielding radiative quantum efficiencies on the order of 90%.³⁴ For comparison, erbium systems exhibit quantum efficiencies of nearly 100% for the 1.5 μm transition. Other considerations such as broadening mechanisms and excited state absorption are analogous to the erbium system. References 1 and 35 are recommended for further reading.

Ytterbium-Doped Fiber Amplifiers (YDFAs)

Ytterbium-doping provides the most efficient fiber amplifier system, as essentially no competing absorption and emission mechanisms exist.⁹ This is because in ytterbium, there are only two energy states that are resonant at the wavelengths of interest. These are the ground state $^2F_{7/2}$ and the excited state $^2F_{5/2}$. When doped into the host material, Stark splitting within these levels occurs as described previously, which leads to strong absorption at wavelengths in the vicinity of 0.92 μm , and emission between 1.0 and 1.1 μm , maximizing at around 1.03 μm . Pump absorption is very high, which makes side-pumping geometries practical. Because of the extremely high gain that is possible, Yb-doped fibers are attractive as power amplifiers for 1.06- μm light, and have been employed in fiber laser configurations. YDFAs have also proven attractive as superfluorescent sources,³⁶ in which the output is simply amplified spontaneous emission, and there is no signal input.

Accompanying the high power levels in YDFAs are unwanted nonlinear effects, which are best reduced (at a given power level) by lowering the fiber mode intensity. This has been accomplished to an extent in special amplifier designs that involve large mode effective areas (A_{eff}). Such designs have been based on either conventional fiber³⁷ or photonic crystal fiber.^{38,39} The best results in both cases involve dual-core configurations, in which the pump light propagates in a large core (or inner cladding) which surrounds a smaller concentric core that contains the dopant ions, and that propagates the signal to be amplified. The large inner cladding region facilitates the input coupling of high-power diode pump lasers that have large output beam cross sections. Such designs have proven successful with other amplifiers (including erbium-doped) as well.

Erbium/Ytterbium-Doped Fiber Amplifiers (EYDFAs)

Erbium/ytterbium codoping takes advantage of the strong absorption of ytterbium at the conventional 0.98 μm erbium pump wavelength. When codoped with erbium, ytterbium ions in their excited state transfer their energy to the erbium ions, and gain between 1.53 and 1.56 μm is formed

as before.³ Advantages of such a system include the following: With high pump absorption, side-pumping is possible, thus allowing the use of large-area diode lasers as pumps. In addition, high gain can be established over a shorter propagation distance in the fiber than is possible in a conventional EDFA. As a result, shorter length amplifiers having lower ASE noise can be constructed. An added benefit is that the absorption band allows pumping by high-power lasers such as Nd: YAG (at 1.06 μm) or Nd: YLF (at 1.05 μm), and there is no excited state absorption. Yb-sensitized fibers are attractive for use as C or L band power amplifiers, and in the construction of fiber lasers, in which a short-length, high-gain medium is needed.⁴⁰

14.5 RAMAN FIBER AMPLIFIERS

Amplification by stimulated Raman scattering has proven to be a successful alternative to rare-earth-doped fiber, and has found wide use in long-haul fiber communication systems.^{41,42} Key advantages of Raman amplifiers include (1) improvement in signal-to-noise ratio over rare-earth-doped fiber, and (2) wavelengths to be amplified are not restricted to lie within a specific emission spectrum, but only require a pump wavelength that is separated from the signal wavelength by the Raman resonance. In this way, the entire low-loss spectral range of optical fiber can in principle be covered by Raman amplification, as in, for example, O band applications.⁴³ The main requirement is that a pump laser is available having power output on the order of 0.5 W, and whose frequency is up-shifted from that of the signal by the primary Raman resonance frequency of 440 cm^{-1} or about 13.2 THz. The required pump wavelength, λ_2 , is thus expressed in terms of the signal wavelength, λ_1 , through

$$\lambda_2 = \frac{\lambda_1}{(1 + 0.044\lambda_1)} \quad (1)$$

where the wavelengths are expressed in micrometers.

Another major difference from rare-earth-doped fiber is that Raman amplifiers may typically require lengths on the order of tens of kilometers to achieve the same gain that can be obtained, for example, in 10 m of erbium-doped fiber. The long Raman amplifier span is not necessarily a disadvantage because (1) Raman amplification can be carried out within portions of the existing fiber link, and (2) the long span may contribute to an improvement in the signal-to-noise ratio for the entire link. This happens if Raman amplification is used *after* amplifying using erbium-doped fiber; the Raman amplifier provides gain for the signal, while the long span attenuates the spontaneous emission noise from the EDFA. A Raman amplifier can be implemented at the receiver end of a link by introducing a backward pump at the output end.

The basic configuration of a Raman fiber amplifier is shown in Fig. 3. Pumping can be done in either the forward direction (with input pump power P_{20}) or in the backward direction (with input P_{2L}). The governing equations are Eqs. (10) and (11) in Chap. 10 of this volume, rewritten here in terms of wave power:

$$\frac{dP_1}{dz} = \frac{g_r}{A_{\text{eff}}} P_1 P_2 - \alpha P_1 \quad (2)$$

$$\frac{dP_2}{dz} = \pm \frac{\omega_2}{\omega_1} \frac{g_r}{A_{\text{eff}}} P_1 P_2 \pm \alpha P_2 \quad (3)$$

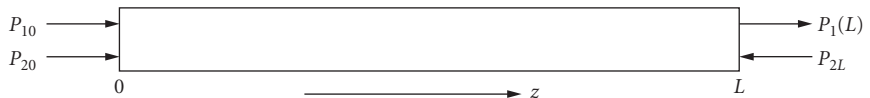


FIGURE 3 Beam configuration for a Raman fiber amplifier using forward or backward pumping.

where the plus and minus signs apply to backward and forward pumping, respectively, and where A_{eff} is the fiber mode cross-sectional area, as before. The peak Raman gain in silica occurs when pump and signal frequencies are spaced by 440 cm^{-1} , corresponding to a wavelength spacing of about $0.1 \text{ }\mu\text{m}$ (see Fig. 2 in Chap. 10 of this volume). The gain in turn is inversely proportional to pump wavelength, λ_2 , and to a good approximation is given by

$$g_r(\lambda_2) = \frac{10^{-11}}{\lambda_2} \text{ cm/W} \quad (4)$$

with λ_2 expressed in micrometers.

The simplest case is the small-signal regime, in which the Stokes power throughout the amplifier is sufficiently low such that negligible pump depletion will occur. The pump power dependence on distance is thus determined by loss in the fiber, and is found by solving Eq. (3) under the assumption that the first term on the right-hand side is negligible. It is further assumed that there is no spontaneous scattering, and that the Stokes and pump fields maintain parallel polarizations. For forward pumping, the solutions of Eqs. (2) and (3) thus simplified are

$$P_1(z) = P_{10} \exp(-\alpha z) \exp \left[\frac{g_r P_{20}}{A_{\text{eff}}} \left(\frac{1 - \exp(-\alpha z)}{\alpha} \right) \right] \quad (5a)$$

$$P_2(z) = P_{20} \exp(-\alpha z) \quad (5b)$$

For backward pumping with no pump depletion, and with fiber length, L :

$$P_1(z) = P_{10} \exp(-\alpha z) \exp \left[\frac{g_r P_{2L}}{A_{\text{eff}}} \exp(-\alpha L) \left(\frac{\exp(\alpha z) - 1}{\alpha} \right) \right] \quad (6a)$$

$$P_2(z) = P_{2L} \exp[\alpha(z - L)] \quad (6b)$$

Implicit throughout is the assumption that there is no spontaneous scattering, and that the Stokes and pump fields maintain parallel polarizations.

The z -dependent signal gain in decibels is $\text{Gain (dB)} = 10 \log_{10}(P_1(z)/P_{10})$. Figure 4⁴⁴ shows this evaluated for several choices of fiber attenuation. It is evident that for forward pumping, an optimum length may exist at which the gain maximizes. For backward pumping, gain will always increase with amplifier length. It is also evident that for a specified fiber length, input pump power, and loss, the same gain is achieved over the *total* length for forward *and* backward pumping, as must be true with no pump depletion.

It is apparent in Eqs. (5a) and (6a) that increased gain is obtained for lower-loss fiber, and by increasing the pump *intensity*, given by P_2/A_{eff} . For a given available pump power, fibers having smaller effective areas, A_{eff} , will yield higher gain, but the increased intensity may result in additional unwanted nonlinear effects.

In actual systems, additional problems arise. Among these is pump depletion, reducing the overall gain as Stokes power levels increase. This is effectively a gain saturation mechanism, and occurs as some of the Stokes power is back-converted to the pump wavelength through the inverse Raman effect. Again, maintaining pump power levels that are significantly higher than the Stokes levels maintains the small signal approximation, and minimizes saturation. In addition, noise may arise from several sources. These include spontaneous Raman scattering, Rayleigh scattering,⁴⁵ pump intensity noise,⁴⁶ and Raman-amplified spontaneous emission from rare-earth-doped fiber amplifiers elsewhere in the link.⁴⁷ Finally, *polarization-dependent gain* (PDG) arises as pump and Stokes field polarizations randomly move in and out of parallelism, owing to the usual changes in fiber birefringence.⁴⁸ This last effect can be reduced by using depolarized pump inputs.

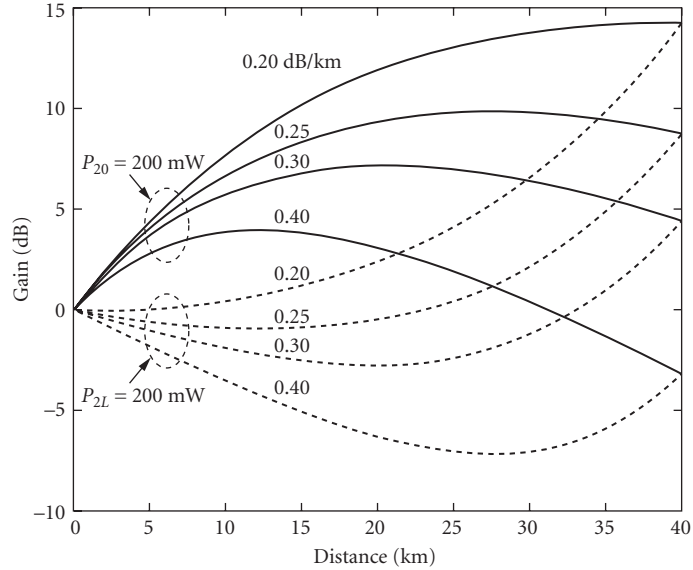


FIGURE 4 Small signal Raman gain as a function of distance z in a single-mode fiber, as calculated using Eqs. (5a) and (6a) for selected values of distributed fiber loss. Plots are shown for cases of forward pumping (solid curves) and backward pumping (dotted curves). Parameter values are $P_{20} = P_{2L} = 200$ mW, $g_r = 7 \times 10^{-12}$ cm/W, and $A_{\text{eff}} = 5 \times 10^{-7}$ cm². (After Ref. 44.)

14.6 PARAMETRIC AMPLIFIERS

Parametric amplification uses the nonlinear four-wave mixing interaction in fiber, as described in Sec. 10.5 (Chap. 10). Two possible configurations are used that involve either a single pump wave, or two pumps at different wavelengths (Fig. 5). The signal to be amplified copropagates with the pumps and is of a different wavelength than either pump. In addition to the amplified signal, the process also generates a fourth wave known as the idler, which is a wavelength-shifted and phase-conjugated replica of the signal. In view of this, parametric amplification is also attractive for use in wavelength conversion applications and—owing to the phase conjugate nature of the idler—in dispersion compensation.

The setup is essentially the same as described in Sec. 10.5, in which we allow the possibility of two distinct pump waves, carrying powers P_1 and P_2 at frequencies ω_1 and ω_2 , or a single pump at frequency ω_0 having power P_0 . The pumps interact with a relatively weak signal, having input power $P_3(0)$, and frequency ω_3 . The signal is amplified as the pump power couples to it, while the idler (power P_4 and frequency ω_4) is generated and amplified. The frequency relations are $\omega_3 + \omega_4 = \omega_1 + \omega_2$ for dual pumps, or $\omega_3 + \omega_4 = 2\omega_0$ for a single pump. In the simple case of a single pump that is nondepleted,

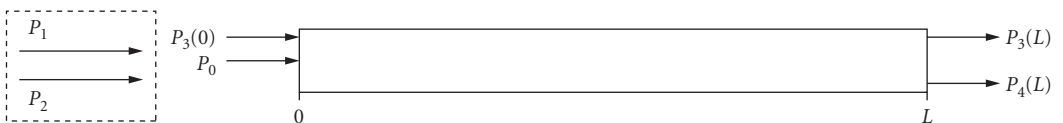


FIGURE 5 Beam configuration for a parametric fiber amplifier using single or dual-wavelength pumping.

and assuming continuous wave operation with a signal input of power $P_3(0)$, the power levels at the amplifier output (length L) are given by:⁴⁹

$$P_3(L) = P_3(0) \left[1 + \left(1 + \frac{\kappa^2}{4g^2} \right) \sinh^2(gL) \right] \quad (7)$$

$$P_4(L) = P_3(0) \left(1 + \frac{\kappa^2}{4g^2} \right) \sinh^2(gL) \quad (8)$$

The parametric gain g is given by

$$g = \left[P_0^2 \left(\frac{2\pi n_2}{\lambda A_{\text{eff}}} \right)^2 - \left(\frac{\kappa}{2} \right)^2 \right]^{1/2} \quad (9)$$

where n_2 is the nonlinear index in m^2/W , λ is the average of the three wavelengths, and A_{eff} is the fiber mode cross-sectional area. The phase mismatch parameter κ includes linear and power-dependent terms:

$$\kappa = \Delta\beta + \frac{2\pi n_2}{\lambda A_{\text{eff}}} P_0 \quad (10)$$

where the linear part, $\Delta\beta = \beta_3 + \beta_4 - 2\beta_0$ has the usual interpretation as the difference in phase constants of a nonlinear polarization wave (at either ω_3 or ω_4) and the field (at the same frequency) radiated by the polarization. The phase constants are expressed in terms of the unperturbed fiber mode indices $\tilde{n}_i$ through $\Delta\beta_i = \tilde{n}_i \omega_i / c$. The second (nonlinear) term on the right hand side of Eq. (10) represents the mode index change arising from the intense pump field through the optical Kerr effect. This uniformly changes the mode indices of all waves, and is thus important to include in the phase mismatch evaluation.

If two pumps are used, having powers P_1 and P_2 , with frequencies ω_1 and ω_2 , Eqs. (9) and (10) are modified by setting $P_0 = P_1 + P_2$. Also, in evaluating $\tilde{P}_0^2$ in Eq. (9), only the cross term ($2P_1P_2$) is retained in the expression. The linear term in Eq. (10) becomes $\Delta\beta = \beta_3 + \beta_4 - \beta_1 - \beta_2$. With these modifications, Eqs. (7) and (8) may not strictly apply because the two-pump interaction is complicated by the generation of multiple idler waves, in addition to contributions from optically induced Bragg gratings, as discussed in Ref. 50. The advantage of using two pumps of different frequencies is that the phase mismatch is possible to reduce significantly over a much broader wavelength spectrum than is possible using a single pump. In this manner, relatively flat gain spectra over several tens of nanometers have been demonstrated with the pump wavelengths positioned on opposite sides of the zero dispersion wavelength.⁵¹ In single-pump operation, gain spectra of widths on the order of 20 nm have been achieved with the pump wavelength positioned at or near the fiber zero dispersion wavelength.⁵² In either pumping scheme, amplification factors on the same order or greater than those available in Raman amplifiers are in principle obtainable, and best results have exceeded 40 dB.⁵³ In practice, random fluctuations in fiber dimensions and in birefringence represent significant challenges in avoiding phase mismatch, and in maintaining alignment of the interacting field polarizations.⁵⁴

14.7 REFERENCES

1. See for example: B. Hoang and O. Perez, "Terabit Networks," www.ieee.org/portal/site/emergingtech/index, 2007.
2. For background on optical methods of dispersion compensation, see the "Special Mini-Issue on Dispersion Compensation," *IEEE Journal of Lightwave Technology* **12**:1706–1765 (1994).
3. B. J. C. Schmidt, A. J. Lowery, and J. Armstrong, "Experimental Demonstration of Electronic Dispersion Compensation for Long-Haul Transmission Using Direct-Detection Optical OFDM," *IEEE Journal of Lightwave Technology* **26**:196–204 (2008).

4. J. Maury, J. L. Auguste, S. Fevrier, and J. M. Blondy, "Conception and Characterization of a Dual-Eccentric-Core Erbium-Doped Dispersion-Compensating Fiber," *Optics Letters* **29**:700–702 (2004).
5. A. C. O. Chan and M. Premaratne, "Dispersion-Compensating Fiber Raman Amplifiers with Step, Parabolic, and Triangular Refractive Index Profiles," *IEEE Journal of Lightwave Technology* **25**:1190–1197 (2007).
6. L. F. Mollenauer and P. V. Mamyshev, "Massive Wavelength-Division Multiplexing with Solitons," *IEEE Journal of Quantum Electronics* **34**:2089–2102 (1998).
7. Alcatel-Lucent 1625 *LambdaXtreme* Transport, www.alcatel-lucent.com, 2006.
8. Y. Ohishi, T. Kanamori, T. Kitagawa, S. Takahashi, E. Snitzer, and G. H. Sigel, Jr., "Pr³⁺-Doped Fluoride Fiber Amplifier Operation at 1.31 μm ," *Optics Letters* **16**:1747–1749 (1991).
9. R. Paschotta, J. Nilsson, A. C. Tropper, D. C. Hanna, "Ytterbium-Doped Fiber Amplifiers," *IEEE Journal of Quantum Electronics* **33**:1049–1056 (1997).
10. J. E. Townsend, K. P. Jedreowski, W. L. Barnes, and S. G. Grubb, "Yb³⁺ Sensitized Er³⁺ Doped Silica Optical Fiber with Ultra High Efficiency and Gain," *Electronics Letters* **27**:1958–1959 (1991).
11. S. Sudo, "Progress in Optical Fiber Amplifiers," in *Current Trends in Optical Amplifiers and Their Applications*, T. P. Lee, ed. (World Scientific, New Jersey, 1996), see pp. 19–21 and references therein.
12. R. H. Stolen, "Phase-Matched Stimulated Four-Photon Mixing in Silica-Fiber Waveguides," *IEEE Journal of Quantum Electronics* **11**:100–103 (1975).
13. Y. Sun, J. L. Zyskind, and A. K. Srivastava, "Average Inversion Level, Modeling, and Physics of Erbium-Doped Fiber Amplifiers," *IEEE Journal of Selected Topics in Quantum Electronics* **3**:991–1007 (1997).
14. P. C. Becker, N. A. Olsson, and J. R. Simpson, *Erbium-Doped Fiber Amplifiers, Fundamentals and Technology* (Academic Press, San Diego, 1999), pp. 139–140.
15. J.-M. P. Delavaux and J. A. Nagel, "Multi-Stage Erbium-Doped Fiber Amplifier Design," *IEEE Journal of Lightwave Technology* **13**:703–720 (1995).
16. E. Desurvire, *Erbium-Doped Fiber Amplifiers, Principles and Applications* (Wiley-Interscience, New York, 1994), pp. 337–340.
17. E. Desurvire, *Erbium-Doped Fiber Amplifiers, Principles and Applications* (Wiley-Interscience, New York, 1994), p. 238.
18. S. Sudo, "Outline of Optical Fiber Amplifiers," in *Optical Fiber Amplifiers: Materials, Devices, and Applications* (Artech House, Norwood, 1997), see pp. 81–83 and references therein.
19. W. J. Miniscalco, "Erbium-Doped Glasses for Fiber Amplifiers at 1500 nm," *IEEE Journal of Lightwave Technology* **9**:234–250 (1991).
20. S. T. Davey and P. W. France, "Rare-Earth-Doped Fluorozirconate Glass for Fibre Devices," *British Telecom Technical Journal* **7**:58 (1989).
21. C. R. Giles and E. Desurvire, "Modeling Erbium-Doped Fiber Amplifiers," *IEEE Journal of Lightwave Technology* **9**:271–283 (1991).
22. E. Desurvire, "Study of the Complex Atomic Susceptibility of Erbium-Doped Fiber Amplifiers," *IEEE Journal of Lightwave Technology* **8**:1517–1527 (1990).
23. Y. Sun, J. L. Zyskind, and A. K. Srivastava, "Average Inversion Level, Modeling, and Physics of Erbium-Doped Fiber Amplifiers," *IEEE Journal of Selected Topics in Quantum Electronics* **3**:991–1007 (1997).
24. M. Horiguchi, K. Yoshino, M. Shimizu, M. Yamada, and H. Hanafusa, "Erbium-Doped Fiber Amplifiers Pumped in the 660- and 820-nm Bands," *IEEE Journal of Lightwave Technology* **12**:810–820 (1994).
25. E. Desurvire, "Analysis of Gain Difference between Forward- and Backward-Pumped Erbium-Doped Fibers in the Saturation Regime," *IEEE Photonics Technology Letters* **4**:711–713 (1992).
26. M. N. Zervas and R. I. Laming, "Rayleigh Scattering Effect on the Gain Efficiency and Noise of Erbium-Doped Fiber Amplifiers," *IEEE Journal of Quantum Electronics* **31**:469–471 (1995).
27. H. A. Haus, "The Noise Figure of Optical Amplifiers," *IEEE Photonics Technology Letters* **10**:1602–1604 (1998).
28. E. Desurvire, D. Bayart, B. Desthieux, and S. Bigo, *Erbium-Doped Fiber Amplifiers, Devices and System Developments* (Wiley-Interscience, Hoboken, 2002), Chap. 2.
29. D. Bayart, B. Clesca, L. Hamon, and J. L. Beylat, "Experimental Investigation of the Gain Flatness Characteristics for 1.55 μm Erbium-Doped Fluoride Fiber Amplifiers," *IEEE Photonics Technology Letters* **6**:613–615 (1994).

30. E. L. Goldstein, L. Eskildsen, C. Lin, and R. E. Tench, "Multiwavelength Propagation in Light-Wave Systems with Strongly-Inverted Fiber Amplifiers," *IEEE Photonics Technology Letters* **6**:266–269 (1994).
31. C. R. Giles, "Lightwave Applications of Fiber Bragg Gratings," *IEEE Journal of Lightwave Technology* **15**:1391–1404 (1997).
32. J.-Y. Pan, M. A. Ali, A. F. Elrefaie, and R. E. Wagner, "Multiwavelength Fiber Amplifier Cascades with Equalization Employing Mach-Zehnder Optical Filters," *IEEE Photonics Technology Letters* **7**:1501–1503 (1995).
33. P. C. Becker, N. A. Olsson, and J. R. Simpson, op. cit., pp. 285–295.
34. D. M. Machewirth, K. Wei, V. Krasteva, R. Datta, E. Snitzer, and G. H. Sigel, Jr., "Optical Characterization of Pr^{3+} and Dy^{3+} Doped Chalcogenide Glasses," *Journal of Noncrystalline Solids* **213–214**:295–303 (1997).
35. T. J. Whitley, "A Review of Recent System Demonstrations Incorporating Praseodymium-Doped Fluoride Fiber Amplifiers," *IEEE Journal of Lightwave Technology* **13**:744–760 (1995).
36. P. Wang and W. A. Clarkson, "High-Power Single-Mode, Linearly Polarized Ytterbium-Doped Fiber Superfluorescent Source," *Optics Letters* **32**:2605–2607 (2007).
37. Y. Jeong, J. Sahu, D. Payne, and J. Nilsson, "Ytterbium-Doped Large-Core Fiber Laser with 1.36 kW Continuous-Wave End-Pumped Optical Power," *Optics Express* **12**:6088–6092 (2004).
38. P. Russel, "Photonic Crystal Fibers," *Science* **299**:358–362 (2003).
39. O. Schmidt J. Rothhardt, T. Eidam, F. Röser, J. Limpert, A. Tünnermann, K. P. Hansen, C. Jakobsen, and J. Broeng, "Single-Polarization Ultra-Large Mode Area Yb-Doped Photonic Crystal Fiber," *Optics Express* **16**:3918–3923 (2008).
40. G. G. Vienne J. E. Caplen, L. Dong, J. D. Minelly, J. Nilsson, and D. N. Payne, "Fabrication and Characterization of $\text{Yb}^{3+}:\text{Er}^{3+}$ Phosphosilicate Fibers for Lasers," *IEEE Journal of Lightwave Technology* **16**:1990–2001 (1998).
41. M. N. Islam, "Raman Amplifiers for Telecommunications," *IEEE Journal of Selected Topics in Quantum Electronics* **8**:548–559 (2002).
42. J. Bromage, "Raman Amplification for Fiber Communication Systems," *IEEE Journal of Lightwave Technology* **22**:79–93 (2004).
43. T. N. Nielsen, P. B. Hansen, A. J. Stentz, V. M. Aguaro, J. R. Pedrazzani, A. A. Abramov, and R. P. Espindola, "8 × 10 Gb/s 1.3- μm Unrepeated Transmission over a Distance of 141 km with Raman Post- and Pre-Amplifiers," *IEEE Photonics Technology Letters* **10**:1492–1494 (1998).
44. J. A. Buck, *Fundamentals of Optical Fibers*, 2nd ed., (Wiley-Interscience, Hoboken, 2004), Chap. 8.
45. P. B. Hansen, L. Eskildsen, A. J. Stentz, T. A. Strasser, J. Judkins, J. J. DeMarco, R. Pedrazzani, and D. J. DiGiovanni, "Rayleigh Scattering Limitations in Distributed Raman Pre-Amplifiers," *IEEE Photonics Technology Letters* **10**:159–161 (1998).
46. C. R. S. Fludger, V. Handerek, and R. J. Mears, "Pump to Signal RIN Transfer in Raman Fiber Amplifiers," *IEEE Journal of Lightwave Technology* **19**:1140–1148 (2001).
47. N. Takachio and H. Suzuki, "Application of Raman-Distributed Amplification to WDM Transmission Systems Using 1.55- μm Dispersion-Shifted Fiber," *IEEE Journal of Lightwave Technology* **19**:60–69 (2001).
48. H. H. Kee, C. R. S. Fludger, and V. Handerek, "Statistical Properties of Polarization Dependent Gain in Fibre Raman Amplifiers," *Optical Fiber Communications Conference*, 2002, paper WB2 (TOPS, vol. 70, Optical Society of America, Washington, D.C.)
49. R. H. Stolen and J. E. Bjorkholm, "Parametric Amplification and Frequency Conversion in Optical Fibers," *IEEE Journal of Quantum Electronics* **18**:1062–1072 (1982).
50. C. J. McKinstrie, S. Radic, and A. R. Chraplyvy, "Parametric Amplifiers Driven by Two Pump Waves," *IEEE Journal of Selected Topics in Quantum Electronics* **8**:538–547 (2002). Erratum, **8**:956.
51. S. Radic, C. J. McKinstrie, R. M. Jopson, J. C. Centanni, Q. Lin, and G. P. Agrawal, "Record Performance of Parametric Amplifier Constructed with Highly-Nonlinear Fibre" *Electronics Letters* **39**:838–839 (2003).
52. J. Hansryd, P. A. Andrekson, M. Westlund, J. Li, and P. O. Hedekvist, "Fiber-Based Optical Parametric Amplifiers and Their Applications," *IEEE Journal of Selected Topics in Quantum Electronics* **8**:506–520 (2002).
53. J. Hansryd and P. A. Andrekson, "Broad-Band Continuous-Wave-Pumped Fiber Optical Parametric Amplifier with 49 dB Gain and Wavelength-Conversion Efficiency," *IEEE Photonics Technology Letters* **13**:194–196 (2001).
54. F. Yaman, Q. Lin, and G. P. Agrawal, "A Novel Design for Polarization-Independent Single-Pump Parametric Amplifiers," *IEEE Photonics Technology Letters* **18**:2335–2337 (2006).

This page intentionally left blank

FIBER OPTIC COMMUNICATION LINKS (TELECOM, DATACOM, AND ANALOG)

Casimer DeCusatis

*IBM Corporation
Poughkeepsie, New York*

Guifang Li

*CREOL, The College of Optics and Photonics
University of Central Florida
Orlando, Florida*

There are many different applications for fiber optic communication systems, each with its own unique performance requirements. For example, analog communication systems may be subject to different types of noise and interference than digital systems, and consequently require different figures of merit to characterize their behavior. At first glance, telecommunication and data communication systems appear to have much in common, as both use digital encoding of data streams; in fact, both types can share a common network infrastructure. Upon closer examination, however, we find important differences between them. First, datacom systems must maintain a much lower bit error rate (BER), defined as the number of transmission errors per second in the communication link (we will discuss BER in more detail in the following sections). For telecom (voice) communications, the ultimate receiver is the human ear and voice signals have a bandwidth of only about 4 kHz; transmission errors often manifest as excessive static noise such as encountered on a mobile phone, and most users can tolerate this level of fidelity. In contrast, the consequences of even a single bit error to a datacom system can be very serious; critical data such as medical or financial records could be corrupted, or large computer systems could be shut down. Typical telecom systems operate at a BER of about 10^{-9} , compared with about 10^{-12} to 10^{-15} for datacom systems. Another unique requirement of datacom systems is eye safety versus distance trade-offs. Most telecommunications equipment is maintained in a restricted environment and accessible only to personnel trained in the proper handling of high power optical sources. Datacom equipment is maintained in a computer center and must comply with international regulations for inherent eye safety; this limits the amount of optical power which can safely be launched into the fiber, and consequently limits the maximum distances which can be achieved without using repeaters or regenerators. For

the same reason, datacom equipment must be rugged enough to withstand casual use while telecom equipment is more often handled by specially trained service personell. Telecom systems also tend to make more extensive use of multiplexing techniques, which are only now being introduced into the data center, and more extensive use of optical repeaters.

In the following sections, we will examine the technical requirements for designing fiber optic communication systems suitable for these different environments. We begin by defining some figures of merit to characterize the system performance. Then, concentrating on digital optical communication systems, we will describe how to design an optical link loss budget and how to account for various types of noise sources in the link.

15.1 FIGURES OF MERIT

There are several possible figures of merit which may be used to characterize the performance of an optical communication system. Furthermore, different figures of merit may be more suitable for different applications, such as analog or digital transmission. In this section, we will describe some of the measurements used to characterize the performance of optical communication systems. Even if we ignore the practical considerations of laser eye safety standards, an optical transmitter is capable of launching a limited amount of optical power into a fiber; similarly, there is a limit as to how weak a signal can be detected by the receiver in the presence of noise and interference. Thus, a fundamental consideration in optical communication systems design is the optical link power budget, or the difference between the transmitted and received optical power levels. Some power will be lost due to connections, splices, and bulk attenuation in the fiber. There may also be optical power penalties due to dispersion, modal noise, or other effects in the fiber and electronics. The optical power levels define the signal-to-noise ratio (SNR) at the receiver, which is often used to characterize the performance of analog communication systems. For digital transmission, the most common figure of merit is the bit error rate (BER), defined as the ratio of received bit errors to the total number of transmitted bits. Signal-to-noise ratio is related to the bit error rate by the Gaussian integral

$$\text{BER} = \frac{1}{\sqrt{2\pi}} \int_Q^{\infty} e^{-Q^2/2} dQ \cong \frac{1}{Q\sqrt{2\pi}} e^{-Q^2/2} \quad (1)$$

where Q represents the SNR for simplicity of notation.¹⁻⁴ From Eq. (1), we see that a plot of BER versus received optical power yields a straight line on semilog scale, as illustrated in Fig. 1. Nominally, the slope is about 1.8 dB/decade; deviations from a straight line may indicate the presence of nonlinear or non-Gaussian noise sources. Some effects, such as fiber attenuation, are linear noise sources; they can be overcome by increasing the received optical power, as seen from Fig. 1, subject to constraints on maximum optical power (laser safety) and the limits of receiver sensitivity. There are other types of noise sources, such as mode partition noise or relative intensity noise (RIN), which are independent of signal strength. When such noise is present, no amount of increase in transmitted signal strength will affect the BER; a noise floor is produced, as shown by curve B in Fig. 1. This type of noise can be a serious limitation on link performance. If we plot BER versus receiver sensitivity for increasing optical power, we obtain a curve similar to Fig. 2 which shows that for very high power levels, the receiver will go into saturation. The characteristic “bathtub”-shaped curve illustrates a window of operation with both upper and lower limits on the received power. There may also be an upper limit on optical power due to eye safety considerations.

We can see from Fig. 1 that receiver sensitivity is specified at a given BER, which is often too low to measure directly in a reasonable amount of time (e.g., a 200 Mbit/s link operating at a BER of 10^{-15} will only take one error every 57 days on average, and several hundred errors are recommended for a reasonable BER measurement). For practical reasons, the BER is typically measured at much higher error rates, where the data can be collected more quickly (such as 10^{-4} to 10^{-8}) and then extrapolated

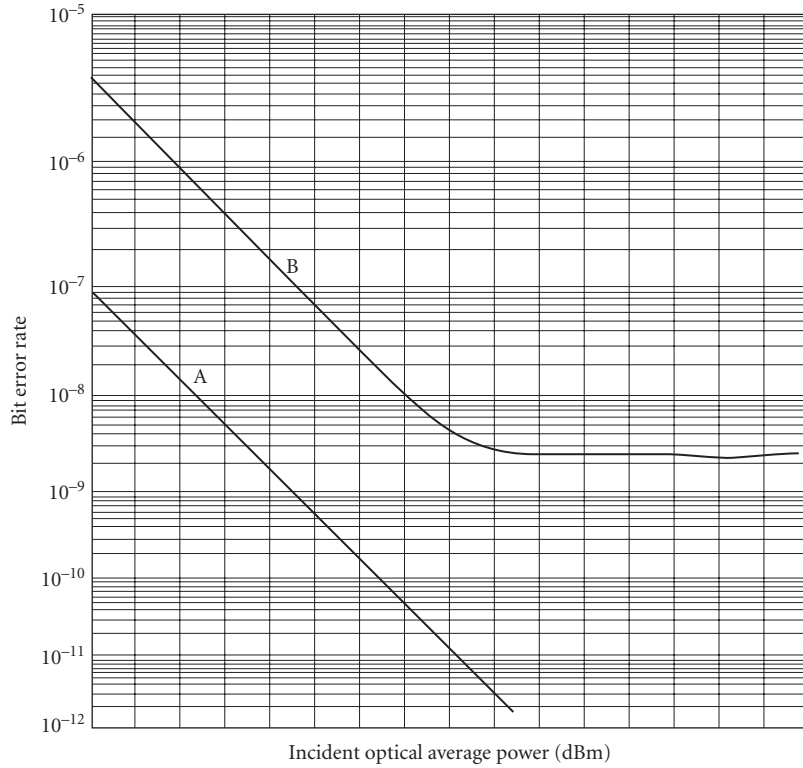


FIGURE 1 Bit error rate as a function of received optical power. Curve A shows typical performance, whereas curve B shows a BER floor.⁵

to find the sensitivity at low BER. This assumes the absence of nonlinear noise floors, as cautioned previously. The relationship between optical input power, in watts, and the BER, is the complimentary Gaussian error function

$$\text{BER} = 1/2 \operatorname{erfc} (P_{\text{out}} - P_{\text{signal}} / \text{RMS noise}) \quad (2)$$

where the error function is an open integral that cannot be solved directly. Several approximations have been developed for this integral, which can be developed into transformation functions that yield a linear least squares fit to the data.¹ The same curve fitting equations can also be used to characterize the eye window performance of optical receivers. Clock position/phase versus BER data are collected for each edge of the eye window; these data sets are then curve fitted with the above expressions to determine the clock position at the desired BER. The difference in the two resulting clock position on either side of the window gives the clear eye opening.¹⁻⁴

In describing Figs. 1 and 2, we have also made some assumptions about the receiver circuit. Most data links are asynchronous, and do not transmit a clock pulse along with the data; instead, a clock is extracted from the incoming data and used to retiming the received data stream. We have made the assumption that the BER is measured with the clock at the center of the received data bit; ideally, this is when we compare the signal with a preset threshold to determine if a logical “1” or “0” was sent. When the clock is recovered from a receiver circuit such as a phase lock loop, there is always some uncertainty about the clock position; even if it is centered on the data bit, the relative clock position

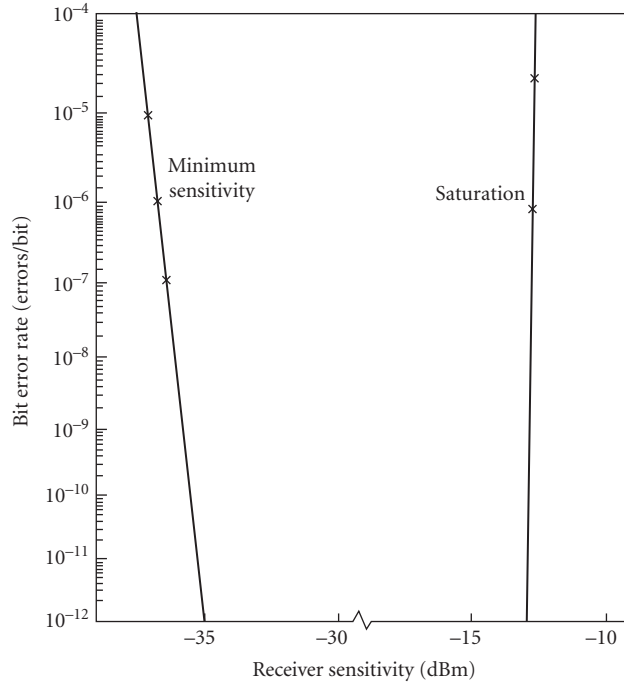


FIGURE 2 Bit error rate as a function of received optical power illustrating range of operation from minimum sensitivity to saturation.

may drift over time. The region of the bit interval in the time domain where the BER is acceptable is called the eyewidth; if the clock timing is swept over the data bit using a delay generator, the BER will degrade near the edges of the eye window. Eyewidth measurements are an important parameter in link design, which will be discussed further in the section on jitter and link budget modeling.

In the design of some analog optical communication systems, as well as some digital television systems (e.g., those based on 64-bit Quadrature Amplitude Modulation), another possible figure of merit is the modulation error ratio (MER). To understand this metric, we will consider the standard definition of the Digital Video Broadcasters (DVB) Measurements Group.⁵ First, the video receiver captures a time record of N received signal coordinate pairs, representing the position of information on a two-dimensional screen. The ideal position coordinates are given by the vector (X_j, Y_j) . For each received symbol, a decision is made as to which symbol was transmitted, and an error vector $(\Delta X_j, \Delta Y_j)$ is defined as the distance from the ideal position to the actual position of the received symbol. The MER is then defined as the sum of the squares of the magnitudes of the ideal symbol vector divided by the sum of the squares of the magnitudes of the symbol error vectors:

$$\text{MER} = 10 \log \frac{\sum_{j=1}^N (X_j^2 + Y_j^2)}{\sum_{j=1}^N (\Delta X_j^2 + \Delta Y_j^2)} \text{dB} \quad (3)$$

when the signal vectors are corrupted by noise, they can be treated as random variables. The denominator in Eq. (3) becomes an estimate of the average power of the error vector (in other words, its second moment) and contains all signal degradation due to noise, reflections, transmitter quadrature errors, etc. If the only significant source of signal degradation is additive white Gaussian noise, then MER and SNR are equivalent. For communication systems which contain other noise sources, MER offers some advantages; in particular, for some digital transmission systems there may be a

very sharp change in BER as a function of SNR (a so-called “cliff effect”) which means that BER alone cannot be used as an early predictor of system failures. MER, on the other hand, can be used to measure signal-to-interference ratios accurately for such systems. Because MER is a statistical measurement, its accuracy is directly related to the number of vectors N used in the computation; an accuracy of 0.14 dB can be obtained with $N = 10,000$, which would require about 2 ms to accumulate at the industry standard digital video rate of 5.057 Msymbols/s.

In order to design a proper optical data link, the contribution of different types of noise sources should be assessed when developing a link budget. There are two basic approaches to link budget modeling. One method is to design the link to operate at the desired BER when all the individual link components assume their worst case performance. This conservative approach is desirable when very high performance is required, or when it is difficult or inconvenient to replace failing components near the end of their useful lifetimes. The resulting design has a high safety margin; in some cases, it may be overdesigned for the required level of performance. Since it is very unlikely that all the elements of the link will assume their worst case performance at the same time, an alternative is to model the link budget statistically. For this method, distributions of transmitter power output, receiver sensitivity, and other parameters are either measured or estimated. They are then combined statistically using an approach such as the Monte Carlo method, in which many possible link combinations are simulated to generate an overall distribution of the available link optical power. A typical approach is the 3-sigma design, in which the combined variations of all link components are not allowed to extend more than 3 standard deviations from the average performance target in either direction. The statistical approach results in greater design flexibility, and generally increased distance compared with a worst-case model at the same BER.

Harmonic Distortions, Intermodulation Distortions, and Dynamic Range

Fiber-optic analog links are in general nonlinear. That is, if the input electrical information is a harmonic signal of frequency f_0 , the output electrical signal will contain the fundamental frequency f_0 as well as high-order harmonics of frequencies nf_0 ($n > 2$). These high-order harmonics comprise the harmonic distortions of analog fiber-optic links.⁶ The nonlinear behavior is caused by nonlinearities in the transmitter, the fiber, and the receiver. The same sources of nonlinearities in the fiber-optic links lead to intermodulation distortions (IMD), which can be best illustrated in a two-tone transmission scenario. If the input electrical information is a superposition of two harmonic signals of frequencies f_1 and f_2 , the output electrical signal will contain second-order intermodulation at frequencies $f_1 + f_2$ and $f_1 - f_2$ as well as third-order intermodulation at frequencies $2f_1 - f_2$ and $2f_2 - f_1$.

Most analog fiber-optic links require bandwidth of less than one octave ($f_{\max} < 2f_{\min}$). As a result harmonic distortions as well as second-order IMD products are not important as they can be filtered out electronically. However, third-order IMD products are in the same frequency range (between $f_{\min}$ and $f_{\max}$) as the signal itself and therefore appear in the output signal as the spurious response. Thus the linearity of analog fiber-optic links is determined by the level of third-order IMD products. In the case of analog links where third-order IMD is eliminated through linearization circuitry, the lowest odd-order IMD determines the linearity of the link.

To quantify IMD distortions, a two-tone experiment (or simulation) is usually conducted where the input RF powers of the two tones are equal. The linear and nonlinear power transfer functions—the output RF power of each of two input tones and the second or third-order IMD product as a function of the input RF power of each input harmonic signal—are schematically presented in Fig. 3. When plotted on a log-log scale, the fundamental power transfer function should be a line with a slope of unity. The second- (third-) order power transfer function should be a line with a slope of two (three). The intersections of the power transfer functions are called second- and third-order intercept points, respectively. Because of the fixed slopes of the power transfer functions, the intercept points can be calculated from measurements obtained at a single input power level. Suppose at a certain input level, the output power of each of the two fundamental tones, the second-order IMD product and third-order

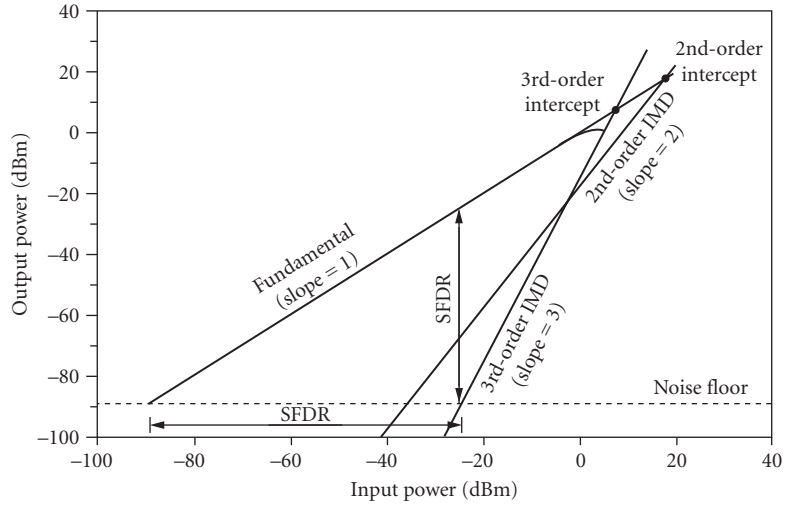


FIGURE 3 Intermodulation and dynamic range of analog fiberoptic links.

IMD products are P_1 , P_2 , and P_3 , respectively. When the power levels are in units of dB or dBm, the second-order and third-order intercept points are

$$IP_2 = 2P_1 - P_2 \quad (4)$$

and

$$IP_3 = (3P_1 - P_3)/2 \quad (5)$$

The dynamic range is a measure of the ability of an analog fiber-optic link to faithfully transmit signals at various power levels. At the low input power end, the analog link can fail due to insufficient power level so that the output power is below the noise level. At the high input power end, the analog link can fail due to the fact that the IMD products become the dominant source of signal degradation. In term of the output power, the dynamic range (of the output power) is defined as the ratio of the fundamental output to the noise power. However, it should be noted that the third-order IMD products increase three times faster than the fundamental signal. After the third-order IMD products exceeds the noise floor, the ratio of the fundamental output to the noise power is meaningless as the dominant degradation of the output signal comes from IMD products. So a more meaningful definition of the dynamic range is the so-called spurious-free dynamic range (SFDR),^{6,7} which is the ratio of the fundamental output to the noise power at the point where the IMD products is at the noise level. The spurious-free dynamic range is then practically the maximum dynamic range. Since the noise floor depends on the bandwidth of interest, the unit for SFDR should be $\text{dB-Hz}^{2/3}$. The dynamic range decreases as the bandwidth of the system is increased. The spurious-free dynamic range is also often defined with reference to the input power, which corresponds to SFDR with reference to the output power if there is no gain compression.

15.2 LINK BUDGET ANALYSIS: INSTALLATION LOSS

It is convenient to break down the link budget into two areas: installation loss and available power. Installation or DC loss refers to optical losses associated with the fiber cable plant, such as connector loss, splice loss, and bandwidth considerations. Available optical power is the difference between the

transmitter output and receiver input powers, minus additional losses due to optical noise sources on the link (also known as AC losses). With this approach, the installation loss budget may be treated statistically and the available power budget as worst case. First, we consider the installation loss budget, which can be broken down into three areas, namely, transmission loss, fiber attenuation as a function of wavelength, and connector or splice losses.

Transmission Loss

Transmission loss is perhaps the most important property of an optical fiber; it affects the link budget and maximum unrepeated distance. Since the maximum optical power launched into an optical fiber is determined by international laser eye safety standards,⁸ the number and separation between optical repeaters and regenerators is largely determined by this loss. The mechanisms responsible for this loss include material absorption as well as both linear and nonlinear scattering of light from impurities in the fiber.¹⁻⁵ Typical loss for single-mode optical fiber is about 2 to 3 dB/km near 800-nm wavelength, 0.5 dB/km near 1300 nm, and 0.25 dB/km near 1550 nm. Multimode fiber loss is slightly higher, and bending loss will only increase the link attenuation further.

Attenuation versus Wavelength

Since fiber loss varies with wavelength, changes in the source wavelength or use of sources with a spectrum of wavelengths will produce additional loss. Transmission loss is minimized near the 1550-nm wavelength band, which unfortunately does not correspond with the dispersion minimum at around 1310 nm. An accurate model for fiber loss as a function of wavelength has been developed by Walker;⁹ this model accounts for the effects of linear scattering, macrobending, and material absorption due to ultraviolet and infrared band edges, hydroxide (OH) absorption, and absorption from common impurities such as phosphorous. Using this model, it is possible to calculate the fiber loss as a function of wavelength for different impurity levels; the fiber properties can be specified along with the acceptable wavelength limits of the source to limit the fiber loss over the entire operating wavelength range. Design tradeoffs are possible between center wavelength and fiber composition to achieve the desired result. Typical loss due to wavelength-dependent attenuation for laser sources on single-mode fiber can be held below 0.1 dB/km.

Connector and Splice Losses

There are also installation losses associated with fiber optic connectors and splices; both of these are inherently statistical in nature and can be characterized by a Gaussian distribution. There are many different kinds of standardized optical connectors, some of which have been discussed previously; some industry standards also specify the type of optical fiber and connectors suitable for a given application.¹⁰ There are also different models which have been published for estimating connection loss due to fiber misalignment;^{11,12} most of these treat loss due to misalignment of fiber cores, offset of fibers on either side of the connector, and angular misalignment of fibers. The loss due to these effects is then combined into an overall estimate of the connector performance. There is no general model available to treat all types of connectors, but typical connector loss values average approximately 0.5 dB worst case for multimode, slightly higher for singlemode (see Table 1).

Optical splices are required for longer links, since fiber is usually available in spools of 1 to 5 km, or to repair broken fibers. There are two basic types, mechanical splices (which involve placing the two fiber ends in a receptacle that holds them close together, usually with epoxy) and the more commonly used fusion splices (in which the fiber are aligned, then heated sufficiently to fuse the two ends together). Typical splice loss values are given in Table 1.

TABLE 1 Typical Cable Plant Optical Losses⁵

Component	Description	Size (μm)	Mean Loss	Variance (dB^2)
Connector*	Physical contact	62.5–62.5	0.40 dB	0.02
		50.0–50.0	0.40 dB	0.02
		9.0–9.0 [†]	0.35 dB	0.06
		62.5–50.0	2.10 dB	0.12
Connector*	Nonphysical contact (multimode only)	50.0–62.5	0.00 dB	0.01
		62.5–62.5	0.70 dB	0.04
		50.0–50.0	0.70 dB	0.04
		62.5–50.0	2.40 dB	0.12
Splice	Mechanical	50.0–62.5	0.30 dB	0.01
		62.5–62.5	0.15 dB	0.01
		50.0–50.0	0.15 dB	0.01
Splice	Fusion	9.0–9.0 [†]	0.15 dB	0.01
		62.5–62.5	0.40 dB	0.01
		50.0–50.0	0.40 dB	0.01
Cable	IBM multimode jumper	9.0–9.0 [†]	0.40 dB	0.01
	IBM multimode jumper	62.5	1.75 dB/km	NA
	IBM multimode jumper	50.0	3.00 dB/km at 850 nm	NA
	IBM single-mode jumper	9.0	0.8 dB/km	NA
	Trunk	62.5	1.00 dB/km	NA
	Trunk	50.0	0.90 dB/km	NA
	Trunk	9.0	0.50 dB/km	NA

*The connector loss value is typical when attaching identical connectors. The loss can vary significantly if attaching different connector types.

[†]Single-mode connectors and splices must meet a minimum return loss specification of 28 dB.

15.3 LINK BUDGET ANALYSIS: OPTICAL POWER PENALTIES

Next, we will consider the assembly loss budget, which is the difference between the transmitter output and receiver input powers, allowing for optical power penalties due to noise sources in the link. We will follow the standard convention in the literature of assuming a digital optical communication link which is best characterized by its BER. Contributing factors to link performance include the following:

- Dispersion (modal and chromatic) or intersymbol interference
- Mode partition noise
- Mode hopping
- Extinction ratio
- Multipath interference
- Relative intensity noise (RIN)
- Timing jitter
- Radiation induced darkening
- Modal noise

Higher order, nonlinear effects including Stimulated Raman and Brillouin scattering and frequency chirping will be discussed elsewhere.

Dispersion

The most important fiber characteristic after transmission loss is dispersion, or intersymbol interference. This refers to the broadening of optical pulses as they propagate along the fiber. As pulses broaden, they tend to interfere with adjacent pulses; this limits the maximum achievable data rate. In multimode fibers, there are two dominant kinds of dispersion, modal and chromatic. Modal dispersion refers to the fact that different modes will travel at different velocities and cause pulse broadening. The fiber's modal bandwidth in units of MHz-km, is specified according to the expression

$$BW_{\text{modal}} = BW_1 / L^\gamma \quad (6)$$

where BW_{modal} is the modal bandwidth for a length L of fiber, BW_1 is the manufacturer-specified modal bandwidth of a 1-km section of fiber, and γ is a constant known as the modal bandwidth concatenation length scaling factor. The term γ usually assumes a value between 0.5 and 1, depending on details of the fiber manufacturing and design as well as the operating wavelength; it is conservative to take $\gamma = 1.0$. Modal bandwidth can be increased by mode mixing, which promotes the interchange of energy between modes to average out the effects of modal dispersion. Fiber splices tend to increase the modal bandwidth, although it is conservative to discard this effect when designing a link.

The other major contribution is chromatic dispersion BW_{chrom} which occurs because different wavelengths of light propagate at different velocities in the fiber. For multimode fiber, this is given by an empirical model of the form

$$BW_{\text{chrom}} = \frac{L^{\gamma_c}}{\sqrt{\lambda_w} (a_o + a_1 |\lambda_c - \lambda_{\text{eff}}|)} \quad (7)$$

where L is the fiber length in km; λ_c is the center wavelength of the source in nm; λ_w is the source FWHM spectral width in nm; γ_c is the chromatic bandwidth length scaling coefficient, a constant; λ_{eff} is the effective wavelength, which combines the effects of the fiber zero dispersion wavelength and spectral loss signature; and the constants a_1 and a_o are determined by a regression fit of measured data. From Ref. 13, the chromatic bandwidth for 62.5/125- μm fiber is empirically given by

$$BW_{\text{chrom}} = \frac{10^4 L^{-0.69}}{\sqrt{\lambda_w} (1.1 + 0.0189 |\lambda_c - 1370|)} \quad (8)$$

For this expression, the center wavelength was 1335 nm and λ_{eff} was chosen midway between λ_c and the water absorption peak at 1390 nm; although λ_{eff} was estimated in this case, the expression still provides a good fit to the data. For 50/125- μm fiber, the expression becomes

$$BW_{\text{chrom}} = \frac{10^4 L^{-0.65}}{\sqrt{\lambda_w} (1.01 + 0.0177 |\lambda_c - 1330|)} \quad (9)$$

For this case, λ_c was 1313 nm and the chromatic bandwidth peaked at $\lambda_{\text{eff}} = 1330$ nm. Recall that this is only one possible model for fiber bandwidth.¹ The total bandwidth capacity of multimode fiber BW_t is obtained by combining the modal and chromatic dispersion contributions, according to

$$\frac{1}{BW_t^2} = \frac{1}{BW_{\text{chrom}}^2} + \frac{1}{BW_{\text{modal}}^2} \quad (10)$$

Once the total bandwidth is known, the dispersion penalty can be calculated for a given data rate. One expression for the dispersion penalty in decibel is

$$P_d = 1.22 \left[\frac{\text{bit rate (Mb/s)}}{BW_t \text{ (MHz)}} \right]^2 \quad (11)$$

For typical telecommunication grade fiber, the dispersion penalty for a 20-km link is about 0.5 dB.

Dispersion is usually minimized at wavelengths near 1310 nm; special types of fiber have been developed which manipulate the index profile across the core to achieve minimal dispersion near 1550 nm, which is also the wavelength region of minimal transmission loss. Unfortunately, this dispersion-shifted fiber suffers from some practical drawbacks, including susceptibility to certain kinds of nonlinear noise and increased interference between adjacent channels in a wavelength multiplexing environment. There is a new type of fiber which minimizes dispersion while reducing the unwanted crosstalk effects, called dispersion optimized fiber. By using a very sophisticated fiber profile, it is possible to minimize dispersion over the entire wavelength range from 1300 nm to 1550 nm, at the expense of very high loss (around 2 dB/km); this is known as dispersion flattened fiber. Yet another approach is called dispersion compensating fiber; this fiber is designed with negative dispersion characteristics, so that when used in series with conventional fiber it will offset the normal fiber dispersion. Dispersion compensating fiber has a much narrower core than standard singlemode fiber, which makes it susceptible to nonlinear effects; it is also birefringent and suffers from polarization mode dispersion, in which different states of polarized light propagate with very different group velocities. Note that standard singlemode fiber does not preserve the polarization state of the incident light; there is yet another type of specialty fiber, with asymmetric core profiles, capable of preserving the polarization of incident light over long distances.

By definition, single-mode fiber does not suffer modal dispersion. Chromatic dispersion is an important effect, though, even given the relatively narrow spectral width of most laser diodes. The dispersion of single-mode fiber corresponds to the first derivative of group velocity τ_g with respect to wavelength, and is given by

$$D = \frac{d\tau_g}{d\lambda} = \frac{S_o}{4} \left(\lambda_c - \frac{\lambda_o^4}{\lambda_c^3} \right) \quad (12)$$

where D is the dispersion in ps/(km-nm) and λ_c is the laser center wavelength. The fiber is characterized by its zero dispersion wavelength, λ_o , and zero dispersion slope, S_o . Usually, both center wavelength and zero dispersion wavelength are specified over a range of values; it is necessary to consider both upper and lower bounds in order to determine the worst case dispersion penalty. This can be seen from Fig. 4 which plots D versus wavelength for some typical values of λ_o and λ_c ; the largest absolute value of D occurs at the extremes of this region. Once the dispersion is determined,

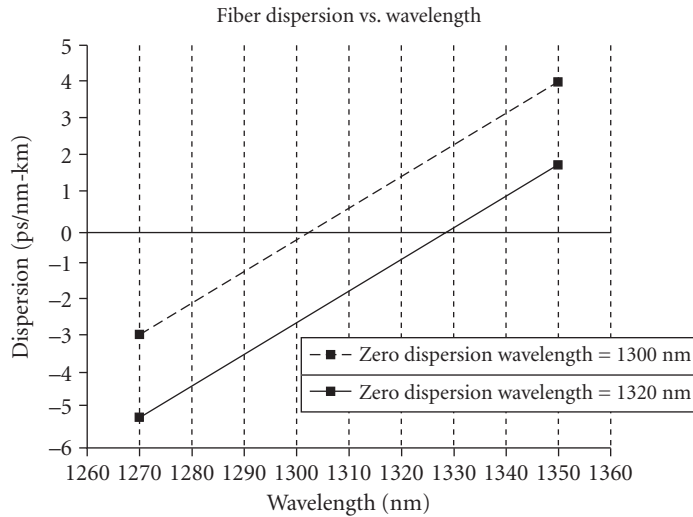


FIGURE 4 Single-mode fiber dispersion as a function of wavelength.⁵

the intersymbol interference penalty as a function of link length L can be determined to a good approximation from a model proposed by Agrawal:¹⁴

$$P_d = 5 \log(1 + 2\pi(BD\Delta\lambda)^2 L^2) \quad (13)$$

where B is the bit rate and $\Delta\lambda$ is the root-mean-square (RMS) spectral width of the source. By maintaining a close match between the operating and zero dispersion wavelengths, this penalty can be kept to a tolerable 0.5 to 1.0 dB in most cases.

Mode Partition Noise

Group velocity dispersion contributes to another optical penalty, which remains the subject of continuing research, mode partition noise and mode hopping. This penalty is related to the properties of a Fabry-Perot type laser diode cavity; although the total optical power output from the laser may remain constant, the optical power distribution among the laser's longitudinal modes will fluctuate. This is illustrated by the model depicted in Fig. 5; when a laser diode is directly modulated with

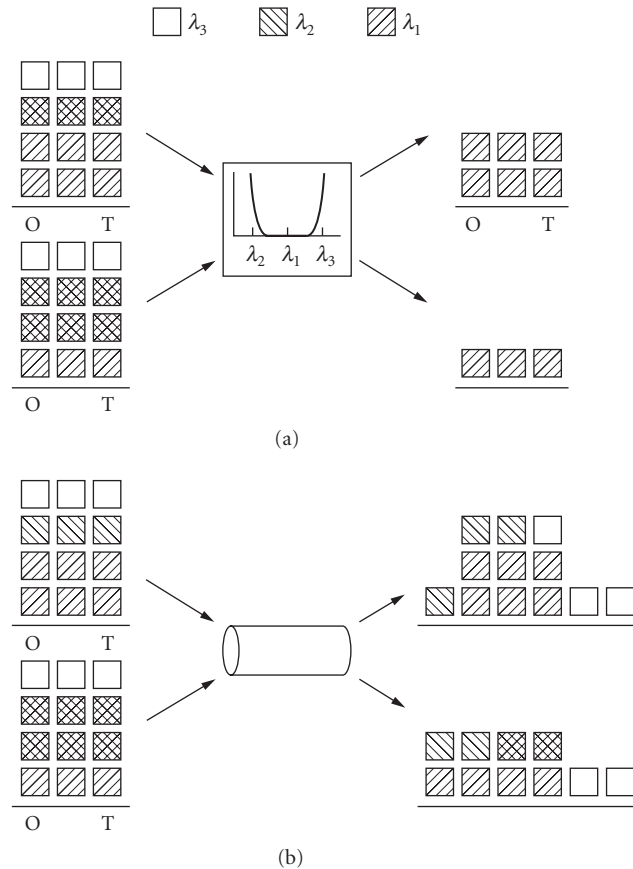


FIGURE 5 Model for mode partition noise; an optical source emits a combination of wavelengths, illustrated by different color blocks: (a) wavelength-dependent loss and (b) chromatic dispersion.

injection current, the total output power stays constant from pulse to pulse; however, the power distribution among several longitudinal modes will vary between pulses. We must be careful to distinguish this behavior of the instantaneous laser spectrum, which varies with time, from the time-averaged spectrum which is normally observed experimentally. The light propagates through a fiber with wavelength-dependent dispersion or attenuation, which deforms the pulse shape. Each mode is delayed by a different amount due to group velocity dispersion in the fiber; this leads to additional signal degradation at the receiver, in addition to the intersymbol interference caused by chromatic dispersion alone, discussed earlier. This is known as mode partition noise; it is capable of generating bit error rate floors, such that additional optical power into the receiver will not improve the link BER. This is because mode partition noise is a function of the laser spectral fluctuations and wavelength-dependent dispersion of the fiber, so the signal-to-noise ratio due to this effect is independent of the signal power. The power penalty due to mode partition noise was first calculated by Ogawa¹⁵ as

$$P_{\text{mp}} = 5 \log(1 - Q^2 \sigma_{\text{mp}}^2) \quad (14)$$

where

$$\sigma_{\text{mp}}^2 = \frac{1}{2} k^2 (\pi B)^4 [A_1^4 \Delta \lambda^4 + 42 A_1^2 A_2^2 \Delta \lambda^6 + 48 A_2^4 \Delta \lambda^8] \quad (15)$$

$$A_1 = DL \quad (16)$$

and

$$A_2 = \frac{A_1}{2(\lambda_c - \lambda_o)} \quad (17)$$

The mode partition coefficient k is a number between 0 and 1 which describes how much of the optical power is randomly shared between modes; it summarizes the statistical nature of mode partition noise. According to Ogawa, k depends on the number of interacting modes and rms spectral width of the source, the exact dependence being complex. However, subsequent work has shown¹⁶ that Ogawa's model tends to underestimate the power penalty due to mode partition noise because it does not consider the variation of longitudinal mode power between successive baud periods, and because it assumes a linear model of chromatic dispersion rather than the nonlinear model given in the above equation. A more detailed model has been proposed by Campbell,¹⁷ which is general enough to include effects of the laser diode spectrum, pulse shaping, transmitter extinction ratio, and statistics of the data stream. While Ogawa's model assumed an equiprobable distribution of zeros and ones in the data stream, Campbell showed that mode partition noise is data dependent as well. Recent work based on this model¹⁸ has re-derived the signal variance:

$$\sigma_{\text{mp}}^2 = E_{\text{av}} (\sigma_o^2 + \sigma_{+1}^2 + \sigma_{-1}^2) \quad (18)$$

where the mode partition noise contributed by adjacent baud periods is defined by

$$\sigma_{+1}^2 + \sigma_{-1}^2 = \frac{1}{2} k^2 (\pi B)^4 [1.25 A_1^4 \Delta \lambda^4 + 40.95 A_1^2 A_2^2 \Delta \lambda^6 + 50.25 A_2^4 \Delta \lambda^8] \quad (19)$$

and the time-average extinction ratio $E_{\text{av}} = 10 \log(P_i/P_o)$, where P_i, P_o represent the optical power by a 1 and 0, respectively. If the operating wavelength is far away from the zero dispersion wavelength, the noise variance simplifies to

$$\sigma_{\text{mp}}^2 = 2.25 \frac{k^2}{2} E_{\text{av}} (1 - e^{-\beta L^2})^2 \quad (20)$$

which is valid provided that,

$$\beta = (\pi BD\Delta\lambda)^2 \ll 1 \quad (21)$$

Many diode lasers exhibit mode hopping or mode splitting in which the spectrum appears to split optical power between 2 or 3 modes for brief periods of time. The exact mechanism is not fully understood, but stable Gaussian spectra are generally only observed for CW operation and temperature stabilized lasers. During these mode hops the above theory does not apply since the spectrum is non-Gaussian, and the model will overpredict the power penalty; hence, it is not possible to model mode hops as mode partitioning with $k = 1$. There is no currently published model describing a treatment of mode hopping noise, although recent papers¹⁹ suggest approximate calculations based on the statistical properties of the laser cavity. In a practical link, some amount of mode hopping is probably unavoidable as a contributor to burst noise; empirical testing of link hardware remains the only reliable way to reduce this effect. A practical rule of thumb is to keep the mode partition noise penalty less than 1.0 dB maximum, provided that this penalty is far away from any noise floors.

Extinction Ratio

The receiver extinction ratio also contributes directly to the link penalties. The receiver BER is a function of the modulated AC signal power; if the laser transmitter has a small extinction ratio, the DC component of total optical power is significant. Gain or loss can be introduced in the link budget if the extinction ratio at which the receiver sensitivity is measured differs from the worst case transmitter extinction ratio. If the extinction ratio E_t at the transmitter is defined as the ratio of optical power when a one is transmitted versus when a zero is transmitted,

$$E_t = \frac{\text{Power}(1)}{\text{Power}(0)} \quad (22)$$

then we can define a modulation index at the transmitter M_t according to

$$M_t = \frac{E_t - 1}{E_t + 1} \quad (23)$$

Similarly, we can measure the linear extinction ratio at the optical receiver input and define a modulation M_r . The extinction ratio penalty is given by

$$P_{\text{er}} = -10 \log \left(\frac{M_t}{M_r} \right) \quad (24)$$

where the subscripts T and R refer to specifications for the transmitter and receiver, respectively. Usually, the extinction ratio is specified to be the same at the transmitter and receiver, and is large enough so that there is no power penalty due to extinction ratio effects.

Multipath Interference

Another important property of the optical link is the amount of reflected light from the fiber end-faces which returns up the link back into the transmitter. Whenever there is a connection or splice in the link, some fraction of the light is reflected back; each connection is thus a potential noise generator, since the reflected fields can interfere with one another to create noise in the detected optical signal. The phenomenon is analogous to the noise caused by multiple atmospheric reflections of radio waves, and is known as *multipath interference noise*. To limit this noise, connectors and splices are specified with a minimum return loss. If there are a total of N reflection points in a link and the geometric mean of the connector reflections is alpha, then based on the model of Duff et al.²⁰ the

power penalty due to multipath interference (adjusted for bit error rate and bandwidth) is closely approximated by

$$P_{\text{mpi}} = 10 \log(1 - 0.7Na) \quad (25)$$

Multipath noise can usually be reduced well below 0.5 dB with available connectors, whose return loss is often better than 25 dB.

Relative Intensity Noise

Stray light reflected back into a Fabry-Perot type laser diode gives rise to intensity fluctuations in the laser output. This is a complicated phenomena, strongly dependent on the type of laser; it is called either reflection-induced intensity noise or relative intensity noise (RIN). This effect is important since it can also generate BER floors. The power penalty due to RIN is the subject of ongoing research; since the reflected light is measured at a specified signal level, RIN is data dependent although it is independent of link length. Since many laser diodes are packaged in windowed containers, it is difficult to correlate the RIN measurements on an unpackaged laser with those of a commercial product. There have been several detailed attempts to characterize RIN;^{21,22} typically, the RIN noise is assumed Gaussian in amplitude and uniform in frequency over the receiver bandwidth of interest. The RIN value is specified for a given laser by measuring changes in the optical power when a controlled amount of light is fed back into the laser; it is signal dependent, and is also influenced by temperature, bias voltage, laser structure, and other factors which typically influence laser output power.²² If we assume that the effect of RIN is to produce an equivalent noise current at the receiver, then the additional receiver noise σ_r may be modeled as

$$\sigma_r = \gamma^2 S^2 g B \quad (26)$$

where S is the signal level during a bit period, B is the bit rate, and g is a noise exponent which defines the amount of signal-dependent noise. If $g = 0$, noise power is independent of the signal, while for $g = 1$ noise power is proportional to the square of the signal strength. The coefficient γ is given by

$$\gamma^2 = S_i^{2(1-g)} 10^{(\text{RIN}_i/10)} \quad (27)$$

where RIN_i is the measured RIN value at the average signal level S_i , including worst case backreflection conditions and operating temperatures. The Gaussian BER probability due to the additional RIN noise current is given by

$$P_{\text{error}} = \frac{1}{2} \left[P_e^1 \left(\frac{S_1 - S_o}{2\sigma_1} \right) + P_e^0 \left(\frac{S_1 - S_o}{2\sigma_o} \right) \right] \quad (28)$$

where σ_1, σ_o represent the total noise current during transmission of a digital 1 and 0, respectively, and P_e^1, P_e^0 are the probabilities of error during transmission of a 1 or 0, respectively. The power penalty due to RIN may then be calculated by determining the additional signal power required to achieve the same BER with RIN noise present as without the RIN contribution. One approximation for the RIN power penalty is given by

$$P_{\text{rin}} = -5 \log \left[1 - Q^2(\text{BW})(1 + M_r)^{2g} (10^{\text{RIN}/10}) \left(\frac{1}{M_r} \right)^2 \right] \quad (29)$$

where the RIN value is specified in dB/Hz, BW is the receiver bandwidth, M_r is the receiver modulation index, and the exponent g is a constant varying between 0 and 1 which relates the magnitude of RIN noise to the optical power level. The maximum RIN noise penalty in a link can usually be kept below 0.5 dB.

Jitter

Although it is not strictly an optical phenomena, another important area in link design deals with the effects of timing jitter on the optical signal. In a typical optical link, a clock is extracted from the incoming data signal which is used to retiming and reshape the received digital pulse; the received pulse is then compared with a threshold to determine if a digital 1 or 0 was transmitted. So far, we have discussed BER testing with the implicit assumption that the measurement was made in the center of the received data bit; to achieve this, a clock transition at the center of the bit is required. When the clock is generated from a receiver timing recovery circuit, it will have some variation in time and the exact location of the clock edge will be uncertain. Even if the clock is positioned at the center of the bit, its position may drift over time. There will be a region of the bit interval, or eye, in the time domain where the BER is acceptable; this region is defined as the eyewidth.¹⁻³ Eyewidth measurements are an important parameter for evaluation of fiber optic links; they are intimately related to the BER, as well as the acceptable clock drift, pulse width distortion, and optical power. At low optical power levels, the receiver signal-to-noise ratio is reduced; increased noise causes amplitude variations in the received signal. These amplitude variations are translated into time domain variations in the receiver decision circuitry, which narrows the eyewidth. At the other extreme, an optical receiver may become saturated at high optical power, reducing the eyewidth and making the system more sensitive to timing jitter. This behavior results in the typical “bathtub” curve shown in Fig. 2; for this measurement, the clock is delayed from one end of the bit cell to the other, with the BER calculated at each position. Near the ends of the cell, a large number of errors occur; toward the center of the cell, the BER decreases to its true value. The eye opening may be defined as the portion of the eye for which the BER remains constant; pulse width distortion occurs near the edges of the eye, which denotes the limits of the valid clock timing. Uncertainty in the data pulse arrival times causes errors to occur by closing the eye window and causing the eye pattern to be sampled away from the center. This is one of the fundamental problems of optical and digital signal processing, and a large body of work has been done in this area.^{23,24} In general, multiple jitter sources will be present in a link; these will tend to be uncorrelated. However, jitter on digital signals, especially resulting from a cascade of repeaters, may be coherent.

International standards on jitter were first published by the CCITT (Central Commission for International Telephony and Telegraphy, now known as the International Telecommunications Union, or ITU). This standards body has adopted a definition of jitter²⁴ as short-term variations of the significant instants (rising or falling edges) of a digital signal from their ideal position in time. Longer-term variations are described as wander; in terms of frequency, the distinction between jitter and wander is somewhat unclear. The predominant sources of jitter include the following:

- Phase noise in receiver clock recovery circuits, particularly crystal-controlled oscillator circuits; this may be aggravated by filters or other components which do not have a linear phase response. Noise in digital logic resulting from restricted rise and fall times may also contribute to jitter.
- Imperfect timing recovery in digital regenerative repeaters, which is usually dependent on the data pattern.
- Different data patterns may contribute to jitter when the clock recovery circuit of a repeater attempts to recover the receive clock from inbound data. Data pattern sensitivity can produce as much as 0.5-dB penalty in receiver sensitivity. Higher data rates are more susceptible (>1 Gbit/s); data patterns with long run lengths of 1s or 0s, or with abrupt phase transitions between consecutive blocks of 1s and 0s, tend to produce worst case jitter.
- At low optical power levels, the receiver signal-to-noise ratio Q is reduced; increased noise causes amplitude variations in the signal, which may be translated into time domain variations by the receiver circuitry.
- Low frequency jitter, also called wander, resulting from instabilities in clock sources and modulation of transmitters.
- Very low frequency jitter caused by variations in the propagation delay of fibers, connectors, etc., typically resulting from small temperature variations. (This can make it especially difficult to perform long-term jitter measurements.)

In general, jitter from each of these sources will be uncorrelated; jitter related to modulation components of the digital signal may be coherent, and cumulative jitter from a series of repeaters or regenerators may also contain some well correlated components.

There are several parameters of interest in characterizing jitter performance. Jitter may be classified as either random or deterministic, depending on whether it is associated with pattern-dependent effects; these are distinct from the duty cycle distortion which often accompanies imperfect signal timing. Each component of the optical link (data source, serializer, transmitter, encoder, fiber, receiver, retiming/clock recovery/deserialization, decision circuit) will contribute some fraction of the total system jitter. If we consider the link to be a “black box” (but not necessarily a linear system) then we can measure the level of output jitter in the absence of input jitter; this is known as the “intrinsic jitter” of the link. The relative importance of jitter from different sources may be evaluated by measuring the spectral density of the jitter. Another approach is the maximum tolerable input jitter (MTIJ) for the link. Finally, since jitter is essentially a stochastic process, we may attempt to characterize the jitter transfer function (JTF) of the link, or estimate the probability density function of the jitter. When multiple traces occur at the edges of the eye, this can indicate the presence of data dependent jitter or duty cycle distortion; a histogram of the edge location will show several distinct peaks. This type of jitter can indicate a design flaw in the transmitter or receiver. By contrast, random jitter typically has a more Gaussian profile and is present to some degree in all data links.

The problem of jitter accumulation in a chain of repeaters becomes increasingly complex; however, we can state some general rules of thumb. It has been shown²⁵ that jitter can be generally divided into two components, one due to repetitive patterns and one due to random data. In receivers with phase-lock loop timing recovery circuits, repetitive data patterns will tend to cause jitter accumulation, especially for long run lengths. This effect is commonly modeled as a second-order receiver transfer function. Jitter will also accumulate when the link is transferring random data; jitter due to random data is of two types, systematic and random. The classic model for systematic jitter accumulation in cascaded repeaters was published by Byrne.²⁶ The Byrne model assumes cascaded identical timing recovery circuits, and then the systematic and random jitter can be combined as rms quantities so that total jitter due to random jitter may be obtained. This model has been generalized to networks consisting of different components,²⁷ and to nonidentical repeaters.²⁸ Despite these considerations, for well designed practical networks the basic results of the Byrne model remain valid for N nominally identical repeaters transmitting random data; systematic jitter accumulates in proportion to $N^{1/2}$ and random jitter accumulates in proportion to $N^{1/4}$. For most applications the maximum timing jitter should be kept below about 30 percent of the maximum receiver eye opening.

Modal Noise

An additional effect of lossy connectors and splices is modal noise. Because high capacity optical links tend to use highly coherent laser transmitters, random coupling between fiber modes causes fluctuations in the optical power coupled through splices and connectors; this phenomena is known as *modal noise*.²⁹ As one might expect, modal noise is worst when using laser sources in conjunction with multimode fiber; recent industry standards have allowed the use of short-wave lasers (750 to 850 nm) on 50 μm fiber which may experience this problem. Modal noise is usually considered to be nonexistent in single-mode systems. However, modal noise in single-mode fibers can arise when higher-order modes are generated at imperfect connections or splices. If the lossy mode is not completely attenuated before it reaches the next connection, interference with the dominant mode may occur. The effects of modal noise have been modeled previously,²⁹ assuming that the only significant interaction occurs between the LP_{01} and LP_{11} modes for a sufficiently coherent laser. For N sections of fiber, each of length L in a single-mode link, the worst case sigma for modal noise can be given by

$$\sigma_m = \sqrt{2N\eta}(1 - \eta)e^{-aL} \quad (30)$$

where a is the attenuation coefficient of the LP_{11} mode, and η is the splice transmission efficiency, given by

$$\eta = 10^{-(\eta_o/10)} \quad (31)$$

where η_o is the mean splice loss (typically, splice transmission efficiency will exceed 90%). The corresponding optical power penalty due to modal noise is given by

$$P = -5 \log(1 - Q^2 \sigma_m^2) \quad (32)$$

where Q corresponds to the desired BER. This power penalty should be kept to less than 0.5 dB.

Radiation Induced Loss

Another important environmental factor as mentioned earlier is exposure of the fiber to ionizing radiation damage. There is a large body of literature concerning the effects of ionizing radiation on fiber links.^{30,31} There are many factors which can affect the radiation susceptibility of optical fiber, including the type of fiber, type of radiation (gamma radiation is usually assumed to be representative), total dose, dose rate (important only for higher exposure levels), prior irradiation history of the fiber, temperature, wavelength, and data rate. Optical fiber with a pure silica core is least susceptible to radiation damage; however, almost all commercial fiber is intentionally doped to control the refractive index of the core and cladding, as well as dispersion properties. Trace impurities are also introduced which become important only under irradiation; among the most important are Ge dopants in the core of graded index (GRIN) fibers, in addition to F, Cl, P, B, OH content, and the alkali metals. In general, radiation sensitivity is worst at lower temperatures, and is also made worse by hydrogen diffusion from materials in the fiber cladding. Because of the many factors involved, there does not exist a comprehensive theory to model radiation damage in optical fibers. The basic physics of the interaction has been described;^{30,31} there are two dominant mechanisms, radiation induced darkening and scintillation. First, high energy radiation can interact with dopants, impurities, or defects in the glass structure to produce color centers which absorb strongly at the operating wavelength. Carriers can also be freed by radiolytic or photochemical processes; some of these become trapped at defect sites, which modifies the band structure of the fiber and causes strong absorption at infrared wavelengths. This radiation-induced darkening increases the fiber attenuation; in some cases, it is partially reversible when the radiation is removed, although high-levels or prolonged exposure will permanently damage the fiber. A second effect is caused if the radiation interacts with impurities to produce stray light, or scintillation. This light is generally broadband, but will tend to degrade the BER at the receiver; scintillation is a weaker effect than radiation-induced darkening. These effects will degrade the BER of a link; they can be prevented by shielding the fiber, or partially overcome by a third mechanism, photobleaching. The presence of intense light at the proper wavelength can partially reverse the effects of darkening in a fiber. It is also possible to treat silica core fibers by briefly exposing them to controlled levels of radiation at controlled temperatures; this increases the fiber loss, but makes the fiber less susceptible to future irradiation. These so-called radiation hardened fibers are often used in environments where radiation is anticipated to play an important role. Recently, several models have been advanced³¹ for the performance of fiber under moderate radiation levels; the effect on BER is a power law model of the form

$$BER = BER_0 + A(\text{dose})^b \quad (33)$$

where BER_0 is the link BER prior to irradiation, the dose is given in rads, and the constants A and b are empirically fitted. The loss due to normal background radiation exposure over a typical link lifetime can be held approximately below 0.5 dB.

15.4 REFERENCES

1. S. E. Miller and A. G. Chynoweth, editors, *Optical Fiber Telecommunications*, Academic Press, Inc., New York, N.Y. (1979).
2. J. Gowar, *Optical Communication Systems*, Prentice Hall, Englewood Cliffs, N.J. (1984).
3. C. DeCusatis, editor, *Handbook of Fiber Optic Data Communication*, Elsevier/Academic Press, New York, N.Y. (first edition 1998, second edition 2002); see also *Optical Engineering* special issue on optical data communication (December 1998).
4. R. Lasky, U. Osterberg, and D. Stigliani, editors, *Optoelectronics for Data Communication*, Academic Press, New York, N.Y. (1995).
5. Digital Video Broadcasting (DVB) Measurement Guidelines for DVB Systems, "European Telecommunications Standards Institute ETSI Technical Report ETR 290, May 1997;" Digital Multi-Programme Systems for Television Sound and Data Services for Cable Distribution, "International Telecommunications Union ITU-T Recommendation J.83, 1995;" Digital Broadcasting System for Television, Sound and Data Services; Framing Structure, Channel Coding and Modulation for Cable Systems, "European Telecommunications Standards Institute ETSI 300 429," 1994.
6. W. E. Stephens and T. R. Hoseph, "System Characteristics of Direct Modulated and Externally Modulated RF Fiber-Optic Links," *IEEE J. Lightwave Technol.* **LT-5**(3):380–387 (1987).
7. C. H. Cox, III and, E. I. Ackerman, "Some Limits on the Performance of an Analog Optical Link," *Proc. SPIE—Int. Soc. Opt. Eng.* **3463**:2–7 (1999).
8. United States laser safety standards are regulated by the Dept. of Health and Human Services (DHHS), Occupational Safety and Health Administration (OSHA), Food and Drug Administration (FDA), Code of Radiological Health (CDRH), 21 Code of Federal Regulations (CFR) subchapter J; the relevant standards are ANSI Z136.1, "Standard for the Safe Use of Lasers" (1993 revision) and ANSI Z136.2, "Standard for the Safe Use of Optical Fiber Communication Systems Utilizing Laser Diodes and LED Sources" (1996–97 revision); elsewhere in the world, the relevant standard is International Electrotechnical Commission (IEC/CEI) 825 (1993 revision).
9. S. S. Walker, "Rapid Modeling and Estimation of Total Spectral Loss in Optical Fibers," *IEEE J. Lightwave Technol.* **4**:1125–1132 (1996).
10. Electronics Industry Association/Telecommunications Industry Association (EIA/TIA) Commercial Building Telecommunications Cabling Standard (EIA/TIA-568-A), Electronics Industry Association/Telecommunications Industry Association (EIA/TIA) Detail Specification for 62.5 micron Core Diameter/125 micron Cladding Diameter Class 1a Multimode Graded Index Optical Waveguide Fibers (EIA/TIA-492AAAA), Electronics Industry Association/Telecommunications Industry Association (EIA/TIA) Detail Specification for Class IV-a Dispersion Unshifted Single-Mode Optical Waveguide Fibers Used in Communications Systems (EIA/TIA-492BAAA), Electronics Industry Association, New York, N.Y.
11. D. Gloge, "Propagation Effects in Optical Fibers," *IEEE Trans. Microwave Theory Technol.* **MTT-23**:106–120 (1975).
12. P. M. Shanker, "Effect of Modal Noise on Single-Mode Fiber Optic Network," *Opt. Comm.* **64**:347–350 (1988).
13. J. J. Refi, "LED Bandwidth of Multimode Fiber as a Function of Source Bandwidth and LED Spectral Characteristics," *IEEE J. Lightwave Technol.* **LT-14**:265–272 (1986).
14. G. P. Agrawal et al., "Dispersion Penalty for 1.3 Micron Lightwave Systems with Multimode Semiconductor Lasers," *IEEE J. Lightwave Technol.* **6**:620–625 (1988).
15. K. Ogawa, "Analysis of Mode Partition Noise in Laser Transmission Systems," *IEEE J. Quantum Elec.* **QE-18**:849–9855 (1982).
16. K. Ogawa, *Semiconductor Laser Noise; Mode Partition Noise*, in *Semiconductors and Semimetals* (R. K. Willardson and A. C. Beer, editors), vol. 22C, Academic Press, New York, N. Y. (1985).
17. J. C. Campbell, "Calculation of the Dispersion Penalty of the Route Design of Single-Mode Systems," *IEEE J. Lightwave Technol.* **6**:564–573 (1988).
18. M. Ohtsu et al., "Mode Stability Analysis of Nearly Single-Mode Semiconductor Laser," *IEEE J. Quantum Elec.* **24**:716–723 (1988).
19. M. Ohtsu and Y. Teramachi, "Analysis of Mode Partition and Mode Hopping in Semiconductor Lasers," *IEEE Quantum Elec.* **25**:31–38 (1989).

20. D. Duff et al., "Measurements and Simulations of Multipath Interference for 1.7 Gbit/s Lightwave Systems Utilizing Single and Multifrequency Lasers," *Proc. OFC* p. 128 (1989).
21. J. Radcliffe, "Fiber Optic Link Performance in the Presence of Internal Noise Sources," *IBM Technical Report*, Glendale Labs, Endicott, New York, N.Y. (1989).
22. L. L. Xiao, C. B. Su, and R. B. Lauer, "Increase in Laser RIN due to Asymmetric Nonlinear Gain, Fiber Dispersion, and Modulation," *IEEE Photon. Tech. Lett.* **4**:774–777 (1992).
23. P. Trischitta and P. Sannuti, "The Accumulation of Pattern Dependent Jitter for a Chain of Fiber Optic Regenerators," *IEEE Trans. Comm.* **36**:761–765 (1988).
24. *CCITT Recommendations G.824, G.823, O.171, and G.703* on timing jitter in digital systems (1984).
25. R. J. S. Bates, "A Model for Jitter Accumulation in Digital Networks," *IEEE Globecom Proc.* pp. 145–149 (1983).
26. C. J. Byrne, B. J. Karafin, and D. B. Robinson, Jr., "Systematic Jitter in a Chain of Digital Regenerators," *Bell Sys. Tech. J.* **43**:2679–2714 (1963).
27. R. J. S. Bates and L. A. Sauer, "Jitter Accumulation in Token Passing Ring LANs," *IBM J. Res. Dev.* **29**:580–587 (1985).
28. C. Chamzas, "Accumulation of Jitter: A Stochastic Model," *AT&T Tech. J.* p. 64 (1985).
29. D. Marcuse and H. M. Presby, "Mode Coupling in an Optical Fiber with Core Distortion," *Bell Sys. Tech. J.* **1**:3 (1975).
30. E. J. Frieble et al., "Effect of Low Dose Rate Irradiation on Doped Silica Core Optical Fibers," *App. Opt.* **23**:4202–4208 (1984).
31. J. B. Haber et al., "Assessment of Radiation Induced Loss for AT&T Fiber Optic Transmission Systems in the Terrestrial Environment," *IEEE J. Lightwave Technol.* **6**:150–154 (1988).

This page intentionally left blank

FIBER-BASED COUPLERS

Daniel Nolan

Corning Inc.

Corning, New York

16.1 INTRODUCTION

Fiber-optic couplers, including splitters and wavelength division multiplexing components, have been used extensively over the last two decades. This use continues to grow both in quantity and in the ways in which the devices are used. The uses today include, among other applications, simple splitting for signal distribution and the wavelength multiplexing and demultiplexing multiple wavelength signals.

Fiber-based splitters and wavelength division multiplexing (WDM) components are among the simplest devices. Other technologies that can be used to fabricate components that exhibit similar functions include the planar waveguide and micro-optic technologies. Planar waveguides are most suitable for highly integrated functions. Micro-optic devices are often used when complex multiple wavelength functionality is required. In this chapter, we will show the large number of optical functions that can be achieved with simple tapered fiber components. We will also describe the physics of the propagation of light through tapers in order to better understand the breadth of components that can be fabricated with this technology. The phenomenon of coupling includes an exchange of power that can depend both on wavelength and on polarization. Beyond the simple 1×2 power splitter, other devices that can be fabricated from tapered fibers include $1 \times N$ devices, wavelength multiplexing, polarization multiplexing, switches, attenuators, and filters.

Fiber-optic couplers have been fabricated since the early 1970s. The fabrication technologies have included fusion tapering,¹⁻³ etching,⁴ and polishing.⁵⁻⁷ The tapered single-mode fiber-optic power splitter is perhaps the most universal of the single-mode tapered devices.⁸ It has been shown that the power transferred during the tapering process involves an initial adiabatic transfer of the power in the input core to the cladding air interface.⁹ The light is then transferred to the adjacent core-cladding mode. During the uptapering process, the input light will transfer back onto the fiber cores. In this case, it is referred to as a cladding mode coupling device. Light that is transferred to a higher-order mode of the core-cladding structure leads to an excess loss. This is because these higher-order modes are not bounded by the core and are readily stripped by the higher index of the fiber coating.

In the tapered fiber coupler process, two fibers are brought into close proximity after the protective plastic jacket is removed. Then, in the presence of a torch, the fibers are fused and stretched (Fig. 1.) The propagation of light through this tapered region is described using Maxwell's vector

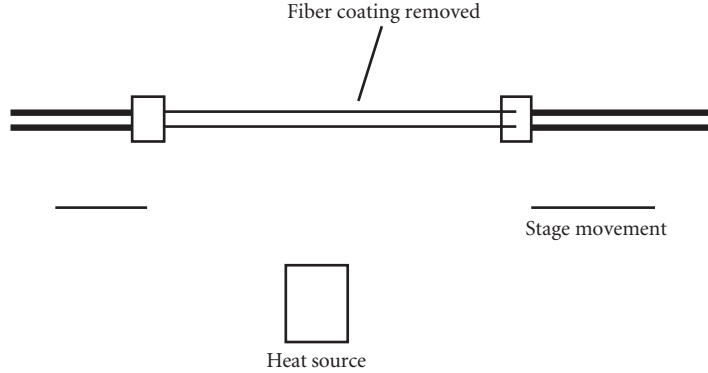


FIGURE 1 Fusing and tapering process.

equations, but to a good approximation, the scalar wave approximation is valid. The scalar wave equation written in cylindrical coordinates is expressed as

$$\left[\frac{1}{r} \frac{\partial}{\partial r} r \frac{\partial}{\partial r} - \nu / r^2 + k^2 n_1^2 - (V/a)^2 f(r/a) \right] \psi = \epsilon \mu \frac{\partial^2 \psi}{\partial t^2} \quad (1)$$

In Eq. (1), n_1 is the index value at $r = 0$, β is the propagation constant, which is to be determined, a is the core radius, $f(r/a)$ is a function describing the index distribution with radius, and V is the modal volume

$$V = \frac{2\pi n_1}{\lambda \sqrt{2\delta}} \quad (2)$$

with

$$\delta = \frac{[n_1^2 - n_2^2]}{2n_1^2} \quad (3)$$

As light propagates in the single-mode fiber, it is not confined to the core region, but extends out into the surrounding region. As the light propagates through the tapered region, it is bounded by the shrinking, air-cladding boundary.

In the simplest case, the coupling from one cladding to the adjacent one can be described by perturbation theory.⁹ In this case, the cladding air boundary is considered as the waveguide outer boundary, and the exchange of power along z is described as

$$P = \sin^2[Cz] \quad (4)$$

where¹⁰

$$C = \frac{[\pi \delta / W d \rho]^2 U^2 \exp(-W d / \rho)}{[V^3 K_1^2(W)]} \quad (5)$$

with

$$\alpha = 2\pi n_1 / \lambda \quad U = \rho(k^2 n_1^2 - \beta^2)^2 \quad W = \rho(\beta^2 - k^2 n_2^2)^2 \quad (6)$$

In Eq. (6), the waveguide parameters are defined in the tapered region. Here the core of each fiber is small and the cladding becomes the effective core, while air becomes the cladding. Also, it

is important to point out that Eqs. (4) and (5) are only a first approximation. These equations are derived using first-order perturbation theory. Also, the scalar wave equation is not strictly valid under the presence of large index differences, such as at a glass-air boundary. However, these equations describe a number of important effects. The sinusoidal dependence of the power coupled with wavelength, as well as the dependence of power transfer with cladding diameter and other dependencies, is well described with the model.

Equation (4) can be described by considering the light input to one core as a superposition of symmetric and antisymmetric.⁹ These modes are eigen solutions to the composite two-core structure. The proper superposition of these two modes enables one to impose input boundary conditions for the case of a two-core structure. The symmetric and antisymmetric modes are written as

$$\psi_s = \frac{\psi_1 + \psi_2}{\sqrt{2}} \quad (7)$$

$$\psi_a = \frac{\psi_1 - \psi_2}{\sqrt{2}} \quad (8)$$

Light input onto one core is described with ψ_1 at $z = 0$,

$$\psi_1 = \frac{\psi_s + \psi_a}{\sqrt{2}} \quad (9)$$

Propagation through the coupler is characterized with the superposition of ψ_s and ψ_a . This superposition describes the power transfer between the two guides along the direction of propagation.¹⁰ The propagation constants of ψ_s and ψ_a are slightly different, and this value can be used to estimate excess loss under certain perturbations.

16.2 ACHROMATICITY

The simple sinusoidal dependence of the coupling with wavelength as described above is not always desired, and often a more achromatic dependence of the coupling is required. This can be achieved when dissimilar fibers¹⁰ are used to fabricate the coupler. Fibers are characterized as dissimilar when the propagation constants of the guides are of different values. When dissimilar fibers are used, Eqs. (4) and (5) can be replaced with

$$P_1(x) = P_1(0) + F^2(P_2(0) - P_1(0) + [\delta\beta/C][P_1(0)P_2(0)]^2)\sin^2(Cz/F) \quad (10)$$

where

$$F = 1/[1 + \delta\beta/(4C^2)] \quad (11)$$

In most cases, the fibers are made dissimilar by changing the cladding diameter of one of the fibers. Etching or pretapering one of the fibers can do this. Another approach is to slightly change the cladding index of one of the fibers.¹¹ When dissimilar fibers are used, the total amount of power coupled is limited. As an example, an achromatic 3-dB coupler is made achromatic by operating at the sinusoidal maximum with wavelength rather than at the power of maximum power change with wavelength. Another approach to achieve achromaticity is to taper the device such that the modes expand well beyond the cladding boundaries.¹² This condition greatly weakens the wavelength dependence of the coupling. This has been achieved by encapsulating the fibers in a third-matrix glass with an index very close to that of the fiber's cladding index. The difference in index between the cladding and the matrix glass is in the order of 0.001. The approach of encapsulating the fibers in a third-index material^{13,14} is also useful for reasons other than achromaticity. One reason is that

the packaging process is simplified. Also, a majority of couplers made for undersea applications use this method because it is a proven approach to ultrahigh reliability.

The wavelength dependence of the couplers described above is most often explained using mode coupling and perturbation theory. Often, numerical analysis is required to explain the effects that the varying taper angles have on the overall coupling. An important numerical approach is the beam propagation method.¹⁵ In this approach, the propagation of light through a device is solved by an expansion of the evolution operator using a Taylor series and with the use of fast Fourier transforms to evaluate the appropriate derivatives. In this way, the propagation of the light can be studied as it couples to the adjacent guides or to higher-order modes.

16.3 WAVELENGTH DIVISION MULTIPLEXING

Besides power splitting, tapered couplers can be used to separate wavelengths. To accomplish this separation, we utilize the wavelength dependence of Eqs. (4) and (5). By proper choice of the device length and taper ratio, two predetermined wavelengths can be put out onto two different ports. Wavelengths from 60 to 600 nanometers can be split using this approach. Applications include the splitting and/or combining of 1480 nm and 1550 nm light, as well as multiplexing 980 and 1550 nm onto an erbium fiber for signal amplification. Also important is the splitting of the 1310- to 1550-nm wavelength bands, which can be achieved using this approach.

16.4 $1 \times N$ POWER SPLITTERS

Often it is desirable to split a signal onto a number of output ports. This can be achieved by concatenating 1×2 power splitters. Alternatively, one can split the input simultaneously onto multiple output ports.^{16,17} Typically, the output ports are of the form 2^N (i.e., 2, 4, 8, 16, . . .). The configuration of the fibers in the tapered region affects the distribution of the output power per port. A good approach to achieve uniform 1×8 splitting is described in Ref. 18.

16.5 SWITCHES AND ATTENUATORS

In a tapered device, the power coupled over to the adjacent core can be significantly affected by bending the device at the midpoint. By encapsulating two fibers before tapering in a third index medium, the device is rigid and can be reliably bent in order to frustrate the coupling.¹⁹ The bending establishes a difference in the propagation constants of the two guiding media, preventing coupling or power transfer.

This approach can be used to fabricate both switches and attenuators. Switches with up to 30 dB crosstalk and attenuators with variable crosstalk up to 30 dB as well over the erbium wavelength band have been fabricated. Displacing one end of a 1 cm taper by 1 mm is enough to alter the crosstalk by the 30-dB value. Applications for attenuators have been increasing significantly over the last few years. An important reason is to maintain the gain in erbium-doped fiber amplifiers. This is achieved by limiting the amount of pump power into the erbium fiber. Over time, as the pump degrades, the power output of the attenuator is increased to compensate for the pump degradation.

16.6 MACH-ZEHNDER DEVICES

Devices to split narrowly spaced wavelengths are very important. As mentioned above, tapers can be designed such that wavelengths from 60 to 600 nm can be split in a tapered device. Dense WDM networks require splitting of wavelengths with separations on the order of nanometers. Fiber-based

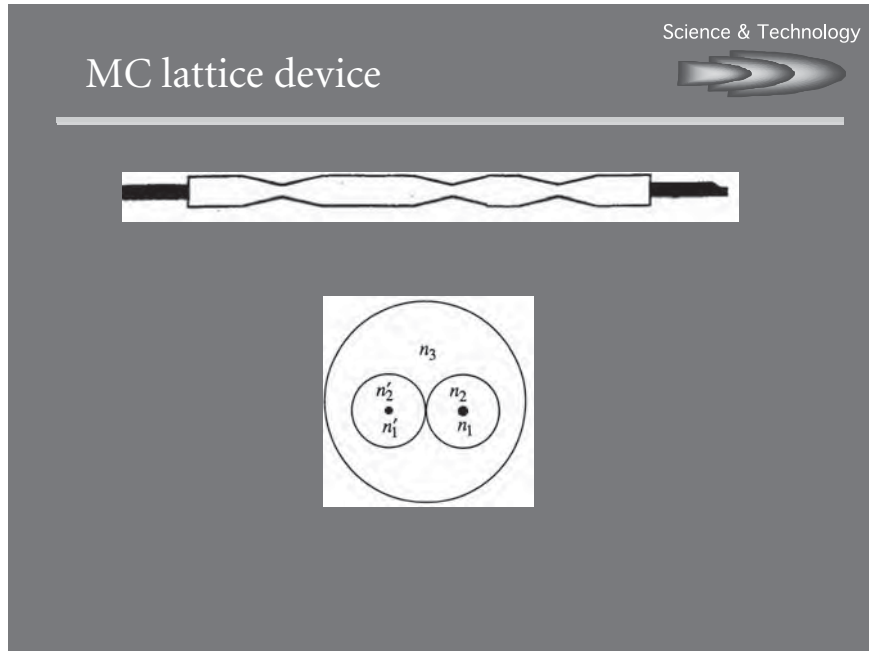


FIGURE 2 Fiber-based Mach-Zehnder devices.²⁴

Mach-Zehnder devices enable such splitting. Monolithic fiber-based Mach-Zehnders can be fabricated using fibers with different cores,^{20,21} i.e., different propagation constants. Two or more tapers can be used to cause light from two different optical paths to interfere (Fig. 2). The dissimilar cores enable light to propagate at different speeds between the tapers, causing the required constructive and destructive interference. These devices are environmentally stable due to the monolithic structure. Mach-Zehnders can also be fabricated using fibers with different lengths between the tapers.²² In this approach, it is the packaging that enables an environmentally stable device.

Mach-Zehnders and lattice filters can also be fabricated by tapering single-fiber devices.^{23,24} In the tapered regions, the light couples to a cladding mode. The cladding mode propagates between tapers since a lower-index overcladding replaces the higher-index coating material. An interesting application for these devices is gain-flattening filters for amplifiers.

16.7 POLARIZATION DEVICES

It is well known that two polarization modes propagate in single-mode fiber. Most optical fiber modules allow both polarizations to propagate, but specify that the performance of the components be insensitive to the polarization states of the propagating light. However, this is often not the situation for fiber-optic sensor applications. Often, the state of polarization is important to the operation of the sensor itself. In these situations, polarization-maintaining fiber is used. Polarization components such as polarization-maintaining couplers and also single polarization devices are used. In polarization-maintaining fiber, a difference in propagation constants of the polarization modes prevents mode coupling or exchange of energy. This is achieved by introducing stress or shape birefringence within the fiber core. A significant difference between the two polarization modes is maintained as the fiber twists in a cable or package.

In many fiber sensor systems, tapered fiber couplers are used to couple light from one core to another. Often the couplers are composed of birefringent fibers.^{24,25} This is done to maintain the

alignment of the polarizations to the incoming and outgoing fibers and also to maintain the polarization states within the device. The axes of the birefringent fibers are aligned before tapering, and care is taken not to excessively twist the fibers during the tapering process.

The birefringent fibers contain stress rods, elliptical core fibers, or inner claddings to maintain the birefringence. The stress rods in some birefringent fibers have an index higher than the silica cladding. In the tapering process, this can cause light to be trapped in these rods, resulting in an excess loss in the device. Stress rods with an index lower than that of silica can be used in these fibers, resulting in very low loss devices.

16.8 SUMMARY

Tapered fiber couplers are extremely useful devices. Such devices include 1×2 and $1 \times N$ power splitters, wavelength division multiplexers and filters, and polarization-maintaining and -splitting components. Removing the fiber's plastic coating and then fusing and tapering two or more fibers in the presence of heat forms these devices. The simplicity and flexibility of this fabrication process is in part responsible for the widespread use of these components. The mechanism involved in the fabrication process is reasonably understood and simple, which is in part responsible for the widespread deployment of these devices. These couplers are found in optical modules for the telecommunication industry and in assemblies for the sensing industry. They are also being deployed as standalone components for fiber-to-home applications.

16.9 REFERENCES

1. T. Ozeki and B. S. Kawaski, "New Star Coupler Compatible with Single Multimode Fiber Links," *Electron. Lett.* **12**:151–152, 1976.
2. B. S. Kawaski and K. O. Hill, "Low Loss Access Coupler for Multimode Optical Fiber Distribution Networks," *Appl. Opt.* **16**:1794–1795, 1977.
3. G. E. Rawson and M. D. Bailey, "Bitaper Star Couplers with up to 100 Fiber Channels," *Electron. Lett.* **15**:432–433, 1975.
4. S. K. Sheem and T. G. Giallorenzi, "Single-Mode Fiber Optical Power Divided; Encapsulated Etching Technique," *Opt. Lett.* **4**:31, 1979.
5. Y. Tsujimoto, H. Serizawa, K. Hatori, and M. Fukai, "Fabrication of Low Loss 3 dB Couplers with Multimode Optical Fibers," *Electron. Lett.* **14**:157–158, 1978.
6. R. A. Bergh, G. Kotler, and H. J. Shaw, "Single-Mode Fiber Optic Directional Coupler," *Electron. Lett.* **16**:260–261, 1980.
7. O. Parriaux, S. Gidon, and A. Kuznetsov, "Distributed Coupler on Polished Single-Mode Fiber," *Appl. Opt.* **20**:2420–2423, 1981.
8. B. S. Kawaski, K. O. Hill, and R. G. Lamont, "Biconical—Taper Single-Mode Fiber Coupler," *Opt. Lett.* **6**:327, 1981.
9. R. G. Lamont, D. C. Johnson, and K. O. Hill, "Power Transfer in Fused Biconical Single Mode Fiber Couplers: Dependence on External Refractive Index," *Appl. Opt.* **24**:327–332, 1984.
10. A. Snyder and J. D. Love, *Optical Waveguide Theory*, London: Chapman and Hall, 1983.
11. W. J. Miller, C. M. Truesdale, D. L. Weidman, and D. R. Young, "Achromatic Fiber Optic Coupler," U.S. Patent 5,011,251, Apr. 1991.
12. D. L. Weidman, "Achromat Overclad Coupler," U.S. Patent, 5,268,979, Dec. 1993.
13. C. M. Truesdale and D. A. Nolan, "Core-Clad Mode Coupling in a New Three-Index Structure," *European Conference on Optical Communications*, Barcelona Spain, 1986.
14. D. B. Keck, A. J. Morrow, D. A. Nolan, and D. A. Thompson, "Passive Optical Components in the Subscriber Loop," *J. Lightwave Technol.* **7**:1623–1633, 1989.

15. M. D. Feit and J. A. Fleck, "Simple Spectral Method for Solving Propagation Problems in Cylindrical Geometry with Fast Fourier Transforms," *Opt. Lett.* **14**:662–664, 1989.
16. D. B. Mortimore and J. W. Arkwright, "Performance of Wavelength-Flattened 1×7 Fused Couplers," Optical Fiber Conference, TUG6, 1990.
17. D. L. Weidman, "A New Approach to Achromaticity in Fused $1 \times N$ Couplers," Optical Fiber Conference, Post Deadline papers, 1994.
18. W. J. Miller, D. A. Nolan, and G. E. Williams, "Method of Making a $1 \times N$ Coupler," US Patent, 5,017,206, 1991.
19. M. A. Newhouse and F. A. Annunziata, "Single-Mode Optical Switch," Technical Digest of the National Fiber Optic Conference, 1990.
20. D. A. Nolan and W. J. Miller, "Wavelength Tunable Mach-Zehnder Device," Optical Conference, 1994.
21. B. Malo, F. Bilodeau, K. O. Hill, and J. Albert, "Unbalanced Dissimilar—Fiber Mach—Zehnder Interferometer: Application as Filter," *Electron. Lett.* **25**:1416, 1989.
22. C. Huang, H. Luo, S. Xu, and P. Chen, "Ultra Low Loss, Temperature Insensitive 16 Channel 100 GHz Dense WDMs Based on Cascaded All Fiber Unbalanced Mach-Zehnder Structure," Optical Fiber Conference, TUH2, 1999.
23. D. A. Nolan, W. J. Miller, and R. Irion, "Fiber Based Band Splitter," Optical Fiber Conference, 1998.
24. D. A. Nolan, W. J. Miller, G. Berkey, and L. Bhagavatula, "Tapered Lattice Filters," Optical Fiber Conference, TUH4, 1999.
25. I. Yokohama, M. Kawachi, K. Okamoto, and J. Noda, *Electron. Lett.* **22**:929, 1986.

This page intentionally left blank

FIBER BRAGG GRATINGS

Kenneth O. Hill

*Communications Research Centre
Ottawa, Ontario, Canada, and
Nu-Wave Photonics
Ottawa, Ontario, Canada*

17.1 GLOSSARY

FBG	fiber Bragg grating
FWHM	full width measured at half-maximum intensity
N_{eff}	effective refractive index for light propagating in a single mode
pps	pulses per second
β	propagation constant of optical fiber mode
Δn	magnitude of photoinduced refractive index change
κ	grating coupling coefficient
Λ	spatial period (or pitch) of spatial feature measured along optical fiber
λ	vacuum wavelength of propagating light
λ_B	Bragg wavelength
L	length of grating

17.2 INTRODUCTION

A fiber Bragg grating (FBG) is a periodic variation of the refractive index of the fiber core along the length of the fiber. The principal property of FBGs is that they reflect light in a narrow bandwidth that is centered about the Bragg wavelength λ_B which is given by $\lambda_B = 2N_{\text{eff}}\Lambda$, where Λ is the spatial period (or pitch) of the periodic variation and N_{eff} is the effective refractive index for light propagating in a single mode, usually the fundamental mode of a monomode optical fiber. The refractive index variations are formed by exposure of the fiber core to an intense optical interference pattern of ultraviolet light. The capability of light to induce permanent refractive index changes in the core of an optical fiber has been named photosensitivity. Photosensitivity was discovered by Hill et al. in 1978 at the Communications Research Centre in Canada (CRC).^{1,2} The discovery has led to techniques for fabricating Bragg gratings in the core of an optical fiber and a means for manufacturing a

wide range of FBG-based devices that have applications in optical fiber communications and optical sensor systems.

This chapter reviews the characteristics of photosensitivity, the properties of Bragg gratings, the techniques for fabricating Bragg gratings in optical fibers, and some FBG devices. More information on FBGs can be found in the following references, which are reviews on Bragg grating technology,^{3,4} the physical mechanisms underlying photosensitivity,⁵ applications for fiber gratings,⁶ and the use of FBGs as sensors.⁷

17.3 PHOTSENSITIVITY

When ultraviolet light radiates an optical fiber, the refractive index of the fiber is changed permanently; the effect is termed *photosensitivity*. The change in refractive index is permanent in the sense that it will last for several years (lifetimes of 25 years are predicted) if the optical waveguide after exposure is annealed appropriately; that is, by heating for a few hours at a temperature of 50°C above its maximum anticipated operating temperature.⁸ Initially, photosensitivity was thought to be a phenomenon that was associated only with germanium-doped-core optical fibers. Subsequently, photosensitivity has been observed in a wide variety of different fibers, many of which do not contain germanium as dopant. Nevertheless, optical fiber with a germanium-doped core remains the most important material for the fabrication of Bragg grating-based devices.

The magnitude of the photoinduced refractive index change (Δn) obtained depends on several different factors: the irradiation conditions (wavelength, intensity, and total dosage of irradiating light), the composition of glassy material forming the fiber core, and any processing of the fiber prior and subsequent to irradiation. A wide variety of different continuous-wave and pulsed-laser light sources, with wavelengths ranging from the visible to the vacuum ultraviolet, have been used to photoinduce refractive index changes in optical fibers. In practice, the most commonly used light sources are KrF and ArF excimer lasers that generate, respectively, 248- and 193-nm light pulses (pulse width ~10 ns) at pulse repetition rates of 50 to 100 pps. Typically, the fiber core is exposed to laser light for a few minutes at pulse levels ranging from 100 to 1000 mJ cm⁻² pulse⁻¹. Under these conditions, Δn is positive in germanium-doped monomode fiber with a magnitude ranging between 10⁻⁵ and 10⁻³.

The refractive index change can be enhanced (photosensitization) by processing the fiber prior to irradiation using such techniques as *hydrogen loading*⁹ or *flame brushing*.¹⁰ In the case of hydrogen loading, a piece of fiber is put in a high-pressure vessel containing hydrogen gas at room temperature; pressures of 100 to 1000 atmospheres (atm; 101 kPa/atm) are applied. After a few days, hydrogen in molecular form has diffused into the silica fiber; at equilibrium the fiber becomes saturated (i.e., loaded) with hydrogen gas. The fiber is then taken out of the high-pressure vessel and irradiated before the hydrogen has had sufficient time to diffuse out. Photoinduced refractive index changes up to 100 times greater are obtained by hydrogen loading a Ge-doped-core optical fiber. In flame brushing, the section of fiber that is to be irradiated is mounted on a jig and a hydrogen-fueled flame is passed back and forth (i.e., brushed) along the length of the fiber. The brushing takes about 10 minutes, and upon irradiation, an increase in the photoinduced refractive index change by about a factor of 10 can be obtained.

Irradiation at intensity levels higher than 1000 mJ/cm² marks the onset of a different non-linear photosensitive process that enables a single irradiating excimer light pulse to photo-induce a large index change in a small localized region near the core/cladding boundary of the fiber. In this case, the refractive index changes are sufficiently large to be observable with a phase contrast microscope and have the appearance of physically damaging the fiber. This phenomenon has been used for the writing of gratings using a single-excimer light pulse.

Another property of the photoinduced refractive index change is *anisotropy*. This characteristic is most easily observed by irradiating the fiber from the side with ultraviolet light that is polarized perpendicular to the fiber axis. The anisotropy in the photoinduced refractive index change results in the fiber becoming birefringent for light propagating through the fiber. The effect is useful for fabricating polarization mode-converting devices or rocking filters.¹¹

The physical processes underlying photosensitivity have not been fully resolved. In the case of germanium-doped glasses, photosensitivity is associated with GeO color center defects that have strong absorption in the ultraviolet (~ 242 nm) wavelength region. Irradiation with ultraviolet light bleaches the color center absorption band and increases absorption at shorter wavelengths, thereby changing the ultraviolet absorption spectrum of the glass. Consequently, as a result of the Kramers-Kronig causality relationship,¹² the refractive index of the glass also changes; the resultant refractive index change can be sensed at wavelengths that are far removed from the ultraviolet region extending to wavelengths in the visible and infrared. The physical processes underlying photosensitivity are, however, probably much more complex than this simple model. There is evidence that ultraviolet light irradiation of Ge-doped optical fiber results in structural rearrangement of the glass matrix leading to densification, thereby providing another mechanism for contributing to the increase in the fiber core refractive index. Furthermore, a physical model for photosensitivity must also account for the small anisotropy in the photoinduced refractive index change and the role that hydrogen loading plays in enhancing the magnitude of the photoinduced refractive change. Although the physical processes underlying photosensitivity are not completely known, the phenomenon of glass-fiber photosensitivity has the practical result of providing a means, using ultraviolet light, for photoinducing permanent changes in the refractive index at wavelengths that are far removed from the wavelength of the irradiating ultraviolet light.

17.4 PROPERTIES OF BRAGG GRATINGS

Bragg gratings have a periodic index structure in the core of the optical fiber. Light propagating in the Bragg grating is backscattered slightly by Fresnel reflection from each successive index perturbation. Normally, the amount of backscattered light is very small except when the light has a wavelength in the region of the Bragg wavelength λ_B , given by

$$\lambda_B = 2N_{\text{eff}}\Lambda$$

where N_{eff} is the modal index and Λ is the grating period. At the Bragg wavelength, each back reflection from successive index perturbations is in phase with the next one. The back reflections add up coherently and a large reflected light signal is obtained. The reflectivity of a strong grating can approach 100 percent at the Bragg wavelength, whereas light at wavelengths longer or shorter than the Bragg wavelength pass through the Bragg grating with negligible loss. It is this wavelength-dependent behavior of Bragg gratings that makes them so useful in optical communications applications. Furthermore, the optical pitch ($N_{\text{eff}}\Lambda$) of a Bragg grating contained in a strand of fiber is changed by applying longitudinal stress to the fiber strand. This effect provides a simple means for sensing strain optically by monitoring the concomitant change in the Bragg resonant wavelength.

Bragg gratings can be described theoretically by using coupled-mode equations.^{4,6,13} Here, we summarize the relevant formulas for tightly bound monomode light propagating through a uniform grating. The grating is assumed to have a sinusoidal perturbation of constant amplitude Δn . The reflectivity of the grating is determined by three parameters: (1) the coupling coefficient κ (2) the mode propagation constant $\beta = 2\pi N_{\text{eff}}/\lambda$, and (3) the grating length L . The coupling coefficient κ which depends only on the operating wavelength of the light and the amplitude of the index perturbation Δn is given by $\kappa = (\pi/\lambda)\Delta n$. The most interesting case is when the wavelength of the light corresponds to the Bragg wavelength. The grating reflectivity R of the grating is then given by the simple expression, $R = \tan^2(h^2(\kappa L))$, where κ is the coupling coefficient at the Bragg wavelength and L is the length of the grating. Thus, the product κL can be used as a measure of grating strength. For $\kappa L = 1, 2, 3$, the grating reflectivity is, respectively, 58, 93, and 99 percent. A grating with a κL greater than one is termed a strong grating, whereas a weak grating has κL less than one. Figure 1 shows the typical reflection spectra for weak and strong gratings.

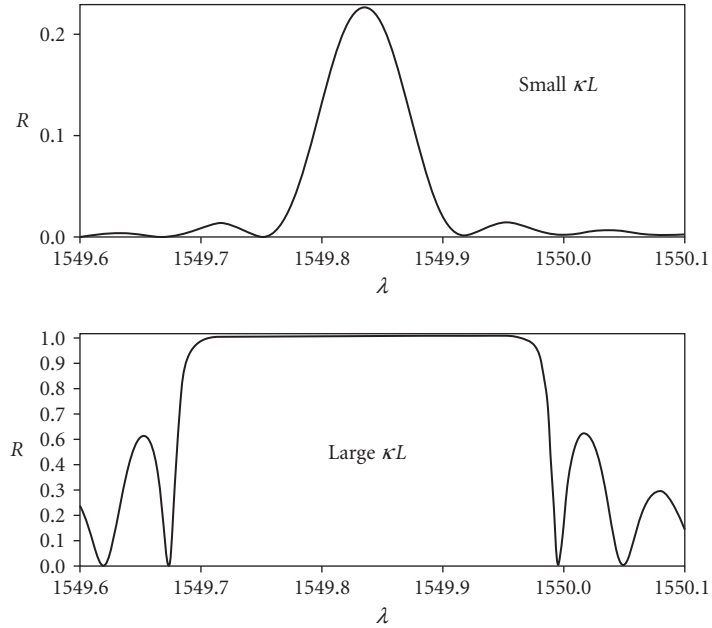


FIGURE 1 Typical reflection spectra for weak (small κL) and strong (large κL) fiber gratings.

The other important property of the grating is its bandwidth, which is a measure of the wavelength range over which the grating reflects light. The bandwidth of a fiber grating that is most easily measured is its full width at half-maximum, $\Delta\lambda_{\text{FWHM}}$, of the central reflection peak, which is defined as the wavelength interval between the 3-dB points. That is the separation in the wavelength between the points on either side of the Bragg wavelength where the reflectivity has decreased to 50 percent of its maximum value. However, a much easier quantity to calculate is the bandwidth, $\Delta\lambda_0 = \lambda_0 - \lambda_B$, where λ_0 is the wavelength where the first zero in the reflection spectra occurs. This bandwidth can be found by calculating the difference in the propagation constants, $\Delta\beta_0 = \beta_0 - \beta_B$, where $\beta_0 = 2\pi N_{\text{eff}}/\lambda_0$ is the propagation constant at wavelength λ_0 for which the reflectivity is first zero, and $\beta_B = 2\pi N_{\text{eff}}/\lambda_B$ is the propagation constant at the Bragg wavelength for which the reflectivity is maximum.

In the case of weak gratings ($\kappa L < 1$), $\Delta\beta_0 = \beta_0 - \beta_B = \pi/L$, from which it can be determined that $\Delta\lambda_{\text{FWHM}} \sim \Delta\lambda_0 = \lambda_B^2/2N_{\text{eff}}L$; the bandwidth of a weak grating is inversely proportional to the grating length L . Thus, long, weak gratings can have very narrow bandwidths. The first Bragg grating written in fibers^{1,2} was more than 1 m long and had a bandwidth less than 100 MHz, which is an astonishingly narrow bandwidth for a reflector of visible light. On the other hand, in the case of a strong grating ($\kappa L > 1$), $\Delta\beta_0 = \beta_0 - \beta_B = 4\kappa$ and $\Delta\lambda_{\text{FWHM}} \sim 2\Delta\lambda_0 = 4\lambda_B^2\kappa/\pi N_{\text{eff}}$. For strong gratings, the bandwidth is directly proportional to the coupling coefficient κ and is independent of the grating length.

17.5 FABRICATION OF FIBER GRATINGS

Writing a fiber grating optically in the core of an optical fiber requires irradiating the core with a periodic interference pattern. Historically, this was first achieved by interfering light that propagated in a forward direction along an optical fiber with light that was reflected from the fiber end and propagated in a backward direction.¹ This method for forming fiber gratings is known as the *internal*

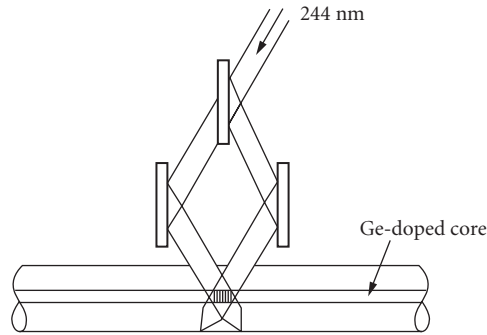


FIGURE 2 Schematic diagram illustrating the writing of an FBG using the transverse holographic technique.

writing technique, and the gratings were referred to as *Hill gratings*. The Bragg gratings, formed by internal writing, suffer from the limitation that the wavelength of the reflected light is close to the wavelength at which they were written (i.e., at a wavelength in the blue-green spectral region).

A second method for fabricating fiber gratings is the *transverse holographic technique*,¹⁴ which is shown schematically in Fig. 2. The light from an ultraviolet source is split into two beams that are brought together so that they intersect at an angle θ . As Fig. 2 shows, the intersecting light beams form an interference pattern that is focused using cylindrical lenses (not shown) on the core of the optical fiber. Unlike the internal writing technique, the fiber core is irradiated from the side, thus giving rise to its name *transverse holographic technique*. The technique works because the fiber cladding is transparent to the ultraviolet light, whereas the core absorbs the light strongly. Since the period Λ of the grating depends on the angle θ between the two interfering coherent beams through the relationship $\Lambda = \lambda_{UV} / 2 \sin(\theta/2)$, Bragg gratings can be made that reflect light at much longer wavelengths than the ultraviolet light that is used in the fabrication of the grating. Most important, FBGs can be made that function in the spectral regions that are of interest for fiber-optic communication and optical sensing.

A third technique for FBG fabrication is the *phase mask technique*,¹⁵ which is illustrated in Fig. 3. The phase mask is made from a flat slab of silica glass, which is transparent to ultraviolet light. On one of the flat surfaces, a one-dimensional periodic surface relief structure is etched using photolithographic techniques. The shape of the periodic pattern approximates a square wave in profile. The optical fiber is placed almost in contact with and at right angles to the corrugations of the phase mask, as shown in Fig. 3. Ultraviolet light, which is incident normal to the phase mask, passes through and is diffracted by the periodic corrugations of the phase mask. Normally, most of the diffracted light is contained in the 0, +1, and -1 diffracted orders. However, the phase mask is designed to suppress the diffraction into the zero order by controlling the depth of the corrugations in the phase mask. In practice, the amount of light in the zero order can be reduced to less than 5 percent with approximately 80 percent of the total light intensity divided equally in the ± 1 orders. The two ± 1 diffracted-order beams interfere to produce a periodic pattern that photoimprints a corresponding grating in the optical fiber. If the period of the phase mask grating is Λ_{mask} , the period of the photoimprinted index grating is $\Lambda_{\text{mask}}/2$. Note that this period is independent of the wavelength of ultraviolet light that irradiates the phase mask.

The phase mask technique has the advantage of greatly simplifying the manufacturing process for Bragg gratings, while yielding high-performance gratings. In comparison with the holographic technique, the phase mask technique offers easier alignment of the fiber for photoimprinting, reduced stability requirements on the photoimprinting apparatus, and lower coherence requirements on the ultraviolet laser beam, thereby permitting the use of a cheaper ultraviolet excimer laser source. Furthermore, there is the possibility of manufacturing several gratings at once in a single exposure by irradiating parallel fibers through the phase mask. The capability to manufacture high-performance gratings at a

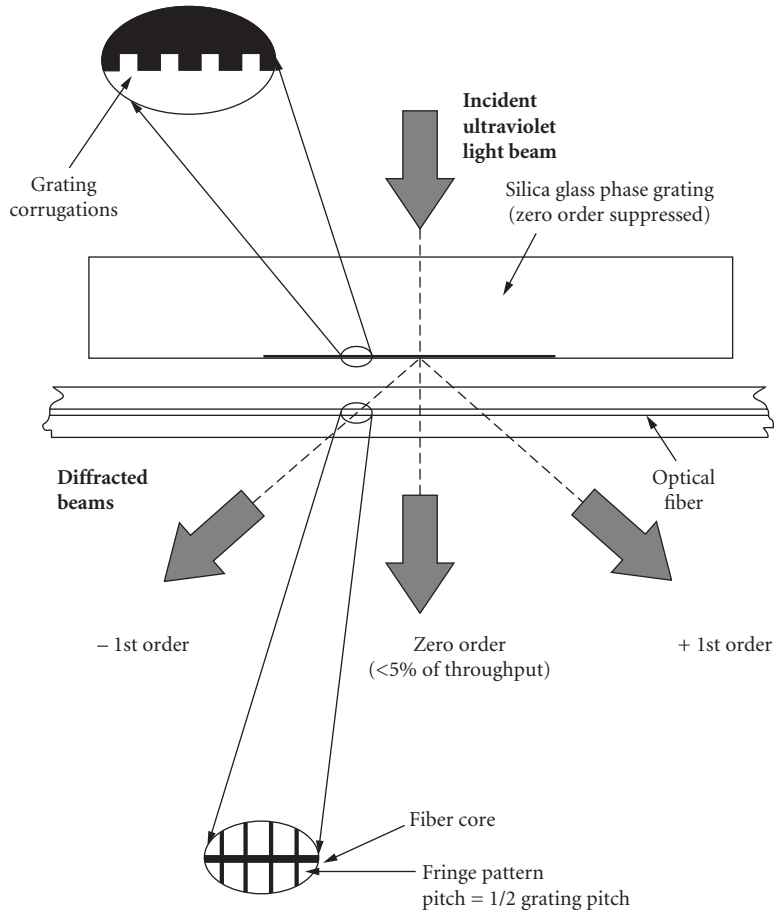


FIGURE 3 Schematic diagram of the phase mask technique for the manufacture of fiber Bragg gratings.

low per-unit grating cost is critical for the economic viability of using gratings in some applications. A draw-back of the phase mask technique is that a separate phase mask is required for each different Bragg wavelength. However, some wavelength tuning is possible by applying tension to the fiber during the photoimprinting process; the Bragg wavelength of the relaxed fiber will shift by ~ 2 nm.

The phase mask technique not only yields high-performance devices, but is also very flexible in that it can be used to fabricate gratings with controlled spectral response characteristics. For instance, the typical spectral response of a finite-length grating with a uniform index modulation along the fiber length has secondary maxima on both sides of the main reflection peak. In applications like wavelength-division multiplexing, this type of response is not desirable. However, if the profile of the index modulation Δn along the fiber length is given a bell-like functional shape, these secondary maxima can be suppressed.¹⁶ The procedure is called *apodization*. Apodized fiber gratings have been fabricated using the phase mask technique, and suppressions of the sidelobes of 30 to 40 dB have been achieved,^{17,18}

Figure 4 shows the spectral response of two Bragg gratings with the same full width at half-maximum (FWHM). One grating exhibits large sidebands, whereas the other has much-reduced sidebands. The one with the reduced sidebands is a little longer and has a coupling coefficient κ apodized as a second-degree cosine ($\cos^2$) along its length. Apodization has one disadvantage:

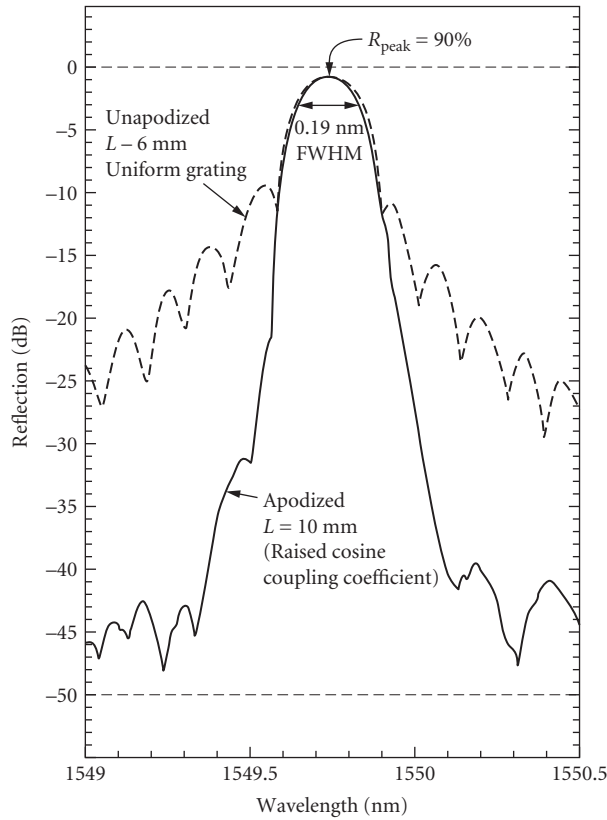


FIGURE 4 Comparison of an unapodized fiber grating's spectral response with that of an apodized fiber grating having the same bandwidth (FWHM).

It decreases the effective length of the Bragg grating. Therefore, to obtain fiber gratings having the same FWHM, the apodized fiber grating has a longer length than the equivalent-bandwidth unapodized fiber grating.

The phase mask technique has been extended to the fabrication of chirped or aperiodic fiber gratings. *Chirping* means varying the grating period along the length of the grating in order to broaden its spectral response. Aperiodic or chirped gratings are desirable for making dispersion compensators¹⁹ or filters having broad spectral responses. The first chirped fiber gratings were made using a double-exposure technique.²⁰ In the first exposure, an opaque mask is positioned between the fiber and the ultraviolet beam blocking the light from irradiating the fiber. The mask is then moved slowly out of the beam at a constant velocity to increase continuously the length of the fiber that is exposed to the ultraviolet light. A continuous change in the photoinduced refractive index is produced that varies linearly along the fiber length with the largest index change occurring in the section of fiber that is exposed to ultraviolet light for the longest duration. In a second exposure, a fiber grating is photoimprinted in the fiber by using the standard phase mask technique. Because the optical pitch of a fiber grating depends on both the refractive index and the mechanical pitch (i.e., optical pitch = $N_{\text{eff}}\Lambda$), the pitch of the photoimprinted grating is effectively chirped, even though its mechanical period is constant. Following this demonstration, a variety of other methods have been developed to manufacture gratings that are chirped permanently^{21,22} or that have an adjustable chirp.^{23,24}

The phase mask technique can also be used to fabricate tilted or blazed gratings. Usually, the corrugations of the phase mask are oriented normal to the fiber axis, as shown in Fig. 3. However, if the corrugations of the phase mask are oriented at an angle to the axis of the fiber, the photoimprinted grating is tilted or blazed. Such fiber gratings couple light out from the bound modes of the fiber to either the cladding modes or the radiation modes. Tilted gratings have applications in fabricating fiber taps.²⁵ If the grating is simultaneously blazed and chirped, it can be used to fabricate an optical spectrum analyzer.²⁶

Another approach to grating fabrication is the *point-by-point technique*,²⁷ also developed at CRC. In this method, each index perturbation of the grating is written point by point. For gratings with many index perturbations, the method is not very efficient. However, it has been used to fabricate micro-Bragg gratings in optical fibers,²⁸ but it is most useful for making coarse gratings with pitches of the order of 100 μm that are required for LP_{01} to LP_{11} mode converters²⁷ and polarization mode converters.¹¹ The interest in coarse period gratings has increased lately because of their use in long-period fiber-grating band-rejection filters²⁹ and fiber-amplifier gain equalizers.³⁰

17.6 THE APPLICATION OF FIBER GRATINGS

Hill and Meltz⁶ provide an extensive review of the many potential applications of fiber gratings in lightwave communication systems and in optical sensor systems. Our purpose here is to note that a common problem in using FBGs is that a transmission device is usually desired, whereas FBGs function as reflection devices. Thus, means are required to convert the reflection spectral response into a transmission response. This can be achieved using a Sagnac loop,³¹ a Michelson (or Mach-Zehnder) interferometer,³² or an optical circulator. Figure 5 shows an example of how this is achieved for the case of a multichannel dispersion compensator using chirped or aperiodic fiber gratings.

In Fig. 5a, the dispersion compensator is implemented using a Michelson interferometer. Each wavelength channel ($\lambda_1, \lambda_2, \lambda_3$) requires a pair of identically matched FBGs, one in each arm of the interferometer. Since it is difficult to fabricate identical Bragg gratings (i.e., having the same

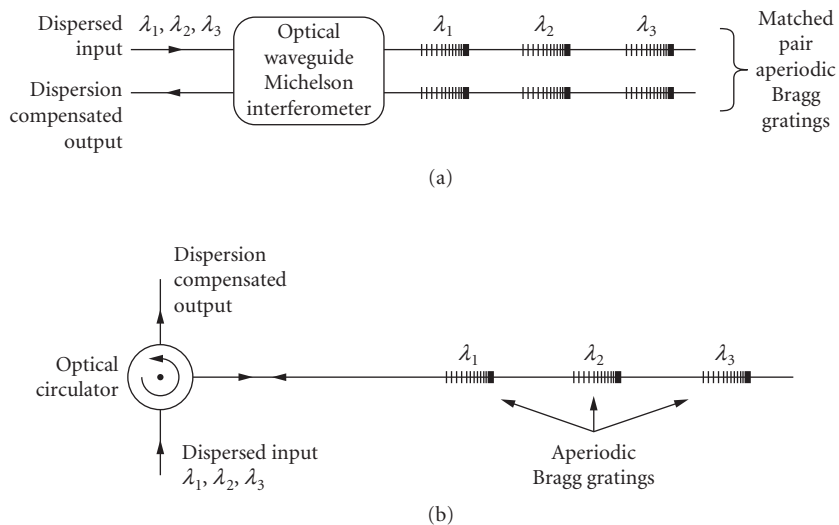


FIGURE 5 Schematic diagram of a multichannel dispersion compensator that is formed by using (a) a Michelson interferometer and (b) an optical circulator.

resonant wavelength and chirp), this configuration for the dispersion compensator has not yet been demonstrated. However, a wavelength-selective device that requires matched grating pairs has been demonstrated.^{33,34} An additional disadvantage of the Michelson interferometer configuration being an interferometric device is that it would require temperature compensation. The advantage of using a Michelson interferometer is that it can be implemented in all-fiber or planar-integrated optics versions.

Figure 5b shows the dispersion compensator implemented using an optical circulator. In operation, light that enters through the input port is routed by the circulator to the port with the Bragg gratings. All of the light that is reflected by the FBGs is routed to the output channel. This configuration requires only one chirped FBG per wavelength channel and is the preferred method for implementing dispersion compensators using FBGs. The only disadvantage of this configuration is that the optical circulator is a bulk optic device (or microoptic device) that is relatively expensive compared with the all-fiber Michelson interferometer.

17.7 REFERENCES

1. K. O. Hill, Y. Fujii, D. C. Johnson, et al., "Photosensitivity in Optical Fiber Waveguides: Application to Reflection Filter Fabrication," *Applied Physics Letters* **32**(10):647–649 (1978).
2. B. S. Kawasaki, K. O. Hill, D. C. Johnson, et al., "Narrow-Band Bragg Reflectors in Optical Fibers," *Optics Letters* **3**(8):66–68 (1978).
3. K. O. Hill, B. Malo, F. Bilodeau, et al., "Photosensitivity in Optical Fibers," *Annual Review of Material Science* **23**:125–157 (1993).
4. I. Bennion, J. A. R. Williams, L. Zhang, et al., "Tutorial Review, UV-Written In-Fibre Bragg Gratings," *Optical and Quantum Electronics* **28**:93–135 (1996).
5. B. Pommellec, P. Niay, M. Douay, et al., "The UV-Induced Refractive Index Grating in Ge:SiO₂ Preforms: Additional CW Experiments and the Macroscopic Origin of the Change in Index," *Journal of Physics D, Applied Physics* **29**:1842–1856 (1996).
6. Kenneth O. Hill and Gerald Meltz, "Fiber Bragg Grating Technology Fundamentals and Overview," *Journal of Lightwave Technology* **15**(8):1263–1276 (1997).
7. A. D. Kersey, M. A. Davis, H. J. Patrick, et al., "Fiber Grating Sensors," *Journal of Lightwave Technology* **15**(8):1442–1463 (1997).
8. T. Erdogan, V. Mizrahi, P. J. Lemaire, et al., "Decay of Ultraviolet-Induced Fiber Bragg Gratings," *Journal of Applied Physics* **76**(1):73–80 (1994).
9. P. J. Lemaire, R. M. Atkins, V. Mizrahi, et al., "High Pressure H₂ Loading as a Technique for Achieving Ultrahigh UV Photosensitivity and Thermal Sensitivity in GeO₂ Doped Optical Fibres," *Electronics Letters* **29**(13):1191–1193 (1993).
10. F. Bilodeau, B. Malo, J. Albert, et al., "Photosensitization of Optical Fiber and Silica-on-Silicon/Silica Waveguides," *Optics Letters* **18**(12):953–955 (1993).
11. K. O. Hill, F. Bilodeau, B. Malo, et al., "Birefringent Photosensitivity in Monomode Optical Fibre: Application to the External Writing of Rocking Filters," *Electronic Letters* **27**(17):1548–1550 (1991).
12. Alan Miller, "Fundamental Optical Properties of Solids," in *Handbook of Optics*, edited by M. Bass, McGraw-Hill, New York, 1995, vol. 1, pp. 9–15.
13. D. K. W. Lam and B. K. Garside, "Characterization of Single-Mode Optical Fiber Filters," *Applied Optics* **20**(3):440–445 (1981).
14. G. Meltz, W. W. Morey, and W. H. Glenn, "Formation of Bragg Gratings in Optical Fibers by a Transverse Holographic Method," *Optics Letters* **14**(15):823–825 (1989).
15. K. O. Hill, B. Malo, F. Bilodeau, et al., "Bragg Gratings Fabricated in Monomode Photosensitive Optical Fiber by UV Exposure Through a Phase Mask," *Applied Physics Letters* **62**(10):1035–1037 (1993).
16. M. Matsuhara and K. O. Hill, "Optical-Waveguide Band-Rejection Filters: Design," *Applied Optics* **13**(12):2886–2888 (1974).

17. B. Malo, S. Thériault, D. C. Johnson, et al., "Apodised In-Fibre Bragg Grating Reflectors Photoimprinted Using a Phase Mask," *Electronics Letters* **31**(3):223–224 (1995).
18. J. Albert, K. O. Hill, B. Malo, et al., "Apodisation of the Spectral Response of Fibre Bragg Gratings Using a Phase Mask with Variable Diffraction Efficiency," *Electronics Letters* **31**(3):222–223 (1995).
19. K. O. Hill, "Aperiodic Distributed-Parameter Waveguides for Integrated Optics," *Applied Optics* **13**(8): 1853–1856 (1974).
20. K. O. Hill, F. Bilodeau, B. Malo, et al., "Chirped In-Fibre Bragg Grating for Compensation of Optical-Fiber Dispersion," *Optics Letters* **19**(17):1314–1316 (1994).
21. K. Sugden, I. Bennion, A. Molony, et al., "Chirped Gratings Produced in Photosensitive Optical Fibres by Fibre Deformation during Exposure," *Electronics Letters* **30**(5):440–442 (1994).
22. K. C. Byron and H. N. Rourke, "Fabrication of Chirped Fibre Gratings by Novel Stretch and Write Techniques," *Electronics Letters* **31**(1):60–61 (1995).
23. D. Garthe, R. E. Epworth, W. S. Lee, et al., "Adjustable Dispersion Equaliser for 10 and 20 Gbit/s over Distances up to 160 km," *Electronics Letters* **30**(25):2159–2160 (1994).
24. M. M. Ohn, A. T. Alavie, R. Maaskant, et al., "Dispersion Variable Fibre Bragg Grating Using a Piezoelectric Stack," *Electronics Letters* **32**(21):2000–2001 (1996).
25. G. Meltz, W. W. Morey, and W. H. Glenn, "In-Fiber Bragg Grating Tap," presented at the *Conference on Optical Fiber Communications*, OFC'90, San Francisco, CA, 1990 (unpublished).
26. J. L. Wagener, T. A. Strasser, J. R. Pedrazzani, et al., "Fiber Grating Optical Spectrum Analyzer Tap," presented at the *IOOC-ECOC'97*, Edinburgh, UK, 1997 (unpublished).
27. K. O. Hill, B. Malo, K. A. Vineberg, et al., "Efficient Mode Conversion in Telecommunication Fibre Using Externally Written Gratings," *Electronics Letters* **26**(16):1270–1272 (1990).
28. B. Malo, K. O. Hill, F. Bilodeau, et al., "Point-by-Point Fabrication of Micro-Bragg Gratings in Photosensitive Fibre Using Single Excimer Pulse Refractive Index Modification Techniques," *Electronic Letters* **29**(18):1668–1669 (1993).
29. A. M. Vengsarkar, P. J. Lemaire, J. B. Judkins, et al., "Long-Period Fiber Gratings as Band-Rejection Filters," presented at the *Optical Fiber Communication conference*, OFC'95, San Diego, CA, 1995 (unpublished).
30. A. M. Vengsarkar, J. R. Pedrazzani, J. B. Judkins, et al., "Long-Period Fiber-Grating-Based Gain Equalizers," *Optics Letters* **21**(5):336–338 (1996).
31. K. O. Hill, D. C. Johnson, F. Bilodeau, et al., "Narrow-Bandwidth Optical Waveguide Transmission Filters: A New Design Concept and Applications to Optical Fibre Communications," *Electronics Letters* **23**(9):465–466 (1987).
32. D. C. Johnson, K. O. Hill, F. Bilodeau, et al., "New Design Concept for a Narrowband Wavelength-Selective Optical Tap and Combiner," *Electronics Letters* **23**(13):668–669 (1987).
33. F. Bilodeau, K. O. Hill, B. Malo, et al., "High-Return-Loss Narrowband All-Fiber Bandpass Bragg Transmission Filter," *IEEE Photonics Technology Letters* **6**(1):80–82 (1994).
34. F. Bilodeau, D. C. Johnson, S. Thériault, et al., "An All-Fiber Dense-Wavelength-Division Multiplexer/Demultiplexer Using Photoimprinted Bragg Gratings," *IEEE Photonics Technology Letters* **7**(4):388–390 (1995).

MICRO-OPTICS-BASED COMPONENTS FOR NETWORKING

Joseph C. Palais

Ira A. Fulton School of Engineering
Arizona State University
Tempe, Arizona

18.1 INTRODUCTION

The optical portion of many fiber networks requires a number of functional devices, some of which can be fabricated using small optical components (so-called *micro-optic* components). Micro-optic components are made up of parts which have linear dimensions on the order of a few millimeters. The completed functional device may occupy a space a few centimeters on a side. Components to be described in this chapter have the common feature that the fiber transmission link is opened and small (micro-optic) devices are inserted into the gap between the fiber ends to produce a function component. Network components constructed entirely of fibers or constructed in integrated-optic form are described elsewhere in this handbook.

The following sections describe, in order: a generalized component, specific useful network functions, micro-optic subcomponents required to make up the final component, and complete components.

18.2 GENERALIZED COMPONENTS

A generalized fiber-optic component is drawn in Fig. 1. As indicated, input fibers are on the left and output fibers are on the right. Although some components have only a single input port and a single output port, many applications require more than one input and/or output ports. In fact, the number of ports in some devices can be more than 100. The *coupling loss* between any two ports is given, in decibels, by

$$L = -10 \log(P_{\text{out}}/P_{\text{in}}) \quad (1)$$

With respect to Fig. 1, P_{in} refers to the input power at any of the ports on the left and P_{out} refers to the output power at any of the ports on the right. Because we are only considering passive components in this section, P_{out} will be less than P_{in} and the loss will be a positive number.

Insertion loss refers to the coupling loss between any two ports where coupling is desired and *isolation* (or *directionality*) refers to the coupling loss between any two ports where coupling is unwanted. *Excess loss* is the fraction of input power that does not emerge from any of the desired output ports, as expressed in decibels. It is the sum of all the useful power out divided by the input power.



FIGURE 1 The generalized component.

18.3 NETWORK FUNCTIONS

Functions useful for many fiber-optic communications applications are described in the following sections.

Attenuators

Attenuators reduce the amount of power flowing through the fiber system. Both fixed and variable attenuators are available. The applications include testing of receiver sensitivities (varying the attenuation changes the amount of power incident on the receiver) and protecting a receiver from saturating due to excess incident power. Attenuation from a few tenths of a decibel to more than 50 dB are sometimes required.

Power Splitters and Directional Couplers

These devices distribute input power from a single fiber to two or more output fibers. The component design controls the fraction of power delivered to each of the output ports. Applications include power distribution in local area networks (LANs) and in subscriber networks. The most common splitters and couplers have a single input and equal distribution of power among each of two outputs, as shown schematically in Fig. 2a. For an ideal three-port splitter (one with no excess loss), half the input power emerges from each of the two output ports. The insertion loss, as calculated from Eq. (1) with a ratio of powers of 0.5, yields a 3-dB loss to each of the two output ports. Any excess loss is added to the 3 dB.

A splitter with more than two output ports can be constructed by connecting several three-port couplers in a tree pattern as indicated schematically in Fig. 2b. Adding more splitters in the same manner allows coupling from one input port to 8, 16, 32 (and more) output ports.

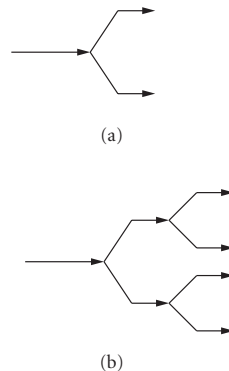


FIGURE 2 Power splitters. (a) 1:2 split and (b) 1:4 split.

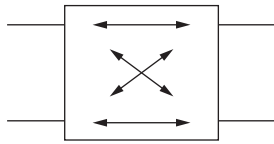


FIGURE 3 Four-port directional coupler.

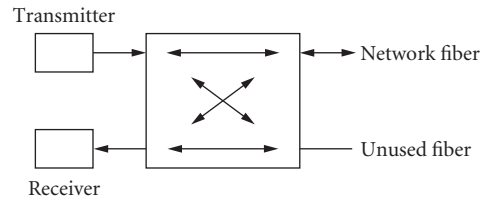


FIGURE 4 LAN terminal illustrating application of the directional coupler.

Adding a fourth port, as in Fig. 3, creates a *directional coupler*. The arrows in the figure show the allowed directions of wave travel through the coupler. An input beam is split between two output ports and is isolated from the fourth. By proper component design, any desired power-splitting ratio can be obtained. One application of the directional coupler is to the distribution network of a local area network, where simultaneous transmission and reception are required. Figure 4 illustrates this usage at one LAN terminal.

Isolators

An isolator is a one-way transmission line. It permits the flow of optical power in just one direction (the forward direction). Applications include protection of a transmitting laser diode from back reflections. Such reflections increase the noise in the system by disrupting the diode's operation. Isolators also improve the stability of fiber amplifiers by minimizing the possibility of feedback, which causes unwanted oscillations in such devices.

Circulators

In a circulator, power into the first port emerges from the second, while power into the second port emerges from a third. This behavior repeats at each successive input port until power into the last port emerges from the first. Practical circulators are typically three- or four-port devices.

Using a circulator, efficient two-way transmission (*full-duplex*) along a single fiber at a single wavelength is possible. The circulator separates the transmitting and receiving beams of light at each terminal, as illustrated in Fig. 5.

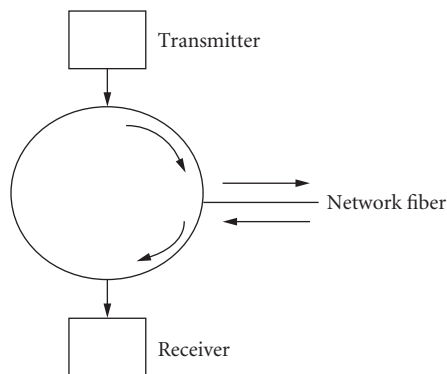


FIGURE 5 An optical circulator separates transmitted and received messages at a terminal.

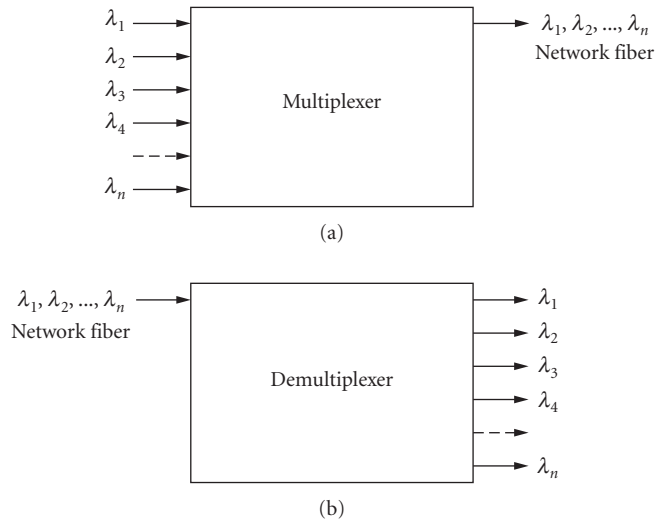


FIGURE 6 (a) A multiplexer combines different wavelength channels onto a single fiber for transmission. (b) A demultiplexer separates several incoming channels at different wavelengths and directs them to separate receivers.

Multiplexers/Demultiplexers/Duplexers

The multiplexer and demultiplexer are heavily utilized in fiber-optic wavelength-division multiplexed (WDM) systems. The *multiplexer* combines beams of light from the different transmitters (each at a slightly shifted wavelength) onto the single transmission fiber. The *demultiplexer* separates the individual wavelengths transmitted and guides the separate channels to the appropriate optical receivers. These functions are illustrated in Fig. 6. Requirements for multiplexers/demultiplexers include combining and separating independent channels less than a nanometer apart, accommodating numerous (in some cases over 100) channels. A frequency spacing between adjacent channels of 100 GHz corresponds to a wavelength spacing of 0.8 nm for wavelengths near 1.55 μm . Insertion losses can be as low as a few tenths of a decibel and isolations of 40 dB or more.

The *duplexer* allows for simultaneous two-way transmission along a single fiber. The wavelengths are different for the transmitting and receiving light beam. The duplexer separates the beams as indicated in Fig. 7, where λ_1 is the transmitting wavelength and λ_2 is the receiving wavelength.

Mechanical Switches

Operationally, an optical switch acts just like an electrical switch. Mechanical movement of some part (as implied schematically in Fig. 8) causes power entering one port to be directed to one of two

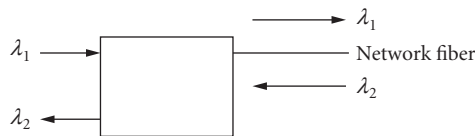


FIGURE 7 A duplexer allows two-way transmission along a single network fiber.

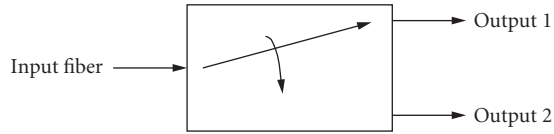


FIGURE 8 Mechanical optical switch.

or more output ports. Such devices are useful in testing of fiber components and systems and in other applications, such as bypassing inoperative nodes in a local area network. Insertion losses less than 0.10 dB and isolations greater than 50 dB are reasonable requirements.

18.4 SUBCOMPONENTS

Micro-optic subcomponents which form part of the design of many complete micro-optic component are described in this section.

Prisms

Because of the dispersion in glass prisms, they can operate as multiplexers, demultiplexers, and duplexers. The dispersive property is illustrated in Fig. 9.

Right-angle glass prisms also act as excellent reflectors, as shown in Fig. 10, owing to perfect reflection (total internal reflection) at the glass-to-air interface. The critical angle for the glass-to-air interface is about 41° and the incident ray is beyond that at 45° .

The beam-splitting cube, drawn in Fig. 11, consists of two right-angle prisms cemented together with a thin reflective layer between them. This beam splitter has the advantage over a flat reflective plate in that no angular displacement occurs between the input and output beam directions. This simplifies the alignment of the splitter with the input and output fibers.

Gratings

Ruled reflection gratings are also used in multiplexers and demultiplexers. As illustrated in Fig. 12, the dispersion characteristics of the grating perform the wavelength separation function required of a demultiplexer. The grating has much greater dispersive power than a prism, permitting increased

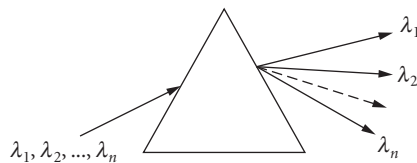


FIGURE 9 A dispersive prism spatially separates different wavelengths. This represents demultiplexing. Reversing the directions of the arrows illustrates combining of different wavelengths. This is multiplexing.

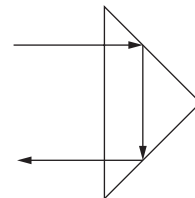
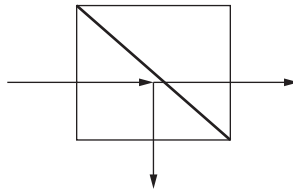
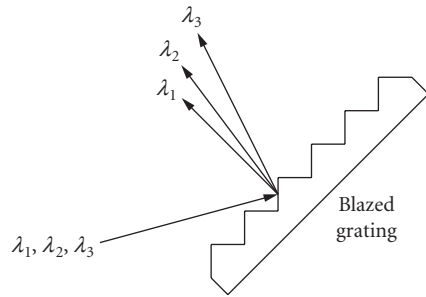


FIGURE 10 Totally reflecting prism.

**FIGURE 11** Beam-splitting cube.**FIGURE 12** Blazed reflection grating operated as a demultiplexer.

wavelength spatial separation. The relationship between the incident and reflected beams, for an incident collimated light beam, is given by the diffraction equation

$$\sin \theta_i + \sin \theta_r = m\lambda/d \quad (2)$$

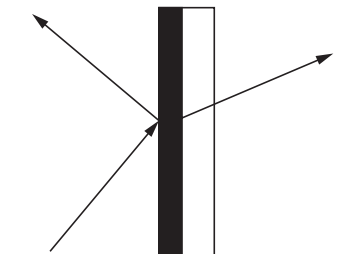
where θ_i and θ_r are the incident and reflected beam angles, d is the separation between adjacent reflecting surfaces, and m is the *order* of the diffraction. Typically, gratings are blazed so as to maximize the power into the first-order beams. As deduced from Eq. (2) for $m = 1$, the diffracted peak occurs at a different angle for different wavelengths. This feature produces the demultiplexing function needed in WDM systems. Reversing the arrows in Fig. 12 illustrates the multiplexing capability of the grating.

Filters

Dielectric-layered filters, consisting of very thin layers of various dielectrics deposited onto a glass substrate, are used to construct multiplexers, demultiplexers, and duplexers. Filters have unique reflectance/transmittance characteristics. They can be designed to reflect at certain wavelengths and transmit at others, thus spatially separating (or combining) different wavelengths as required for WDM applications.

Beam Splitters

A beam-splitting plate, shown in Fig. 13, is a partially silvered glass plate. The thickness of the silvered layer determines the fraction of light transmitted and reflected. In this way, the input beam can be divided in two parts of any desired ratio.

**FIGURE 13** Beam-splitting plate.

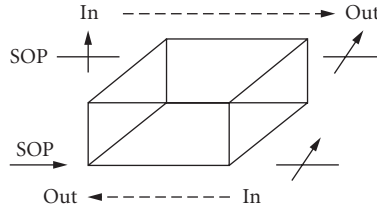


FIGURE 14 Faraday rotator. The dashed arrows indicate the direction of beam travel. The solid arrows represent the wave polarization in the plane perpendicular to the direction of wave travel.

Faraday Rotators

The Faraday rotator produces a nonreciprocal rotation of the plane of polarization. The amount of rotation is given by

$$\theta = VHL \quad (3)$$

where θ is the rotation angle, V is the *Verdet constant* (a measure of the strength of the Faraday effect), H is the applied magnetic field, and L is the length of the rotator. A commonly used rotator material is YIG (yttrium-iron garnet), which has a high value of V .

Figure 14 illustrates the nonreciprocal rotation of the state of polarization (SOP) of the wave. The rotation of a beam traveling from left to right is 45° , while the rotation for a beam traveling from right to left is an additional 45° .

The Faraday rotator is used in the isolator and the circulator.

Polarizers Polarizers based upon dichroic absorbers and polarization prisms using birefringent materials are common. The polarizing beam splitter, illustrated in Fig. 15, is useful in micro-optics applications, such as the optical circulator. The polarizing splitter separates two orthogonally polarized beams.

GRIN-Rod Lens

The subcomponents discussed in the last few sections perform the operations indicated in their descriptions. The problem is that they cannot be directly inserted into a fiber transmission line. To insert one of the subcomponents into the fiber link requires that the fiber be opened to produce a gap. The subcomponent would then fit into the gap. Because the light emerging from a fiber diverges, with

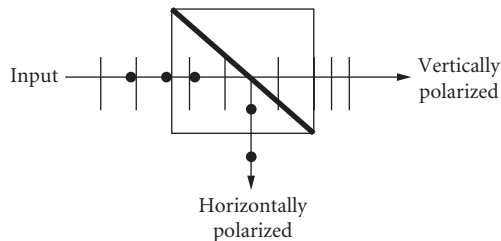


FIGURE 15 Polarizing beam splitter.

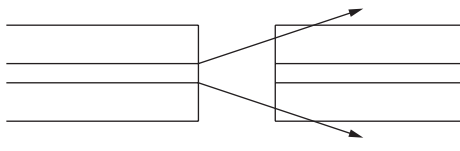


FIGURE 16 Diverging wave emitted from an open fiber couples poorly to the receiving fiber.

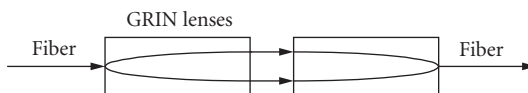


FIGURE 17 Collimating light between two fibers using GRIN-rod lenses.

a gap present the receiving fiber does not capture much of the transmitted light. This situation is illustrated in Fig. 16. The emitted diverging light must be collimated, the required subcomponent (e.g., beamsplitter, grating, etc.) inserted, and the light refocused. A commonly used device for performing this function is the *graded-index rod lens* (GRIN-rod lens). Its use is illustrated in Fig. 17. The diverging light emitted by the transmitting fiber is collimated by the first GRIN-rod lens. The collimated beam is refocused onto the receiving fiber by the second GRIN-rod lens. The collimation is sufficient such that a gap of 20 mm introduces less than 0.5 dB excess loss.¹ This allows for the insertion of beam-modifying devices of the types described in the preceding sections (e.g., prisms, gratings, and beamsplitters) in the gap with minimum added loss.

MEMS Mirrors

Tiny mirrors are the foundation of the *micro-electromechanical systems* (MEMS) optical switch. These mirrors can be fabricated in several ways. Two examples are thin-film mirrors and bulk mirrors, sketched respectively in Figs. 18 and 19. In the thin-film mirror, epitaxial layers are deposited on a silicon substrate. The moveable mirror is formed by removing material from underneath these layers but leaving silicon hinges as indicated in the figure. The bulk mirror is formed by etching it from the silicon substrate.

Mirror movement can be controlled by electrostatic, electromagnetic, piezoelectric, and thermal effects. In electrostatic control, two plates are oppositely charged by placing a voltage across. The resulting attraction causes them to attract and move toward each other. Electromagnetic control uses the forces of attraction between two magnetic circuits. Piezoelectric control is obtained when a voltage placed across a material causes the dimensions of that body to change. Thermal control uses the deformation of a resistive body that occurs when it is heated by passing a current through it.

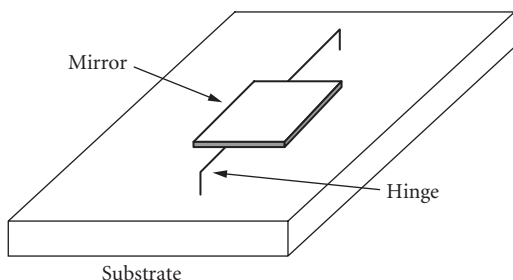


FIGURE 18 Thin-film MEMS mirror.

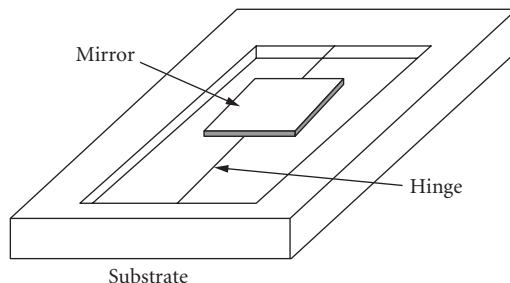


FIGURE 19 Bulk-silicon MEMS mirror.

18.5 COMPONENTS

The subcomponents introduced in the last section are combined into useful fiber devices in the manner described in this section.

Attenuators

The simplest attenuator is produced by a gap introduced between two fibers, as in Fig. 20. As the gap length increases, so does the loss. Loss is also introduced by a lateral displacement. A variable attenuator is produced by allowing the gap (or the lateral offset) to be changeable. A disc whose absorption differs over different parts may also be placed between the fibers. The attenuation is varied by rotating the disk.

In another attenuator design, a small thin flat reflector is inserted at variable amounts into the gap to produce the desired amount of loss.²

Power Splitters and Directional Couplers

A power splitter³ can be constructed as illustrated in Fig. 21. A beam-splitting cube (or a beam-splitting plate) is placed in the gap between two GRIN-rod lenses to connect ports 1 and 2. A third combination of lens and fiber collects the reflected light at port 3. The division of power between the two output fibers is determined by the reflective properties of the splitter itself. Any desired ratio of outputs can be obtained.

If a fourth port is added (Port 4 in Fig. 21) the device is a four-port directional coupler.

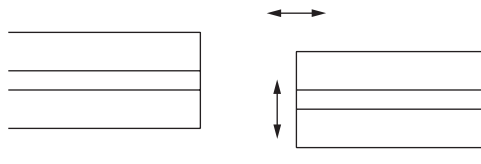


FIGURE 20 Gap attenuator showing relative displacement of the fibers to vary the insertion loss.

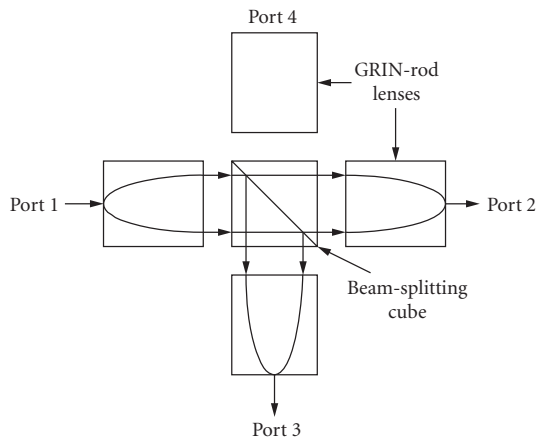


FIGURE 21 Power splitter and (with Port 4 added) four-port directional coupler.

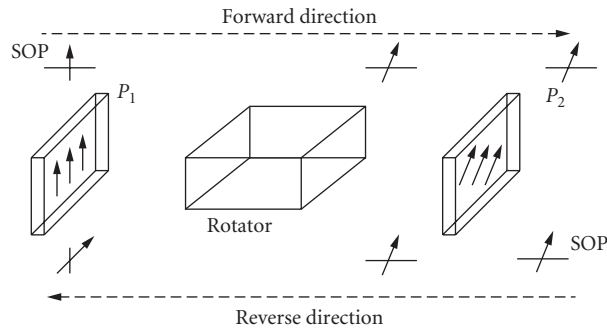


FIGURE 22 Optical isolator. P_1 and P_2 are polarizers.

Isolators and Circulators

The isolator combines the Faraday rotator and two polarizers⁴ as indicated in Fig. 22. The input and output fibers can be coupled to the isolator using GRIN lenses. The vertically polarized beam at the input is rotated by 45° and passed through the output polarizer. Any reflected light is rotated an additional 45° , emerging cross-polarized with respect to the polarizer on the left. In this state, the reflected light will not pass back into the transmitting fiber. Similarly, a light beam traveling from right to left will be cross-polarized at the input polarizer and will not travel further in that direction. The polarizers can be polarizing beam splitters, dichroic polarizers, or polarizing fibers.

A circulator also requires a Faraday rotator and polarizers (polarizing beam splitters or polarizing fiber). Additional components include reflecting prisms, reciprocal 45° rotators, and fiber-coupling devices such as GRIN-rod lenses.⁵

Multiplexers/Demultiplexers/Duplexers

The multiplexer, demultiplexer, and duplexers are fundamentally the same device. The application determines which of the three descriptions is most appropriate. One embodiment is illustrated in Fig. 23 for a two-channel device. As a demultiplexer, the GRIN lens collimates the diverging beam from the network fiber and guides it onto the diffraction grating. The grating redirects the beam according to its wavelength. The GRIN lens then focuses the various wavelengths onto the output fibers for reception. As a multiplexer, the operation is just reversed with the “receiver fibers” replaced by transmitter fibers and all arrows reversed. As a duplexers, one of the two “receiver fibers” becomes a transmitter fiber.

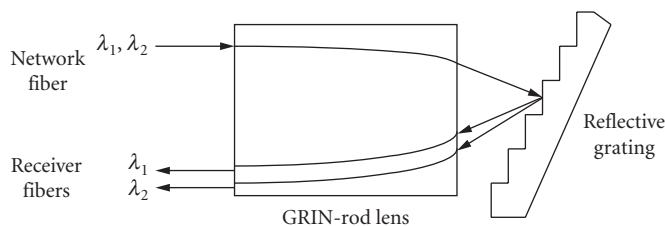


FIGURE 23 Two-channel demultiplexer. Only the beam's central rays are drawn. To operate as a multiplexer the arrows are reversed. To operate as a duplexers, the arrows for just one of the two wavelengths is reversed.

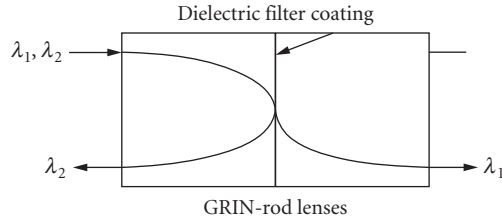


FIGURE 24 Filter-based multiplexer/demultiplexer.

Other configurations also use the diffraction grating, including one incorporating a concave reflector for properly collimating and focusing the beams between input and output fibers.⁶ Micro-optic grating-based devices can accommodate more than 100 WDM channels, with wavelength spacing on the order of 0.4 nm.

A filter-based multiplexer/demultiplexer appears in Fig. 24. The reflective coating transmits wavelength λ_1 and reflects wavelength λ_2 . The device is illustrated as a demultiplexer. Again, by reversing the directions of the arrows, the device becomes a multiplexer. Filter-based multiplexers/demultiplexers can be extended to several channels in the micro-optical form, essentially by cascading several devices of the type just described.

Mechanical Switches⁷

Switching the light beam from one fiber to another one is basically easy. Simply move the transmitting fiber to mechanically align it with the desired receiving fiber. The problem is that even very small misalignments between the two fiber cores introduce unacceptable transmission losses. Several construction strategies have been utilized. Some incorporate a moving fiber and other incorporate a moveable reflector.⁸ In a moveable fiber switch, the fiber can be positioned either manually or by using an electromagnetic force. The switching action in Fig. 25 occurs when the totally reflecting prism moves to align the beam with one or the other of the two output fibers.

A two-dimensional MEMS switch can be constructed as drawn in Fig. 26.⁹ As indicated, light enters the switch from any of the fibers on the left. The mirrors are arranged in an array. Each mirror can be raised into the light path to deflect the beam or can be lowered to allow beam passage. In this manner, light entering any of the input fibers can be directed to any of the output fibers. Collimators (either GRIN or conventional lenses) collimate the entering beams and refocus those exiting.

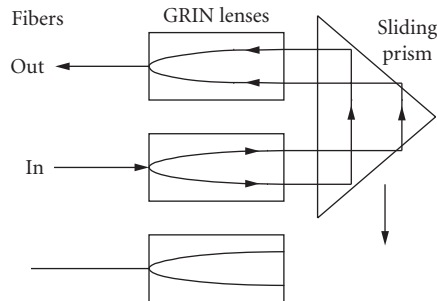


FIGURE 25 Moveable reflecting prism switch.

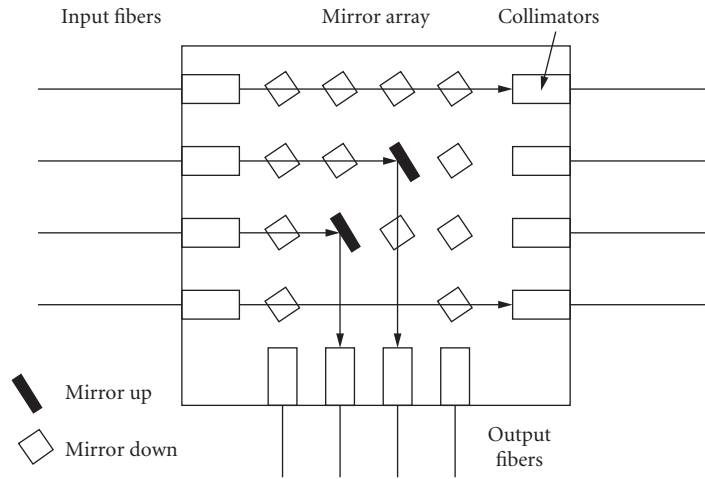


FIGURE 26 Two-dimensional MEMs switch.

18.6 REFERENCES

1. R. W. Gilsdorf, and J. C. Palais, "Single-Mode Fiber Coupling Efficiency with Graded-Index Rod Lenses," *Appl. Opt.* **33**:3440–3445 (1994).
2. C. Marxer, P. Griss, and N. F. de Rooij, "A Variable Optical Attenuator Based on Silicon Micromechanics," *IEEE Photon. Technol. Lett.* **11**:233–235 (1999).
3. C.-L. Chen, *Elements of Optoelectronics and Fiber Optics* (Irwin, Chicago, 1996).
4. R. Ramaswami and K. N. Sivarajan, *Optical Networks: A Practical Perspective* (Morgan Kaufmann, San Francisco, 1998).
5. N. Kashima, *Passive Optical Components for Optical Fiber Transmission* (Artech House, Boston, 1995).
6. J. P. Laude and J. M. Lerner, "Wavelength Division Multiplexing/Demultiplexing (WDM) Using Diffraction Gratings," *SPIE-Application, Theory and Fabrication of Periodic Structures* **503**:22–28 (1984).
7. J. C. Palais, *Fiber Optic Communications*, 5th ed. (Prentice-Hall, Upper Saddle River, N. J., 2005).
8. W. J. Tomlinson, "Applications of GRIN-Rod Lenses in Optical Fiber Communications Systems," *Appl. Opt.* **19**:1123–1138 (1980).
9. Special Issue on Optical MEMS and its Future Trends. *J. Lightwave Tech.* **21**(3) (March 2003).

SEMICONDUCTOR OPTICAL AMPLIFIERS

Jay M. Wiesenfeld

*Bell Laboratories, Alcatel-Lucent
Murray Hill, New Jersey*

Leo H. Spiekman

*Alphion Corp.
Princeton Junction, New Jersey*

19.1 INTRODUCTION

Amplification is a useful, even necessary, function for optical systems, and laser amplifiers have been considered since the early days of quantum electronics. In photonic systems, semiconductor lasers are a fundamental light source. They have the advantage of compactness and efficiency and the ability to emit at many wavelengths, depending on the material system from which they are fabricated. It is therefore natural to make optical amplifiers based on the technology created for semiconductor lasers. Indeed, the first optical amplifier based on a semiconductor laser was reported as early as 1966.¹

A semiconductor optical amplifier (SOA) is a device that provides optical gain based on inversion in a semiconductor medium. Most SOAs are electrically pumped and rely on a p-n junction structure to create a spatial region of optical inversion. As is generally the case for a semiconductor laser, the optical field in a SOA is confined within an optical waveguide, which is an integral part of the structure. In general, the SOA functions as a gain block within an optical system. Again sharing common attributes with the semiconductor laser, SOAs are compact and can be integrated with other elements into photonic integrated circuits (PICs).

SOAs were studied in the early 1980s in the AlGaAs/GaAs material system.² As optical communication systems shifted to the “long wavelength” 1300- and 1550-nm bands, SOAs were intensely studied in the InGaAsP/InP material system.^{3–5} This chapter concentrates on SOAs for telecommunication systems, and so on devices based on the InGaAsP/InP material system. However, the general principles and features discussed herein apply to other material systems as well.

In this chapter, Secs. 19.2 to 19.4 describe the basic principles and properties of individual SOAs. Section 19.2 covers the basic device physics. Section 19.3 describes the fabrication of devices and Sec. 19.4 provides a description of the characterization of the properties of SOAs. Sections 19.5 to 19.8 describe some applications for SOAs, primarily related to optical communication systems and networks. Applications related to simple gain for optical communication signals are covered in Sec. 19.6. Section 19.7 covers applications where SOAs are used as gates, with moderate (for switching) and fast (for modulation) gating speeds. Section 19.8 covers some applications that are enabled by the nonlinear gain and index properties that occur within the SOAs.

The illustrative examples in this chapter correspond to devices that operate within the 1300- or 1550-nm regions. In this chapter, when there is a difference between upper- and lower-case symbols, lower case corresponds to linear units and upper case corresponds to decibel (dB) units.

19.2 DEVICE BASICS

The amplification mechanism in the semiconductor optical amplifier is based on the stimulated emission of photons. Just like in any laser, the arrival of a photon in the excited medium prompts the generation of a second photon—a perfect copy in wavelength, phase, and polarization. The basic mechanism of all stimulated emission is the same, but the physical interaction with matter depends on the specific system. For example, in neodymium-doped YAG lasers a transition between different energy levels of an excited Nd ion is used as the energy source for the new photon, and the emission wavelength is determined by the energy difference of those levels. In the SOA on the other hand, the light quantum is generated by an electron in the conduction band of a semiconductor that recombines with a hole in the valence band. For this reason, the wavelength at which photons can be amplified in a SOA is determined fully by the band structure of the semiconductor material. The energy of emitted photons is in a range beginning at the *band gap energy*, as illustrated in Fig. 1. A larger band gap results in shorter wavelength emission, and vice versa.

In order for amplification to occur, a *population inversion* must be present, that is, more carriers must be in the excited state (electrons in the conduction band, leaving holes in the valence band) than in the ground state. Otherwise, emitted photons would be readily reabsorbed. This inversion is usually accomplished by electrical pumping, applying a forward current to a semiconductor diode: electrons are injected from the n-doped side and holes from the p-doped side, and where they meet, they recombine.

In a good optical amplifier this recombination needs to occur in a region of the device through which the optical input signal is propagating, so that it can be efficiently amplified. To this end, the SOA consists of an optical waveguide in which the input signal is confined to a core with a higher index of refraction than the surrounding material. This higher index core has at the same time a lower band gap than the cladding, so that injected carriers are trapped in a potential well and overlap spatially with the signal photons (see Fig. 2). The configuration just described is equivalent to that of a semiconductor laser, and in fact a SOA can be viewed as a minor variation on a laser.

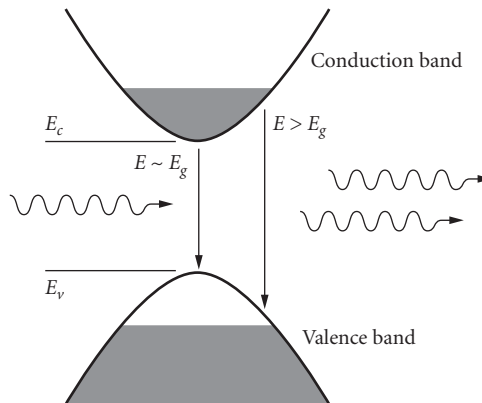


FIGURE 1 Stimulated emission of photons in a semiconductor as a result of population inversion. Recombination of electrons and holes close to the band edges results in emission of photons with an energy close to that of the band gap, while recombination of carriers from higher occupied states within the bands produces photons with a shorter wavelength.

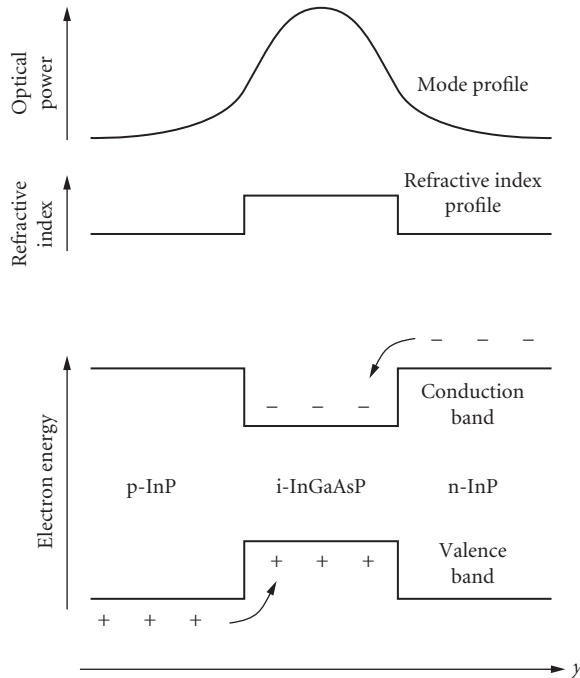


FIGURE 2 Confinement of carriers and photons in a *double heterostructure* waveguide consisting of a lower band gap, higher refractive index material, the *active layer*, sandwiched between layers of higher band gap material with a lower index. By appropriately applying p- and n-doping, a diode structure is formed in which excited states are easily created by injecting a forward current. Note that the lower index cladding material also has a larger band gap and is generally transparent to the emission from the active region.

ASE Noise

In the absence of an input signal, photons are generated in an excited medium by *spontaneous emission*. In a pumped semiconductor this occurs due to the spontaneous recombination of electron-hole pairs. Without these random events, in a laser the lasing action would never start; in a SOA they are a source of noise.

Spontaneous emission occurs over a range of wavelengths corresponding to the occupied excited states of the semiconductor bands, and in all spatial directions. The fraction that couples to the waveguide will subsequently give rise to stimulated emission, and for this reason we speak of *amplified spontaneous emission* (ASE).

In a laser, a feedback mechanism is present that causes the initial ASE to make round trips through the device. When enough current is injected to make this process self-sustaining, lasing action starts. In an optical amplifier, on the other hand, we go to great lengths to avoid optical feedback, so that amplification occurs in a single pass through the device, in a strictly *traveling-wave* fashion. In this case, an ASE spectrum emanates from the device without signs of lasing. An example is shown in Fig. 3. Any residual feedback, for example in the form of reflections from both ends of the amplifier, appears in the spectrum as a ripple, caused by constructive and destructive resonance.

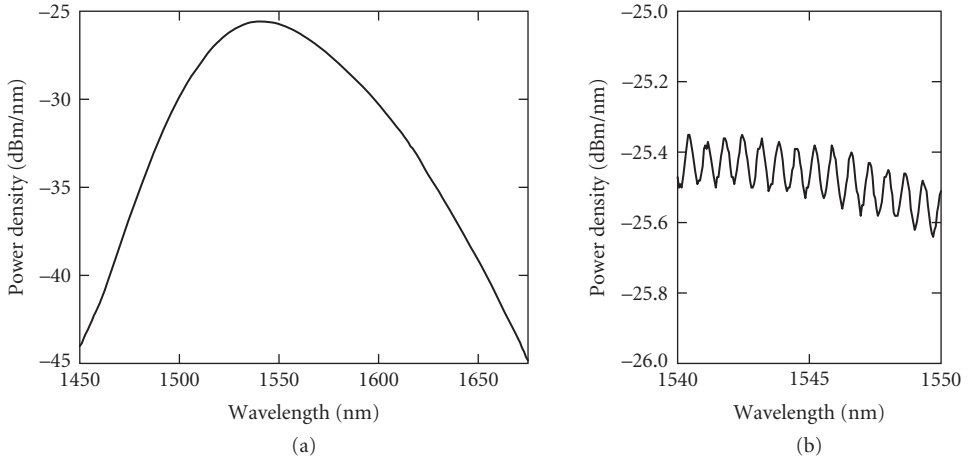


FIGURE 3 (a) Typical amplified spontaneous emission spectrum of a traveling wave SOA. (b) ASE ripple caused by residual reflections from the SOA chip facets.

Gain

A complete SOA typically consists of a semiconductor chip with a waveguide in which the amplification occurs, and two fibers that couple the signal into and out of the chip using lenses, as shown in Fig. 4. A signal coupled into the chip experiences gain as it propagates along the waveguide, according to the process of stimulated emission described above. The chip gain can be written as

$$g_{\text{chip}} = \frac{p_{\text{out}}}{p_{\text{in}}} = e^{(g_{\text{wg}} - \alpha)L} \quad (1)$$

with p_{in} and p_{out} the chip-coupled input and output powers, g_{wg} the gain of the active waveguide per unit length, α a loss term that includes propagation loss and absorption through mechanisms other than producing electron-hole pairs in the active layer, and L the length of the waveguide.

We already saw in Fig. 1 that gain at different wavelengths is generated by electron-hole pairs from different occupied states in the bands. The gain spectrum of the SOA is determined by the semiconductor band structure, and the extent to which it is filled with free carriers.

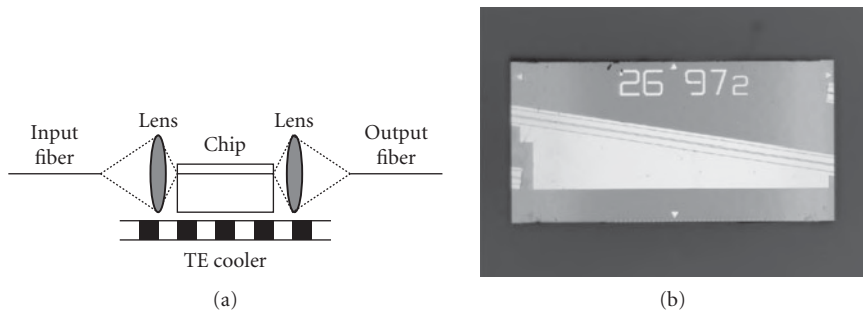


FIGURE 4 (a) Typical semiconductor optical amplifier configuration: lenses or lensed fibers couple the signal into/out of the SOA chip. A thermoelectric cooler (TEC) controls the operating temperature. (b) Photograph of a SOA chip, with the active waveguide visible, as well as the p-side metallization.

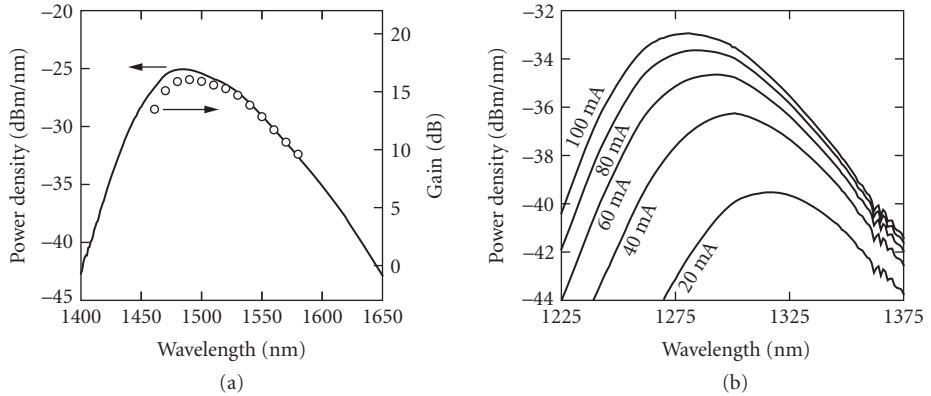


FIGURE 5 (a) Gain and ASE spectra of a SOA plotted to the same scale; only a vertical translation has been applied to match the curves. The mismatch toward the left conveys the larger noise figure of the device at shorter wavelengths. (b) ASE spectrum (of a different SOA) as a function of injected current. The gain near the band edge wavelength hardly changes, but higher current induces gain at shorter wavelengths.

Since the ASE spectrum represents spontaneous emission that has been amplified by the same gain that amplifies incoming signal, ASE spectrum and gain spectrum are strongly related. Figure 5 shows the two plotted together. Gain spectrum and ASE spectrum depend on injected current, with higher current filling more states higher in the bands, which extends the gain to shorter wavelengths.

The gain of a SOA depends strongly on temperature. This is the reason why a thermoelectric cooler (TEC) is applied to keep the chip at a nominal operating temperature, often 20 or 25°C (see Fig. 4). At high temperature, free carriers higher in the bands can be ejected out of the potential well formed by the double heterostructure without recombining radiatively (Fig. 2), which has the same effect as lowering the injection current. Figure 6 shows the effect on gain and peak wavelength of varying the chip temperature.

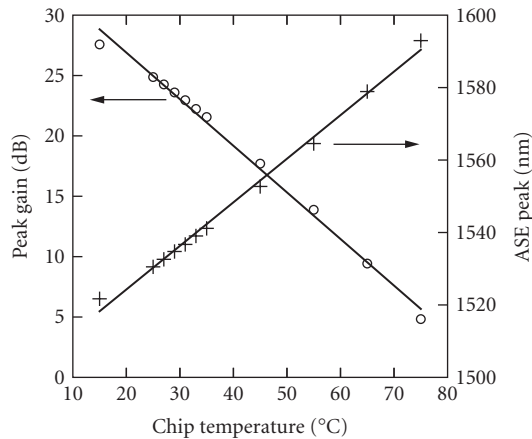


FIGURE 6 Variation of the gain and the ASE peak wavelength with chip temperature. Typical coefficients of $-0.4 \text{ dB}/^\circ\text{C}$ and $1 \text{ nm}/^\circ\text{C}$ are found in a SOA fabricated in the InGaAsP/InP material system with a bulk active layer.

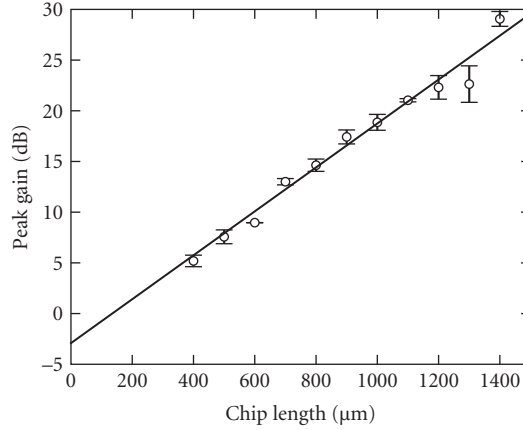


FIGURE 7 Gain versus chip length, for equal chip design and injection current density. The slope of the fit shows that, for this particular chip design and injection current, the on-chip gain is 2.2 dB per 100 μm , and at the 0 μm mark we find the loss caused by fiber-chip coupling, which in this case is about 1.5 dB per side.

In Figs. 5 and 6, gain has been plotted as a fiber-to-fiber number. The on-chip gain is a more fundamental quantity, but it cannot be readily measured without knowledge of the optical loss incurred by coupling a signal from the fiber into the chip and vice versa. Figure 7 shows fiber-to-fiber gain for many chips of different length but otherwise equal design, which allows us to deduce on-chip gain per unit length, as well as obtain an estimate for the fiber-chip coupling loss.

Confinement Factor

The gain per unit length as mentioned in the preceding paragraph is related to the material gain of the active region through the *confinement factor* Γ . This represents the fraction of the total power propagating in the waveguide that is confined to the active region, that is, it can be written as

$$\Gamma = \frac{\text{power in active region}}{\text{total power}} \quad (2)$$

The confinement factor thus relates the waveguide gain to the material gain as $g_{\text{wg}} = \Gamma g_{\text{mat}}$. In the example shown in Fig. 7, the confinement factor is $\Gamma = 0.1$. Therefore the waveguide gain of 22 dB/mm implies a material gain of 220 dB/mm.

The relation between material gain and the free carrier density n can be written in first order approximation as

$$g_{\text{mat}} = g_0(n - n_0) \quad (3)$$

where $g_0 = dg/dn$ is the differential gain, and n_0 is the free carrier density needed for material transparency, that is, the number of free carriers for which the rate of absorption just equals the rate of stimulated emission. g_0 and n_0 can vary with wavelength and temperature. The waveguide loss α varies only very weakly with wavelength.

Polarization Dependence

A SOA does not always amplify light in different polarization states by the same amount. The reason for this is a dependence on the polarization state of the confinement factor, which in turn is caused by the different waveguide boundary conditions for the two polarization directions giving rise to different mode profiles.

The principal polarization states of a planar waveguide are those in which the light is linearly polarized in the horizontal and vertical direction. The waveguide mode in which the electric field vector is predominantly in the plane of the substrate of the device is called the transverse electric (TE) polarization, while the mode in which the electric field is normal to the substrate is called transverse magnetic (TM).

In isotropic active material, such as the zincblende structure for common III-V crystals, the material gain is independent of the polarization direction, and is the same for the TE and TM modes. However, the rectangular shape of the waveguide, which is usually much more wide than it is high (typically around $2\text{ }\mu\text{m}$ wide while only around 100 nm thick), causes the confinement factor to be smaller for the TM mode than for the TE mode, sometimes by 50 percent or more. Without any mitigating measures, this would result in a significant polarization-dependent gain (PDG) (see Fig. 8).

Several methods exist with which polarization-independent gain can be achieved. The most straightforward one is to use a square active waveguide. This ensures symmetry between the TE and TM modes, and thus equal gain for both.⁶ This approach is not very practical, though, because the waveguide dimensions have to be kept very small (around $0.5 \times 0.5\text{ }\mu\text{m}$) to keep it from becoming multimode, and even though thin layers can be produced in crystal growth with high accuracy, the same accuracy is not available in the lithographic processes that define the waveguide width.

Another method is to introduce a material anisotropy that causes the material gain to become more favorable for the TM polarization direction in the exact amount needed to compensate for the waveguide anisotropy. This can be done by introducing tensile crystal strain in the active layer during material growth, a fact that was first discovered when lasers based on tensile-strained quantum wells were found to emit in the TM polarization direction.⁷

Introducing tensile strain in bulk active material modifies the shape of the light-hole and the heavy-hole bands comprising the valence band of the semiconductor in such a way that TE gain is somewhat reduced, while TM gain stays more or less constant. As a result, with increasing strain the PDG is reduced, and it can even overshoot, yielding devices that exhibit a TM gain higher than their TE gain. Appropriately optimized, this method can yield devices with a PDG close to 0 dB.⁸

In a multiquantum well (MQW) active layer, the same method can be used by introducing tensile strain into the quantum wells.⁹ Alternatively, a stack of QWs alternating between tensile and compressive strain can be used. The compressive wells amplify only TE, while the tensile wells predominantly amplify TM. This way, the gain of TE and TM can be separately optimized, and low-PDG structures can be obtained.^{10,11}

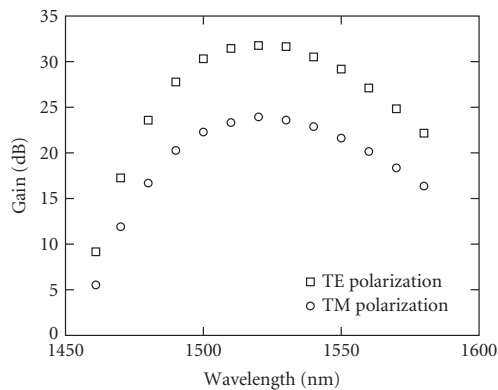


FIGURE 8 Gain versus wavelength in a SOA with significantly undercompensated polarization dependence.

Gain Ripple and Feedback Reduction

Reflections from the chip facets can cause resonant or antiresonant amplification, depending on whether a whole number of wavelengths fit in the cavity. This behavior shows up in the ASE spectrum, as shown in Fig. 3*b*. The depth of this gain ripple is given by⁴

$$\text{Ripple} = \frac{(1 + gr)^2}{(1 - gr)^2} \quad (4)$$

in which g is the on-chip gain experienced by the guided mode, and r is the facet reflectivity. Obviously, gain ripple becomes more of a problem for high-gain SOAs. A device with an on-chip gain of 30 dB will need the facet reflectivities to be suppressed to as low as 5×10^{-6} in order to show less than 0.1 dB ripple.

Several methods are used to suppress facet reflections (see Fig. 9); the most well-known one is to apply antireflection (AR) coatings onto the facets. An AR coating is a dielectric layer or stack of layers that is designed such that destructive interference occurs among the reflections of all its interfaces.

For a planar wave, a quarter-wave layer with a refractive index that is the geometric mean of the indices of the two regions it separates is a perfect AR coating. Such a design is also a reasonable first approximation for guided waves, but for ultralow reflectivity, careful optimization to the actual mode field needs to be done, taking account of the fact that the optimum for the TE and TM modes may be different.

Since any light that is reflected at an interface is not transmitted through it, AR coatings also help lowering the coupling loss to fiber. The approximately 30 percent reflection of an InP-air interface in absence of a coating would represent a loss of 1.5 dB.

A second method is to angle the SOA waveguide on the chip. Such an *angled stripe* design makes the reflected field propagate backward at double the angle with respect to the waveguide, rather than being reflected directly into the waveguide, and therefore only a small fraction will couple back into the guided mode. Note that a consequence of using an angled stripe is that the output light will emanate from the SOA at a larger angle according to Snell's law, and an appropriate angle of the fiber assemblies will have to be provisioned, which can be as large as 35° for an on-chip angle of 10° .

Another way to reduce reflections back into the waveguide is to end the waveguide a few micrometers before the facet and continue with only cladding material; this is often done in combination with a taper that enhances mode matching to the fiber. This so-called *window structure* allows the modal field to diverge before it hits the facet, so that the reflected field couples poorly back into the waveguide.

For low-gain SOAs, applying only an AR coating suffices. It has also been shown that only an angled stripe (without AR coating) can be sufficient.¹² But when ultralow reflectivity is needed, it is common to find an AR coating being combined with an angled stripe, or a window region, or both. Reflectivities as low as 2×10^{-6} have been obtained in this way.¹³

It has to be noted that mode-matching techniques to enhance fiber-chip coupling efficiency, such as lateral tapers, have an effect on reflections. A larger (usually better-matched) mode diffracts at a smaller angle, improving the effectiveness of an angled stripe, but reducing the effectiveness of a window region.

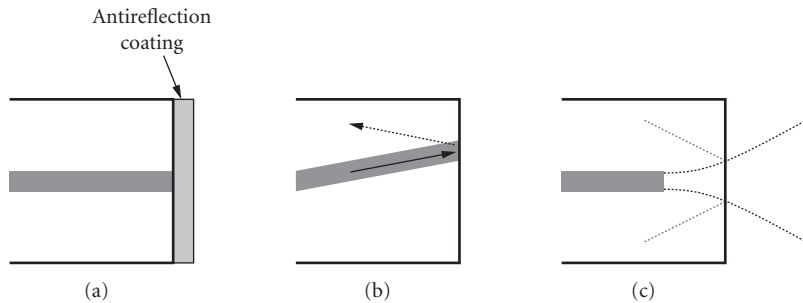


FIGURE 9 Suppressing facet reflections: (a) antireflection coating; (b) angled stripe; and (c) window structure.

Noise Figure

An optical amplifier's noise figure depends on its inversion factor as $nf = 2n_{sp}/\eta_i$, with n_{sp} the inversion factor (equal to one for full carrier inversion) and η_i the optical transmission ($1 - \text{loss}$) at the input of the amplifier. An ideal fully inverted amplifier with zero input loss ($\eta_i = 1$) would have a $nf = 2$ (or, expressed in decibels, $NF = 3$ dB). In a SOA, η_i consists of the fiber-chip coupling coefficient (see Fig. 4a), which can be significantly different from unity. This is the reason why NF is usually somewhat higher for SOAs compared to fiber amplifiers.

The conventional interpretation of the noise figure is that it is the signal-to-noise ratio at an amplifier's output divided by that at its input, for a shot-noise-limited input signal. For optical amplifiers this definition is not very practical, since signals in optical networks are seldom shot-noise-limited.

A more practical definition is based on the approximation that in an optically amplified, optically filtered transmission line, the noise in the receiver is dominated by signal-spontaneous emission beat noise.¹⁴ This results in a definition of noise figure as

$$nf = \frac{2\rho_{ASE||}}{gh\nu} \quad (5)$$

in which g is the gain, $\rho_{ASE||}$ is the power spectral density of the amplifier's ASE noise, and $h\nu$ is the photon energy. Only the noise power copolarized with the signal is taken into account, since noise with a polarization orthogonal to that of the signal does not give rise to beat noise in the detector. All quantities in this expression can be easily measured, which allows for straightforward characterization of a device's noise figure (see "Noise Figure" in Sec. 19.4).

Saturation

In an amplifier, the gain depends on the amplified signal power, which at high values causes the output power to saturate. A strong input signal causes the stimulated emission to reduce the carrier density, which decreases the gain and at the same time shifts the gain peak to longer wavelengths, closer to the band gap emission wavelength of the active stripe. This gain compression can be written as a function of the output power p_o as follows:¹⁵

$$g = g_{ss} e^{-p_o/p_{sat}} \quad (6)$$

with g_{ss} the small-signal gain (assumed to be $\gg 1$), and

$$p_{sat} = \frac{h\nu A \eta_o}{\tau \Gamma dg/dn} \quad (7)$$

the characteristic saturation output power, which depends on the carrier lifetime τ , the confinement factor Γ , the differential gain dg/dn , the cross-section area A of the active stripe, and the output coupling efficiency η_o .

A convenient description of the saturation power of an optical amplifier is given by the output power at which the gain is reduced by a factor of two, or 3 dB. This so-called 3-dB saturation output power can now simply be written as

$$p_{3dB} = \ln 2 \cdot p_{sat} \quad (8)$$

Figure 10 shows an example of a measured gain versus output power curve, from which the small-signal gain of the amplifier and the 3-dB saturation power can be directly determined.

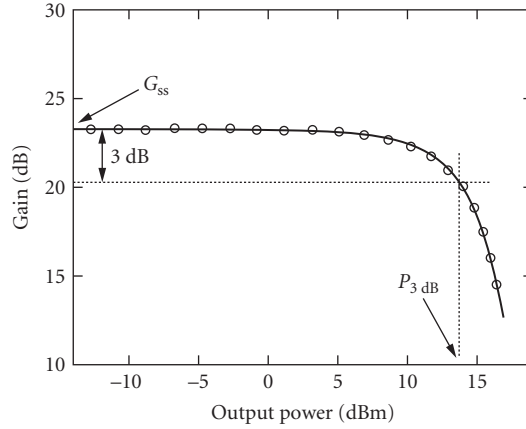


FIGURE 10 Gain compression curve of a SOA. At small powers, the gain approaches the value of the small-signal gain G_{ss} . The output power corresponding to a gain compression of a factor of two is the 3-dB saturation output power $P_{3\text{ dB}}$.

Increasing the injection current into the SOA increases the saturation power through reduction of the carrier lifetime τ and reduction of the differential gain dg/dn . An increase can also be accomplished by using a SOA with gain peak significantly shorter than the signal wavelength. Due to band filling, the differential gain is smaller on the red side of the gain peak, which causes the saturation power to be larger at longer wavelengths (see Fig. 11).

Structurally, the saturation power of a SOA can be increased by reducing the thickness of the active layer.¹⁶ The optical field expands widely in the vertical direction, which decreases the optical confinement factor Γ much faster than it decreases the active cross-section A .

In the horizontal direction, the field is usually much better confined. Therefore another effective way to increase P_{sat} is to use a flared gain stripe, that tapers to a much larger width at the output. This increases the active cross-section much faster than it does the confinement factor. Since the waveguide at the output will typically be multimode, care has to be taken in the design of the taper.

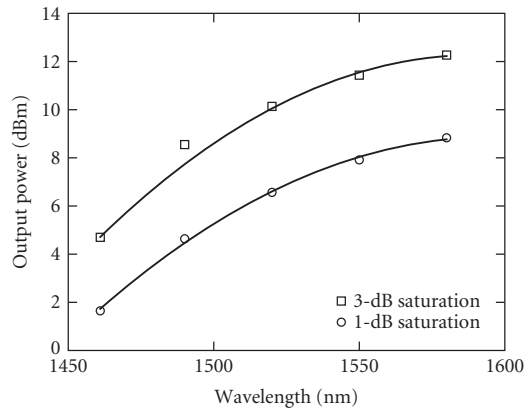


FIGURE 11 Saturation power of a SOA versus signal wavelength. The smaller differential gain at longer wavelengths causes an increase in P_{sat} .

Material Systems

Semiconductor optical amplifiers are most commonly fabricated in the InGaAsP/InP material system. The active and other waveguiding layers, as well as electrical contacting layers are epitaxially grown on an InP substrate. InGaAsP is chosen because it allows the emission wavelength to be chosen in the range 1250 to 1650 nm, which contains a number of bands that are important for telecommunications.

Since the gallium atom is slightly smaller than the indium atom, whereas the arsenic atom is slightly larger than the phosphorus atom, by choosing the element ratios In:Ga and As:P properly, a crystal lattice can be formed with the same lattice constant as InP. The remaining degree of freedom among these *lattice-matched* compositions is used to tune the band gap, which is direct over the full range from binary InP to ternary InGaAs, hence the ability to form 1250 to 1650 nm emitters. For emission at wavelengths such as 850 or 980 nm, the GaAs/AlGaAs material is commonly used.

A *quantum well* may be created by sandwiching a thin layer of material between two layers with wider band gap. This forms a potential well in which the free carriers may be confined, leaving them only the plane of the quantum-well layer to move freely (see Fig. 12). Even though quantum wells are used almost exclusively for the fabrication of semiconductor lasers, both quantum well and bulk

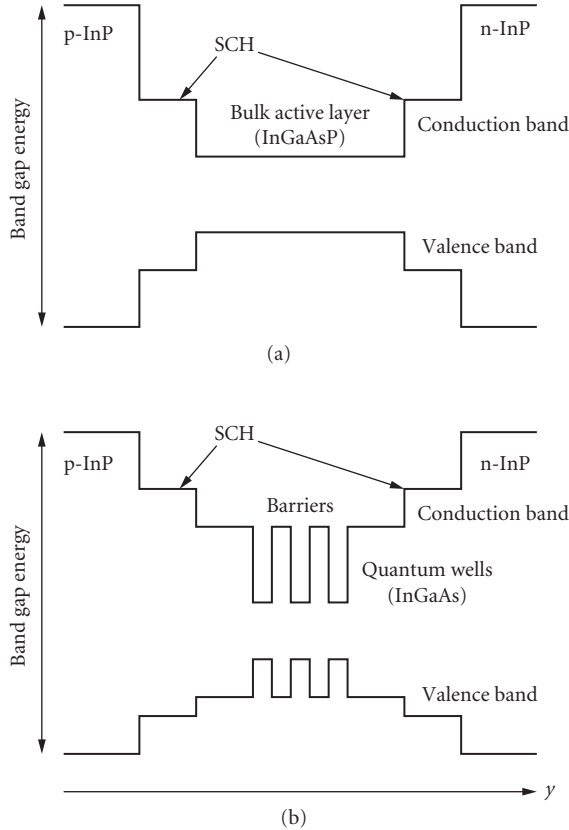


FIGURE 12 (a) Layer structure of a SOA waveguide with a bulk active layer and (b) structure of a MQW SOA. Note that the band offsets in the conduction and valence bands are not drawn to scale.

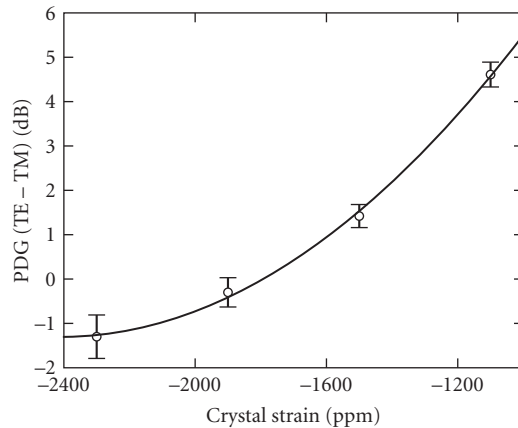


FIGURE 13 Polarization dependence in SOAs with bulk active layers with varying amounts of tensile strain.

active structures are used for SOAs, as the advantages of quantum wells are less pronounced in amplifiers than they are in lasers. Recently, active structures have been demonstrated based on *quantum dots*. These dots confine the electrons not in one, but in all three spatial directions, giving rise to delta function-like density of states. This property is expected to lead to devices with reduced temperature dependence with respect to bulk or quantum well devices.

As mentioned earlier, strain can be used in the active layer to tune the polarization-dependent gain of a SOA. Strain is introduced by deviating from the layer compositions that would yield the active layer lattice-matched, either by introducing a larger fraction of the larger elements, to produce compressive strain, or by emphasizing the smaller atoms, to produce tensile strain. Figure 13 shows the effect of strain on PDG in a bulk active SOA.

Gain Dynamics

The dynamic behavior of a SOA is governed by the time constants associated with the various processes its free carriers can undergo. The carrier lifetime τ has already been mentioned. This is the characteristic time associated with *interband* processes such as spontaneous emission and the electrical pumping of the active layer, that is, with the movement of electrons between the valence and the conduction band. The carrier lifetime is of the order of 25 to 250 ps.

Intraband processes such as spectral hole burning and carrier heating, on the other hand, govern the (re)distribution of carriers inside the semiconductor bands. These processes are much faster than the carrier lifetime.¹⁷

Dynamic effects can be a nuisance when one only wants to amplify modulated signals, because they introduce nonlinear behavior that leads to intersymbol interference. But they can be used advantageously in various forms of all-optical processing. Using a strong signal to influence the gain of the amplifier, one can affect the amplitude of other signals being amplified at the same time. An example of this *cross-gain modulation* (XGM) is shown in Fig. 14. In the gain-recovery measurement, the gain of the SOA is reduced almost instantaneously as the pump pulse sweeps the free carriers out of the active region. After the pulse has passed, the gain slowly recovers back to its original value.

The gain-recovery time depends on the design of the SOA and the injection current, as shown in Fig. 15. Cross-gain modulation can support all-optical processing for signals with data rates higher than 100 Gb/s.

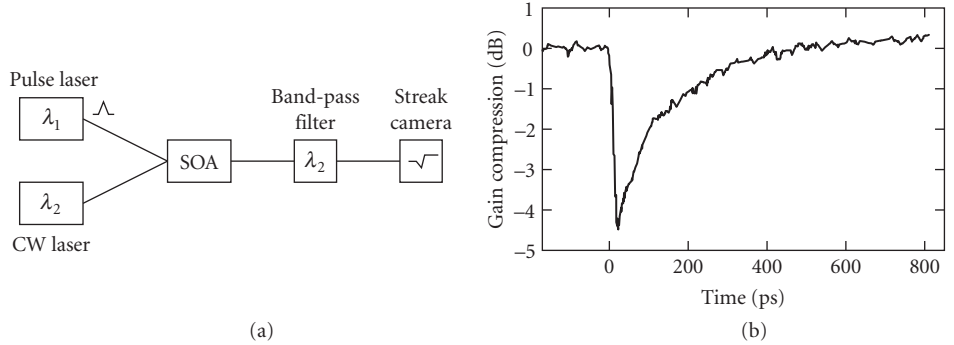


FIGURE 14 (a) Gain recovery experiment in which an intense pulse (the pump) compresses the gain of a SOA, which is measured by a weak probe beam and (b) gain compression and recovery at λ_2 .

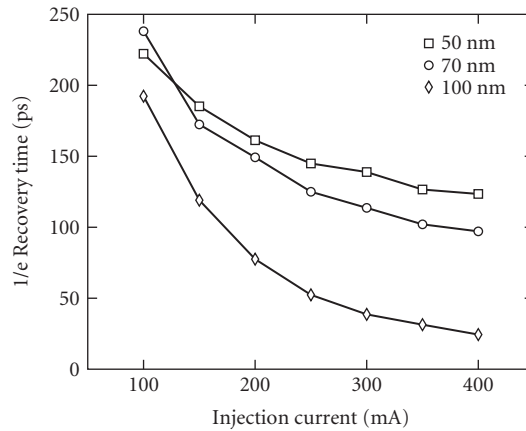


FIGURE 15 Cross-gain modulation recovery times versus active-layer thickness and current injection. The SOAs have a bulk active layer and a gain peak at 1550 nm. Chip length is 1 mm.

Along with the gain change caused by a strong input signal, there is a phase change associated with the refractive index difference caused by the removal of free carriers, which results in heavy chirping of signals optically modulated by the XGM. However, this *cross-phase modulation* (XPM) can also be used to advantage. Only a small gain change is needed to obtain a π phase shift, so all-optical phase modulation can be obtained without adding much amplitude modulation. Using a waveguide interferometer, this phase modulation can be converted back to on-off keying.

Intraband processes give rise to effects like four-wave mixing (FWM). This is an interaction between wavelengths injected into a SOA that creates photons at different wavelengths (see Fig. 16). A straightforward way to understand FWM is as follows. Two injected pump beams create a moving beat pattern of intensity hills and valleys, which interacts with the SOA nonlinearities to set up a moving grating of minima and maxima in refractive index. Photons in either beam can be scattered by that moving grating, creating beams at lower or higher frequency, spaced by the frequency difference between the two pump beams.

More detail on applications of the nonlinearities will be described in Sec. 19.8.

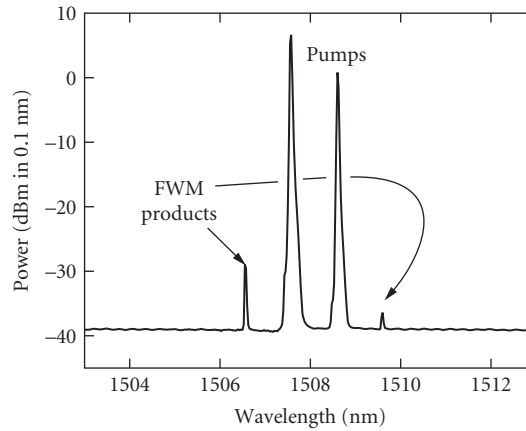


FIGURE 16 Four-wave mixing in a SOA. The two center pump beams give rise to mixing products on both sides.

Gain Clamping

One possible solution to limit intersymbol interference caused by SOA nonlinearities is to resort to gain clamping. When controlled lasing is introduced in a SOA, the gain is clamped by virtue of the lasing condition, and no gain variations are caused by modulated input signals.

Lasing can be introduced in a SOA by etching short gratings at both ends of the active waveguide.^{18,19} The feedback wavelength is set to fall outside the wavelength band of interest for amplification, and the grating strength defines the round-trip loss, and therefore the level at which the gain is clamped (see Fig. 17). Now when an input signal is introduced into the gain-clamped SOA (GC-SOA), the gain will not change as long as the device is lasing. As the amplified input signal takes up more power, less carriers are available to support the lasing action. Only when so many carriers are used by the signal to make the laser go below threshold, will gain start to drop.

The steady-state picture sketched needs to be augmented to account for the dynamic behavior of the clamping laser. Relaxation oscillations limit the effectiveness of gain clamping. In order to support amplification of 10-Gb/s NRZ on-off keying modulated signals, the GC-SOA is designed with its relaxation oscillation peak at 10 GHz, where the modulation spectrum has a null.

A different type of GC-SOA has its clamping laser operating vertically. This device, called linear optical amplifier (LOA),²⁰ has a vertical cavity surface emitting laser (VCSEL) integrated along the full length of the active stripe. This design has the advantage that the clamping laser line is not present in the amplifier output, as it emits orthogonally to the propagation direction of the amplified signals.

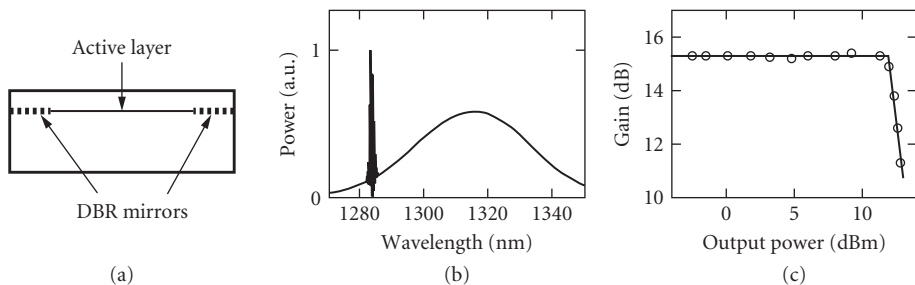


FIGURE 17 Gain-clamped SOA from Ref. 19: (a) schematic; (b) output spectrum; and (c) gain versus output power curve.

The gain clamping laser in this case has a relaxation oscillation that varies over the length of the device. For this reason, no hard relaxation oscillation peak is observed. At the same time, the large spontaneous emission factor of the vertical laser makes the clamping level less well-defined: rather than staying absolutely constant up to the point of going below threshold, it causes a softer knee in the gain versus output power curve, somewhat in between the horizontal GC-SOA case (Fig. 17c) and the case of an unclamped SOA (Fig. 10).

In practice, the “sweet spot” for amplification of digitally modulated signals is at an output power corresponding to around 1 to 3 dB of gain compression,²¹ depending on data rate and modulation format. At this point in the gain versus output power curve, the LOA has no output power advantage over standard, unclamped SOAs. For analog transmission, the gain has to stay absolutely constant; this has only been attempted with horizontally clamped GC-SOAs.²²

19.3 FABRICATION

The fabrication of SOAs is a wafer-scale process that is very similar to the manufacturing of semiconductor laser diodes. First, epitaxial layers are grown on a semiconductor substrate. Then, waveguides are formed by etching, followed by one or more optional regrowth steps. Finally, p- and n-side metalization is applied.

Waveguide Processing

InGaAsP/InP SOAs, like their laser counterparts, mostly use buried waveguide structures that are fabricated in a standard buried heterostructure (BH) process: A mesa is etched in the epitaxial layer stack containing the gain stripe using a dielectric mask. Using this mask, selective regrowth is performed in the regions next to the waveguide, in order to form current blocking layers that force all injection current to flow through the active stripe. This is accomplished either using semi-insulating material, for example, Fe:InP, or with a p-n structure that forms a reverse-biased diode. After removing the etching and regrowth mask, a p-doped InP top layer is grown to provide for the p-contact (see Fig. 18).

GaAs/AlGaAs structures are not easily overgrown. In this material system, ridge waveguides are usually used. Usually the epi layers are grown in a single growth step with a thick p-doped top layer already in place. Waveguides are formed by etching, after which the whole structure is covered with a passivation layer, in which contact openings are etched on top of the ridge. (See Fig. 19 for an example of a ridge structure grown on InP.)

The waveguide pattern on a SOA wafer usually includes angled stripes to reduce feedback into the waveguide from facet reflections. For the same reason, a buried waveguide may contain window structures, in which the waveguide ends a few micrometers before the facet, the remaining distance

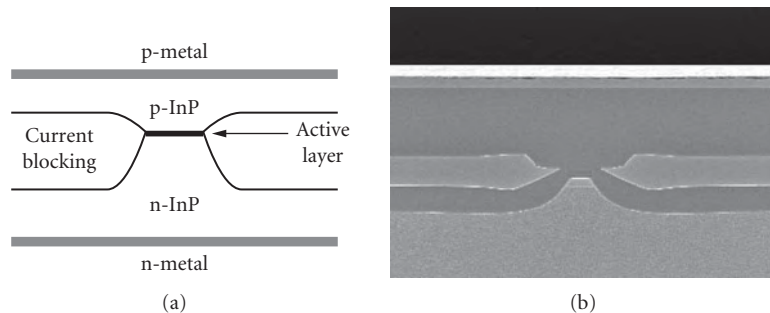


FIGURE 18 (a) Schematic of planar buried heterostructure and (b) SEM photograph of BH waveguide.

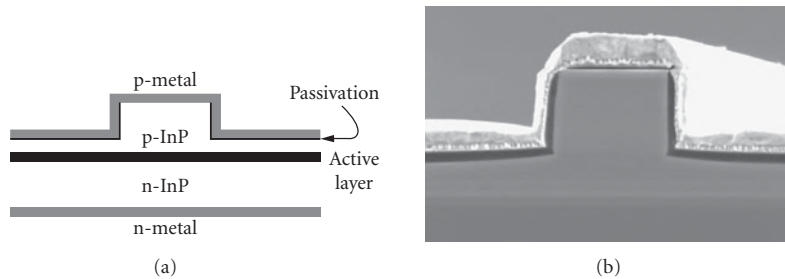


FIGURE 19 (a) Schematic of ridge waveguide structure and (b) SEM photograph of ridge waveguide.

being bridged by nonguided propagation in the regrown InP. The waveguides may contain tapers near the facets to shape the mode for improved fiber-chip coupling efficiency, and may be flared as described earlier to obtain higher saturation power.

Metallization

After waveguide processing, p-side metal is applied on top of the wafer, which is patterned to form contact pads. Before n-side metal is applied to the back side of the wafer, it is usually thinned in a lapping and polishing process to improve cleaving yield, as well as thermal conductivity between the active layer and the heatsink on which the devices will be mounted. Figure 20 shows a photograph of a finished SOA chip on submount.

Postprocessing and Package Assembly

A fully processed 2-in wafer can contain thousands of individual SOA devices. Before these can be used, a facet coating needs to be applied for passivation and to reduce reflections. To this end, the wafer is cleaved into bars containing many SOAs, which are subsequently antireflection coated. An AR coating is nominally a quarter-wave layer of dielectric material, which reduces reflections from the active waveguide back into the chip. In combination with an angled stripe and possibly a window region, both of

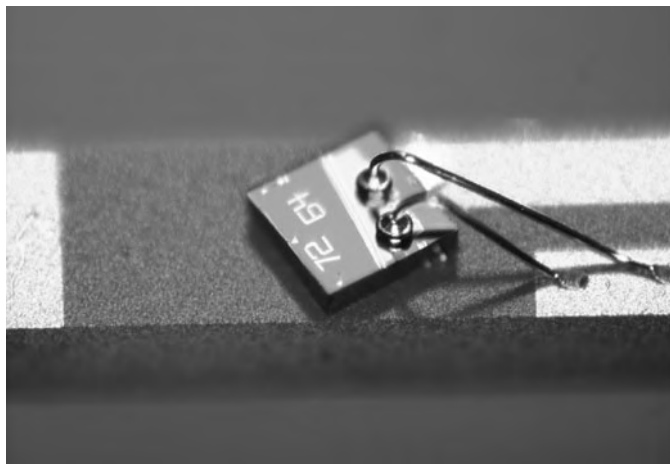


FIGURE 20 SOA chip-mounted on a heatsink.

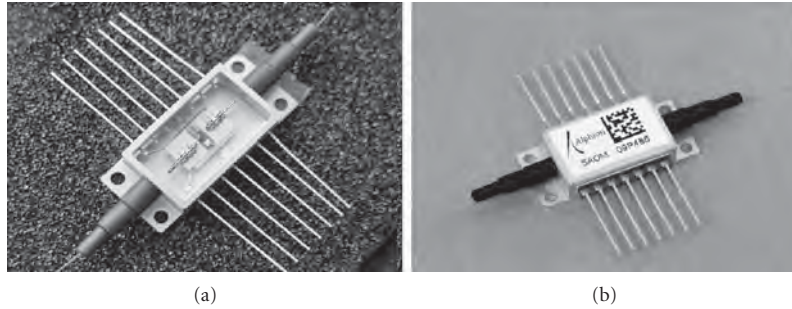


FIGURE 21 (a) Fiber-coupled chip and (b) SOA in industry-standard butterfly package.

which reduce the fraction of backreflected light that is coupled back into the waveguide, the effective facet reflectivity may be lower than 10^{-5} .¹³ Finally, the coated bars are diced into individual chips.

Most modern SOAs come packaged in a 14-pin industry standard butterfly package (see Fig 21). To increase fiber-chip coupling efficiency, either lensed fibers are used, or separate lenses are aligned between chip and fibers. In packaging schemes with multiple lenses per side, a collimated free-space section can be provided that allows for inclusion of polarization-independent isolators inside the package.

19.4 DEVICE CHARACTERIZATION

Chip Screening

The initial characterization of a SOA chip comprises the diode characteristic and the ASE light output. This is accomplished using a L-I-V measurement, which measures the light output versus current (L-I) and the I-V curve at the same time.

The measurement procedure is identical to the L-I-V measurement that is used to determine the threshold of a laser diode, the main differences being that no threshold is measured for properly antireflection coated SOA chips, and facet emission is monitored on both sides. Figure 22 shows an example measurement: A current sweep is applied to the chip, and the voltage across the chip is measured, while facet emission is monitored using broad-area photodiodes.

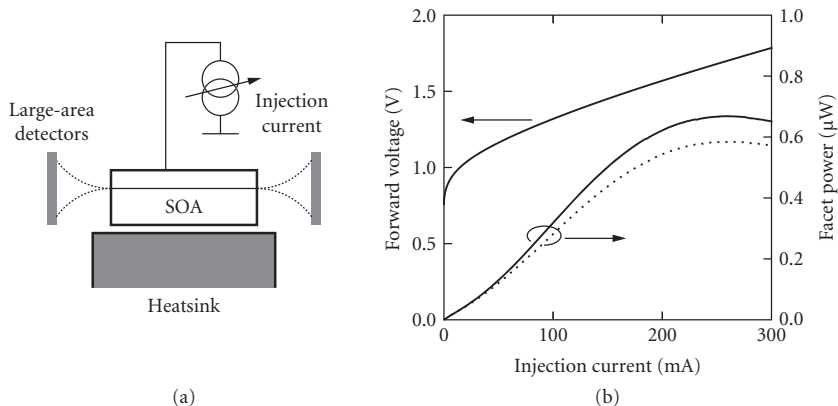


FIGURE 22 (a) Schematic of L-I-V measurement setup and (b) example measurement.

Gain Measurement

Measuring device gain requires coupling an input signal into the chip, and measuring output power divided by input power. This is easily done in a packaged device. Measuring an unpackaged chip requires aligning lenses or lensed fibers in front of the facets using nanopositioning stages.

In small signal gain measurements, the ASE noise the amplifier produces will typically overwhelm the output signal power. Therefore, a filter has to be used in the output. Depending on filter width, signal power, and accuracy required, it might even be necessary to measure noise level and subtract out the noise power transmitted through the filter, to avoid overestimating the gain.

A gain versus output power measurement like Fig. 10 is easily accomplished by stepping the input power. From Eq. (6) it appears that small signal gain and 3-dB saturation power can be easily extracted by a linear fit to

$$G = G_{ss} - 3p_o/p_{3\text{dB}} \quad (9)$$

where the measured gain values G are expressed in decibels and the measured output powers p_o are expressed in milliwatts. The fit parameter G_{ss} is the small-signal gain and $p_{3\text{dB}}$ is the 3-dB saturation output power as defined before. However, this expression assumes that the differential gain and the carrier lifetime remain constant upon saturation, which is not the case in practice. A second-order fit is necessary to accurately characterize the saturation behavior of the amplifier, in which case the 1-dB and 3-dB saturation powers can be used as figures of merit.¹³

Polarization-Dependent Gain Measurement

Signals of controlled polarization need to be applied to the SOA in order to measure its PDG. If a device has been pigtailed using polarization-maintaining fiber, all that is needed is to connect a horizontally polarized signal and then a vertically polarized signal. However, standard single-mode fiber does not preserve the polarization state, so more complicated measurement techniques are required in most cases.

Methods and systems available to measure polarization-dependent loss (PDL) in passive devices are usually applicable to PDG measurements, with one caveat: the device under test generates ASE noise, which has to be filtered out prior to the detector stage of the system. Because one is usually interested in the absolute value of the gain in addition to the PDG, the insertion loss of the filter needs to be well calibrated.

An often used method is polarization scrambling. A fast polarization scrambler scans through many different polarization states, evenly covering the Poincaré sphere. By measuring the output power of the SOA using a fast power meter with a min/max hold function, the highest and lowest gain over all polarizations can be found, and thus the PDG. Note that if an optical spectrum analyzer (OSA) is used as the power meter (in which case the detector is the filter), often a min/max function is not available, and having to coordinate the scans of the polarization scrambler and the OSA makes this method of PDG measurement inconvenient.

The *Mueller matrix method*²³ is an alternative that does not suffer from this problem. In this case, a polarization controller is needed that can synthesize any desired polarization state in the input signal, although a limited variant of the method works with a polarization controller that can access only the principal states on the Poincaré sphere: horizontal, vertical, 45°, -45°, right-hand circular, left-hand circular. Note that these states are defined as launched in the fiber directly following the polarization controller. Due to the nonpolarization-maintaining nature of the fiber, the states launched into the SOA chip are unknown, although their relative properties are maintained, that is, orthogonal states remain orthogonal, etc.

The essence of the method is to first measure the gain corresponding to an unpolarized signal by averaging the gain of two orthogonally polarized states, for example horizontal and vertical.

Next, gain is measured for three states that are orthogonal to each other *on the Poincare sphere*, for example, horizontal, 45° , and right-hand circular. The differences between these three gain values on the one hand, and the average gain measured in the first step on the other hand, contain the information needed to determine the polarization transformation the light has undergone in the fiber between the polarization controller and the chip. The gains in the maximum and minimum polarization states can now be calculated, or alternatively those two polarization states can be synthesized in the polarization controller and the corresponding gains measured directly.

In practice, therefore, the Mueller method allows for rapid measurement of both gain and PDG of a SOA by measuring output signal powers for four predetermined polarization states. Note that both this method and the polarization scrambling method only yield the highest gain and lowest gain over polarization; not information on which of these gains belongs to TE and TM polarization in the waveguide, respectively. In other words, the methods yield the absolute value of the PDG but not its sign. The Mueller method, but not the scrambling method, can yield the sign of the PDG by comparing the calculated polarization transformations for a device under test (DUT) relative to a device with a large PDG of which the sign is known, for example, a device with unstrained or compressively strained active layer. The fiber leading up to the facet needs to remain relatively undisturbed when replacing the DUT with the reference chip, in order to preserve the polarization transformation.

Gain Ripple

Gain ripple can be measured either directly, by sampling gain at closely spaced wavelength points, or by proxy by measuring ASE ripple. If the gain ripple period is known (e.g., by deriving it from the length of the chip), the number of gain measurements can be greatly reduced. Since the ripple is caused by a Fabry-Pérot cavity formed by the gain stripe and the two facets, the curve of reciprocal gain ($1/g$) versus wavelength is a sinusoid. Measuring three points on this sinusoid separated by one-third of the gain ripple period, the average and standard deviation are invariant for translation over wavelength. The inverted average now corresponds to the single pass gain, while the standard deviation can be worked into the ripple amplitude¹³ (see Fig. 23a).

Alternatively, the amplitude of the ripple on the ASE spectrum can be found by measuring a small wavelength span with sufficiently narrow resolution bandwidth (see Fig. 23b.)

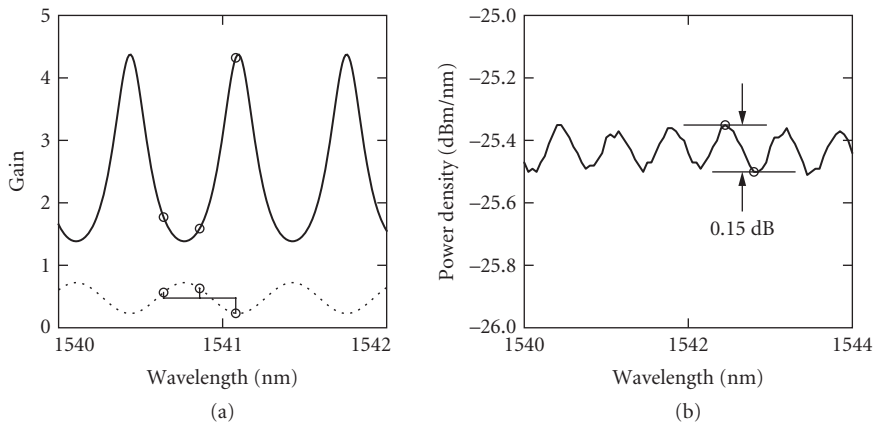


FIGURE 23 (a) Measurement of gain ripple and (b) measurement of ASE ripple.

Noise Figure

If the noise level has been measured during a gain measurement, all ingredients are present to calculate noise figure. If $G = P_o - P_i$ is converted to linear units and ρ_{ASE} is given in watts per hertz, noise figure in linear units is given by

$$\text{nf} = \frac{2(\rho_{\text{ASE}}/2)}{gh\nu} + \frac{1}{g} \quad (10)$$

The rightmost term was neglected in Eq. (5); it is only of importance for low-gain devices. The factor 1/2 expresses the assumption that half the total ASE power is copolarized with the signal. This is a good approximation for a polarization-insensitive SOA. For devices with considerable polarization dependence, the value of the PDG can be used to partition the total ASE power into two polarization components. Combined with the corresponding gain, the above expression will then yield two equal values for the NF. An actual measurement of the NF associated with each polarization state requires measurement of the output signal and ASE power through a polarization analyzer.

Complete Characterization

Figure 24 schematically shows a measurement system that allows full steady-state characterization of a SOA that includes all of the above measurements. The input source is a tunable laser to allow for measurement of gain versus wavelength. A variable attenuator and a polarization controller enable measurement of a saturation curve and of the PDG. Input power is referenced to a power meter, while output power and noise levels are measured on an optical spectrum analyzer.

An additional pair of power meters is provided on both sides for measuring the total ASE power. Comparing the ASE powers thus measured with direct power measurement of the DUT fibers with a reference power meter allows calibration of the loss of the connectors indicated in Fig. 24. This ensures a true fiber-to-fiber measurement. For chip measurements, the two extra power meters are helpful during optimization of the fiber position in front of the facets.

An example of a complete dataset measured for a SOA is shown in Fig. 25.

High-Performance SOA Properties

The gain of a SOA is only limited by the quality of the antireflection coatings, and by the rising ASE noise power which may cause the amplifier to autosaturate. Devices with fiber-to-fiber gain as high as 36 dB (at $\lambda = 1310$ nm, using a strained-layer MQW active layer) have been reported in the literature.¹³ High gain is not hard to obtain, by extending the device length and proportionally increasing the injection current. The difficult parameters to optimize are usually PDG, saturation output power, and noise figure. In this section, we will cite some state of the art results; unless stated otherwise, numbers quoted are fiber-coupled values.

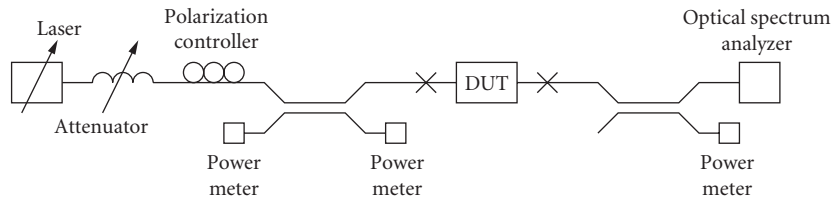


FIGURE 24 Measurement system to characterize gain versus wavelength, polarization, and input power.

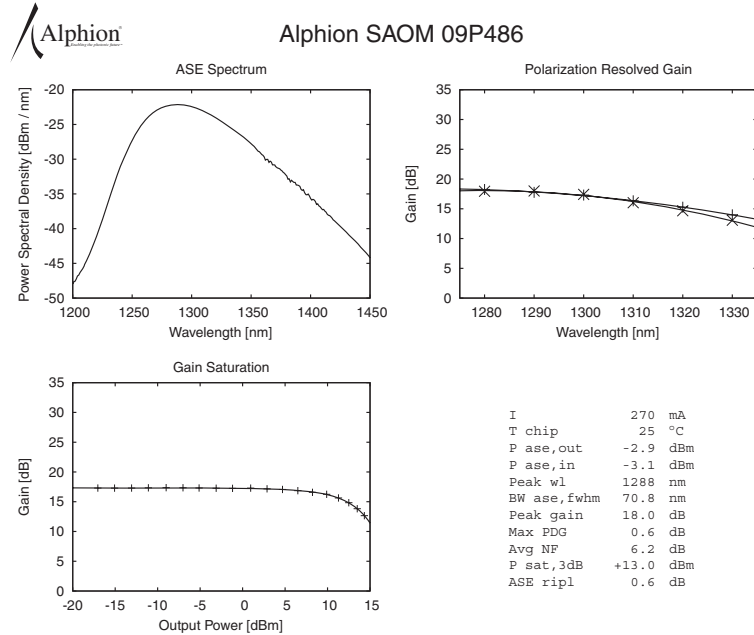


FIGURE 25 Example of a datasheet as delivered with a commercial SOA, showing the complete characterization results of the device.

A tensile strained bulk active device with a saturation output power of 17 dBm at an injection current of $I = 500$ mA was reported, having a gain of 19 dB at $\lambda = 1550$ nm.²⁴ Polarization dependence was 0.2 dB thanks to the tensile strain, and NF was 7 dB. The high saturation power was reached mainly thanks to the thin active layer (50 nm). At 600 mA injection current, a similar device, implemented using a MQW structure with tensile strained barriers and unstrained wells, reached a saturation power of 20 dBm.²⁵ Gain was 11 dB and PDG was 0.6 dB at 1550 nm, and NF was 6 dB.

A compressively strained MQW structure reached a chip output power of 24 dBm at 1590 nm.²⁶ This is a single-polarization device, having a peak chip gain of 18 dB. The 1.8 mm long structure was injected with a current of 1 A. The minimum chip NF was 3.6 dB, only 0.6 dB above the theoretical minimum, thanks to low cladding absorption. This chip was built into a polarization diversity module, in which a gain of 15 dB with a PDG of 0.5 dB was obtained. $P_{3\text{ dB}}$ and NF were 22 dBm and 5.7 dB, respectively.²⁷

A very large saturation output power of 29 dBm has been obtained in a slab-coupled optical waveguide amplifier (SCOWA), thanks to an ultralow confinement factor ($\Gamma < 0.005$).²⁸ At an injection current of 5 A, the 10 mm long device exhibits a gain of 13.8 dB. The single-polarization device has a NF of 5.5 dB.²⁹

The above results have been obtained with MQW or bulk active layers. Quantum dot (QD) active layers show promise because of their ultrafast gain dynamics (a few picoseconds) and ultrawide gain bandwidth. A high-performance device has been demonstrated with a gain of 25 dB, a NF of 5 dB, and a $P_{3\text{ dB}}$ of 19 dBm, all over a bandwidth of 1410 to 1500 nm.³⁰ The 6-mm long device was pumped with a current of 2.5 A. These are chip-referenced numbers, and the device amplifies one polarization only. Recently, a polarization-independent QD device was demonstrated, which uses dots arranged in columns, surrounded by tensile strained barriers, to obtain a PDG of 0.5 dB.³¹ At a length of 6 mm and an injection current of 1.2 A, a fiber-to-fiber gain of 4 dB and a saturation power of 16.5 dBm were obtained. The chip NF of 9.5 dB was relatively high due to the low gain.

19.5 APPLICATIONS

Applications for SOAs make use of the gain provided by the device, as well as the high-speed internal dynamics, which can be used for various optical signal-processing applications. The gain and index dynamics within the SOA must be managed to enhance the desired application. In a first set of applications, the SOA is used for linear amplification. Here, the operating regime of the device is managed to reduce nonlinear effects to acceptable levels, which benefits from devices with large saturation powers. For nonlinear applications, operating conditions are chosen to enhance the nonlinear gain and/or index nonlinearities. Both the SOA design and its configuration within a system are chosen to produce the desired nonlinear functionality. For nonlinear applications, smaller values of saturation power can be a benefit. Also, incorporation of the SOA into a subsystem with other optical elements, such as filters, interferometers, and the like, is used to create the nonlinear functional devices.

In the following sections, we characterize the applications of SOAs under three main headings: amplification of signals, switching and modulation, and nonlinear applications.

19.6 AMPLIFICATION OF SIGNALS

The function of amplification in a transmission system is performed in several places, as shown in Fig. 26, in which the amplifier can be a power amplifier, an in-line amplifier, or a preamplifier. As shown, the transmission system is divided into three parts: the transmitter, the transmission line, and the receiver. The power amplifier is used in the transmitter, the in-line amplifier is used in the transmission line, and the preamplifier is used in the receiver.

Single-Channel Systems

The power amplifier is located in the transmitter to boost the optical signal level before it enters the transmission system. (In Fig. 26, two possible locations are shown before or after the modulator.) In most cases, the power amplifier is placed before the signal modulator, so that the optical signal is either continuous wave (CW) or a pulse train. In this case, the properties of concern for the SOA are its output power and broadband ASE. The signal injected into the SOA is relatively large, so if the broadband ASE is filtered to allow only a narrow band around the signal channel, noise degradation by the SOA is minimal in the power amplifier. Because the polarization of the system is controlled at this location, polarization sensitivity of the amplifier is not of concern. Because the signal is either CW or a train of equally spaced pulses, the gain recovery time will not influence the temporal properties of the amplified signal. The SOA can be operated in saturation to maximize the output power.

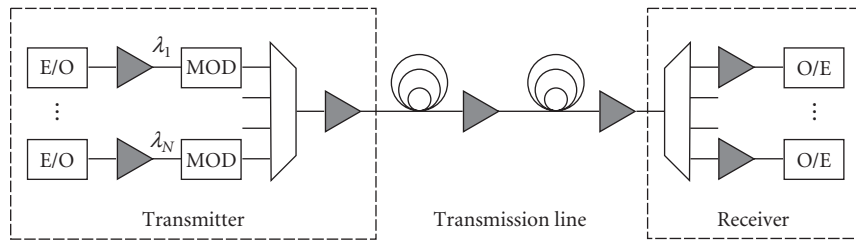


FIGURE 26 Placement of amplifiers in a WDM optical transmission system. The system is divided into transmitter, receiver, and transmission line blocks. SOAs here and in other figures are indicated by shaded triangles. E/O and O/E are electronic-to-optical converters (e.g., lasers) and optical-to-electronic converters (e.g., optical receivers), respectively.

However, when the SOA is used after the signal has been modulated and operated in saturation, it can distort the signal because the finite gain recovery time will introduce a time-dependent gain for each bit, which depends on the pattern of bits preceding it. For single-channel, intensity-modulated systems, this produces intersymbol interference (ISI). Because the gain recovery time is on the order of 100 ps, this will be a problem for multi-gigabit/s systems, where the bit period is comparable to the gain recovery time. An illustration of this is shown in Fig. 27, which shows an RZ bit sequence at 12.5 Gb/s, under conditions of minimal and 6-dB gain compression. The pattern-dependent gain produces significant eye closure, as shown. When the SOA is used as a power amplifier for analog-modulated signals, gain saturation can introduce second and higher-order distortions.¹⁹ To control ISI and analog distortions, such as composite second order (CSO) and composite triple beat (CTB), requires operating the SOA in a region of small gain compression, commensurate with the overall system application. Gain-clamped SOAs have been used for this application.¹⁹ Clearly, SOAs with large saturation powers offer advantages because they operate at higher output power for a given level of gain compression.

The pre-amp SOA is used in front of the receiver, as shown in Fig 26. At this point, the signal level is low so saturation of the amplifier is not an issue. The role of the SOA is to increase the optical power in the signal to make the received electrical power far exceed the thermal noise power in the receiver. Hence, the gain and noise figure are the important issues for the SOA when used as a pre-amp. To achieve maximum receiver sensitivity, the noise in the receiver should be dominated by signal-spontaneous beat noise, which requires a narrow-band optical filter after the SOA that is centered on the signal wavelength. At the receiver, the signal polarization has become randomized, so the polarization-dependence of the gain of the SOA pre-amp should be minimal (practically, less than about 1 dB).

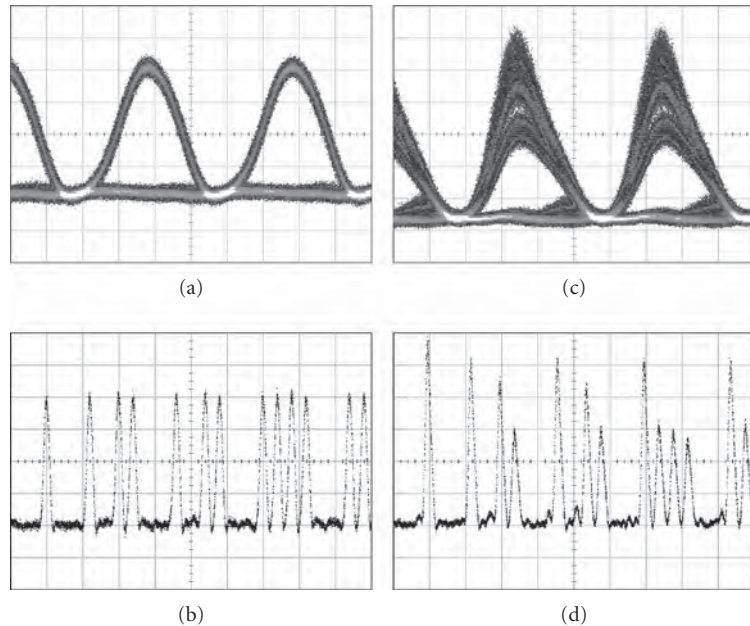


FIGURE 27 Effects of gain compression on bit sequences. Bottom traces show single-channel bit sequences at 12.5 Gb/s and top traces are corresponding eye diagrams. (a) and (b) correspond to the SOA operating under < 0.5 dB gain compression, while (c) and (d) correspond to the case of 6 dB gain compression. The effect of finite gain recovery on the pattern sequence and the corresponding eye diagrams is significant.

The function of the in-line SOA is to compensate for the loss in the fiber and other components in the transmission link. Issues of gain, ASE noise, gain recovery and its effect on ISI, and polarization-dependent gain arise for the in-line amplifiers. The output power level is determined by the operating characteristics for the transmission line, which require power levels compatible with management of fiber nonlinearities. Nevertheless, high-gain permits amplification to overcome the loss in long fiber links. The saturation power should be large, so that the amplifier can be operated at most lightly into the saturation regime. The ASE noise added to the signal by the SOA will degrade the optical signal-to-noise ratio and will ultimately limit the system. The noise figure of the SOA can be a limit on the number of SOAs that could be cascaded.

The ASE noise power in a single polarization from an optical amplifier is given by Eq. (11), which follows from Eq. (5), where

$$p_{\text{ASE}} = \frac{1}{2} n f (g - 1) h \nu B_0 \quad (11)$$

where p_{ASE} is the ASE power ($p_{\text{ASE}} = \rho_{\text{ASE}} B_0$), $n f$ is the noise figure, g is the gain, $h \nu$ is the photon energy, and B_0 is the optical bandwidth within which the ASE power is measured. Figure 28 shows a transmission line, which is composed of n fiber spans and n amplifiers. The loss of each fiber span, L , is exactly compensated by the gain g of each amplifier, so that the signal power is p_s at the output of each amplifier. Each amplifier adds ASE noise given by Eq. (11), so that the optical signal-to-noise ratio (OSNR), at the end of the line is given by

$$\text{OSNR} = \frac{p_s}{n p_{\text{ASE}}} = \frac{p_s}{n n f (g - 1) h \nu B_0} \quad (12)$$

Converting Eq. (12) to logarithmic units and referencing power levels to 1 mW (power is given in dBm), produces the very useful equation:

$$\text{OSNR(dB)} = 58 + P_s - G - \text{NF} - 10 \log n \quad (13)$$

In Eq. (13), P_s is the signal output power in dBm, G is the gain (and span loss L) in decibels, and NF is the noise figure in decibels. In arriving at Eq. (13), the gain is assumed much larger than unity and the optical bandwidth B_0 is taken to be 0.1 nm (at 1550-nm wavelength). Equations (12) and (13) also assume that the dominant optical noise is signal-spontaneous beat noise.¹⁴

Single-channel systems using SOAs to compensate fiber and other loss illustrate the general comments above. An early single-channel demonstration of a link amplified using SOAs was reported in 1996,³² for a 10-Gb/s RZ signal transmitted over 420 km at 1310 nm. In this experiment, there were 10 in-line SOAs (compensating ≈ 17 -dB loss from 40-km spans and other components), a power amplifier, and a preamplifier. The signal level was maintained below 10 dBm, well below the amplifier saturation power of 18 dBm. The signal degradation for this experiment is fully described by noise accumulation from the cascaded SOAs.

Successful demonstration of analog signals has been reported when the signal level is kept well below the saturation power. When used as a preamplifier for a CATV demonstration, which used a single wavelength carrying 23 QPSK subcarriers, a polarization-insensitive non-gain-clamped SOA enabled an 11-dB improvement for an 85-km link, (compared to a receiver without preamplifier).³³ Again, the main degradation arises from the SOA ASE noise.

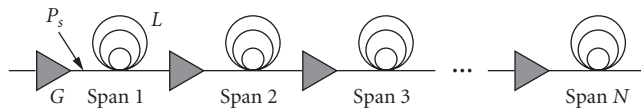


FIGURE 28 An amplified transmission line. Each fiber span has loss L (L in decibels), each SOA has gain g (G in decibels), and the power at the output of each SOA is p_s (P_s in decibels).

DWDM Systems

In DWDM systems, channels are typically spaced by 50 to 200 GHz. In the range around 1.55 μm , an eight-channel system spaced at 200 GHz can cover a spectral wavelength range of 11.2 nm. Over this wavelength range, while the gain and saturation power of the SOA can vary by a few decibels, as can be seen from Figs. 5a and 11, insight can be obtained by considering the gain and saturation powers to have averaged values.

In DWDM systems, power amplifiers can be used before combining channels, so their use is equivalent to the single-channel systems above. If the power amplifier is used after combining intensity-modulated signals, cross-gain modulation from one channel to the others will impress crosstalk onto the signals. This effect is mitigated by operating the amplifier under the conditions of small (at most ≈ 1 dB) gain compression. For equally spaced (in frequency) channels, four-wave mixing can potentially also cause crosstalk between channels. Except for systems that use short pulses at high bit rates (and large peak powers), four-wave mixing in the power amplifier is small compared to cross-gain modulation.³⁴

At the end of the system where the SOA can be used as a preamplifier, it is unlikely that the signals will saturate the amplifier, whether a single SOA is used to amplify many channels before they are separated or used for amplifying a single channel. The issues are the same as for a single-channel preamplifier: noise addition and polarization-dependent gain.

In-line amplifiers compensate for the span losses, which include losses due to dispersion compensation and other elements in nodes. As for the single-channel systems discussed above, the output power should be as large as possible to ensure overall large OSNR for the system. As the output power approaches the saturation power, the amplifier gain dynamics lead to time-dependent gain and crosstalk between channels caused by cross-gain compression. A variety of mitigation techniques have been proposed, including operation in a region of low saturation (≈ 1 dB gain compression),^{21,35,36} use of a reservoir channel that biases the SOA into a region of gain compression for which cross-gain modulation is reduced,^{36,37} use of a gain-clamped amplifier,³⁸ and the use of a pair of complementary signals to maintain a constant signal power into the amplifier.^{39,40}

An illustration of the effects of gain compression on systems using SOAs is shown in Fig. 29, which shows results for a system of eight channels at 40Gb/s, operating over two 80-km spans. As the

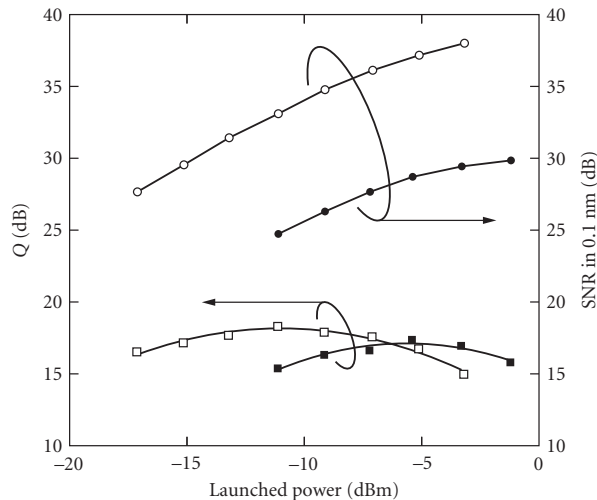


FIGURE 29 System results for 40-Gb/s transmission. The transmission line consists of two amplified 80-km fiber spans. Open symbols correspond to single-channel transmission, closed symbols correspond to transmission of eight channels, spaced at 200 GHz. Optimum Q is achieved in both cases for launch power in a single channel of ≈ -13 dBm.

optical power is increased, the OSNR increases, although the system performance, as measured by the Q-factor⁴¹ reaches a peak and declines at higher power levels, due to cross-gain modulation effects. The peak in input powers for one-channel and eight-channel systems is separated by ≈ 9 dB, and in both cases corresponds to the same total input power to the SOA of -5 dBm, which corresponds to about 1-dB gain compression for the SOA that is used in the experiment.

The system capacity is limited by the OSNR at about 1 dB of gain compression. In a rough way, for similar systems before the onset of nonlinearities, Q is almost linearly related to the OSNR. Also, for different bit rates, the optical bandwidth for the optimized receiver depends linearly on the bit rate. Assuming these linear relationships, and noting that the relevant signal output power in Eq. (13) is the sum of powers for all channels (and assuming a single value for gain and saturation power), Eq. (13) can be manipulated to give

$$10 \log c = 58 - k - P_{1\text{ dB}} - L - \text{NF} - 10 \log n \quad (14)$$

In Eq. (14) c is the overall system capacity in number of channels $\times$ bit-rate, $P_{1\text{ dB}}$ is the 1-dB gain compression power for the SOA, L is the span loss in dB (spans $\times$ loss/span), and k is a constant relating optimal Q to OSNR. The value of $P_{1\text{ dB}}$ for the output power has been used because this is the operating point for which nonlinearities caused by gain compression are minimal. Figure 30 shows a set of experiments for a variety of bit rates and links using similar SOAs. For these experiments, each systems capacity is determined for $Q \approx 16$ dB, corresponding to a bit-error rate $\approx 1 \cdot 10^{-9}$. The horizontal axis is the total link loss, which is the number of spans times the loss per span. In these experiments, the span losses were in the range of 15 to 17 dB and channel separations were either 100 or 200 GHz. The results

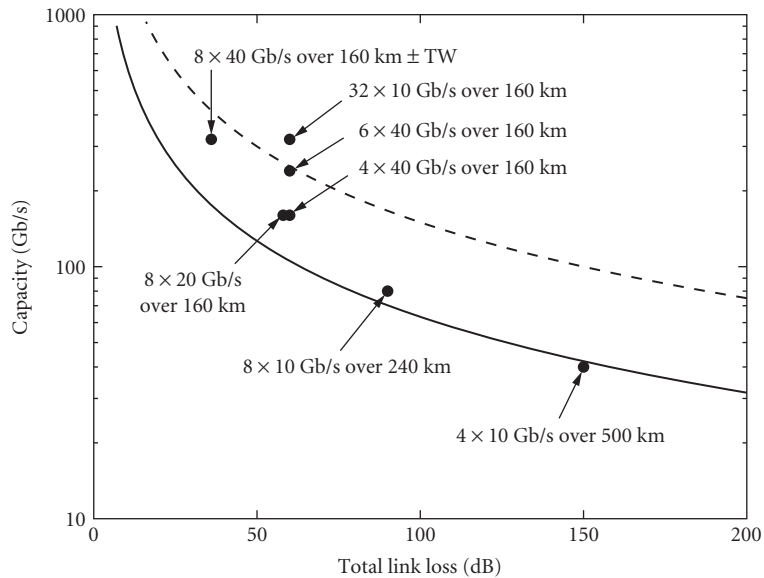


FIGURE 30 Results of a series of DWDM system experiments over multiple fiber spans, using a set of nominally similar SOAs. Overall system capacity, which is the number of channels times the bit-rate per channel, is plotted against the total link loss, which includes loss of dispersion compensating fiber for each span. The systems use NRZ intensity modulation and capacity is determined where all signals achieve a bit-error rate of 10^{-9} ($Q \approx 15.8$ dB). The two curves show expectations for all spans with 20-dB loss (solid line) and 15-dB loss (dashed line). TW: TrueWave fiber.

show that the analysis in terms of OSNR captures the main systems limitations and describes the overall system capacity.

When SOA systems are used with forward error correction (FEC) coding, Q values as low as 8 dB can be corrected to essentially error-free operation. This permits operation at lower OSNR (about 8 dB) and from Eq. (14), this will enable the overall system loss (and therefore length or capacity) to be increased by about 8 dB from the results shown in Fig. 30.

For phase-modulated signals, the lack of temporal intensity-dependence of the input signals will reduce the effects of time-dependent gain and cross-gain compression. In this case, the dominant nonlinearity, which will limit transmission, will be four-wave mixing. Thus, large system capacity and distance have been achieved using differential phase shift keying (DPSK) modulation, but larger channel separation is required to reduce four-wave mixing effects. Using DPSK modulation and SOAs operating under 3-dB gain compression, it was possible to operate an eight-channel, 10.7-Gb/s system over 1050 km, with a FEC margin of 6 dB.⁴² Because of four-wave mixing, the channels were spaced by 400 GHz. Four-wave mixing can also lead to system penalties comparable to those from cross-gain modulation for high bit-rate systems using short RZ pulses.³⁸

CWDM Systems

Coarse WDM (CWDM) systems are envisioned to be inexpensive and environmentally tolerant for access and metro applications, with a reach of up to about 100 km. To this end, the channel spacing is set to 20 nm in a wavelength range that spans the window from about 1250 to 1650 nm—a window that could include 20 channels. The coarse spacing allows for very reduced temperature and wavelength stability requirements on the components, and requires filters with large, almost 20-nm bandwidths. Amplification can extend the reach of such systems.

The broad gain bandwidth of SOAs is advantageous for application to CWDM systems, although a single SOA is incapable of covering the entire 1200- to 1600-nm window. Nevertheless, a broadband LOA has been used to amplify eight CWDM channels covering the range from 1470 to 1610 nm.⁴³ To cover the full range, a set of SOAs designed to amplify different gain bands can be used.

Because the channel wavelength is specified only to within a band of 20 nm, filters that separate the different channels are much wider than the filters used in DWDM systems. Therefore, in addition to signal-spontaneous beat noise, the component of optical noise due to beating of the ASE from the SOA with itself, within the signal bandwidth, (spontaneous-spontaneous beat noise) is also significant. Equations (12) and (13) are no longer valid. For filter bandwidths of 13 nm, the addition of the spontaneous-spontaneous beat noise will reduce the sensitivity of a preamplified receiver by about 3 dB, assuming about 25-dB amplifier gain and a bit rate ≈ 5 Gb/s.¹⁴ Other issues for SOAs used in CWDM systems are the wavelength variation of the gain and saturation powers. Because the gain will vary by more than 3 dB over the SOA bandwidth used, a cascade of amplifiers would quickly lead to very unbalanced channels. With only one amplifier in the system, and only ≈ 100 km of fiber, this does not lead to much difficulty. For a cascade of amplifiers and fiber spans, however, weaker channels would rapidly degrade in OSNR and fiber nonlinearities would build up rapidly for the stronger channels.

Cross-gain modulation becomes an impairment when the amplifier is operated in saturation. All input wavelengths can contribute to saturation and reduction of the carrier density, but for equal input power in each channel, the channels nearest the gain peak will cause the largest effect. Because the differential gain is largest on the short-wavelength side of the gain peak, the carrier density changes will cause the largest gain compression for these channels. Therefore, the short wavelength channels will be most affected by cross-gain modulation. The nonlinearity caused by four-wave mixing will not be significant in CWDM systems, because the wavelengths are so widely separated and the four-wave mixing interaction falls off rapidly with wavelength (see Fig. 37).

For broader bandwidth coverage, a hybrid amplifier composed of an SOA and a fiber Raman amplifier, has been applied to CWDM systems.⁴⁴ Here, the Raman pump is chosen so that the Raman gain peak lies to longer wavelength than the SOA peak. The composite amplifier can have a 3-dB gain bandwidth of over 120 nm. Wireless over fiber transmission has also been demonstrated in a CWDM system, employing a bidirectional SOA.⁴⁵

19.7 SWITCHING AND MODULATION

An unbiased SOA is strongly attenuating. The difference in output of a signal emerging from an SOA with gain compared to one attenuated by an unpumped SOA can be as high as 60 dB. This forms the basis of an optical switch, with very high extinction ratio.⁴⁶ Figure 31 shows a 2×2 switch fabric, where the switching elements are SOAs. For this simple switch matrix, the bar state is enabled by supplying bias, and therefore gain, to SOAs 1 and 4, while SOAs 2 and 3 are unbiased. Thus, the cross state is highly attenuated. For the cross state, the opposite set of biases are applied—SOAs 2 and 3 are biased and SOAs 1 and 4 are not biased.

The SOA-based switch is capable of moderately fast reconfiguration. The temporal response of an SOA between full on and full off states is several times the gain recovery time (to achieve high extinction), but can generally be ≈ 1 ns. To switch the state of the SOA by adjusting its drive current, the response must also include electrical parasitics, which can have bandwidths of up to a few gigahertz (or less), depending on device design. Thus, gigahertz rearrangements for switching fabrics based on SOAs are possible, and this is one of their significant benefits. For a switching application, the signal transmitting the fabric should not be distorted and should have minimal noise degradation. Thus, signals passing through SOAs in a switch fabric are in these regards similar to signals passing through cascades of SOAs in the transmission systems described above.

The fabric above is intensive in the number of SOA gates required per port. Other switch geometries, involving combinations with AWGs, can reduce this requirement.^{47–50} Because of the high-speed response of the SOA switch fabrics, they are often used in space-switching stages of larger all-optical cross-connects, which may have additional features of wavelength interchange (using wavelength converters, such as discussed below). Impressive all-optical nodes, including all-optical packet switching nodes, have been constructed using these switches.^{51–53}

Because the gain of the SOA varies with the bias current, the SOA could be used as an optical modulator. The response speed of an SOA as a modulator is significantly slower than the response of a directly modulated laser, because there is no optical resonator and hence the response has no relaxation oscillation peak. The modulation bandwidth can be ≈ 2 GHz, and direct modulation has been achieved for a 2 Gb/s bit rate.⁵⁴ Of course, the optical signal would be highly chirped. The amplifier does provide gain and its simplicity could be advantageous for some applications. The concept of a reflective SOA (RSOA) as a modulator was introduced for simple access systems, where the RSOA at the user node could amplify and modulate a wavelength that originates from the central office.⁵⁵ Various improvements and applications use RSOAs in access networks with bit rates beyond 1 Gb/s.⁵⁶

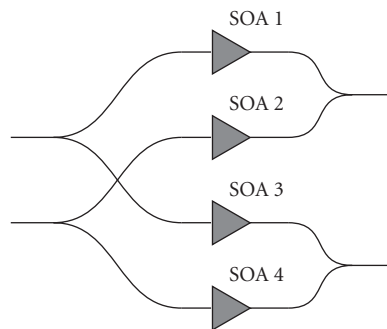


FIGURE 31 2×2 switch fabric based on SOA gates.

19.8 NONLINEAR APPLICATIONS

The gain and index of refraction within the SOA depend on several different physical processes with different time scales. The magnitudes and particular time responses will depend on material, structures (multiple quantum well vs. quantum dot vs. bulk) and operating conditions, such as wavelength, power, bit-rate, modulation format, and the like. There is a rich variety of effects and their applications. The general applications of these nonlinearities lead to optically controlled optical gates. The three most significant nonlinearities in SOAs are cross-gain modulation, cross-phase modulation and four-wave mixing. XGM and XPM depend on the optical intensity, while FWM depends on the optical field. The following sections describe the application of these nonlinearities.

Cross-Gain Modulation

The principle of XGM is shown in Fig. 32. In this process, a strong, saturating input (pump) signal at λ_1 and a weaker input at λ_2 (probe) are simultaneously injected into the SOA. The injection of the strong optical signal at λ_1 reduces the carrier density, which compresses the gain at all wavelengths. When the pump signal is turned off, the gain recovers. Thus, an inverted copy of the intensity modulation on the pump signal is copied onto the probe signal, as in Fig. 32. The device has functioned as a wavelength converter.^{57,58} Note that if probes at several wavelengths are input to the SOA, each probe experiences gain compression and the pump signal can be copied simultaneously onto each.⁵⁹ The schematic shows that the pump and probe signals copropagate and are separated by an external filter. It is possible to avoid the filter by counter-propagating the two signals. This reduces the highest speed capabilities of XGM, however, due to traveling-wave effects.

As the free carrier density is reduced, the gain compression depends on wavelength, as can be seen by comparing ASE curves for different pumping currents (related to carrier density) in Fig. 5b.

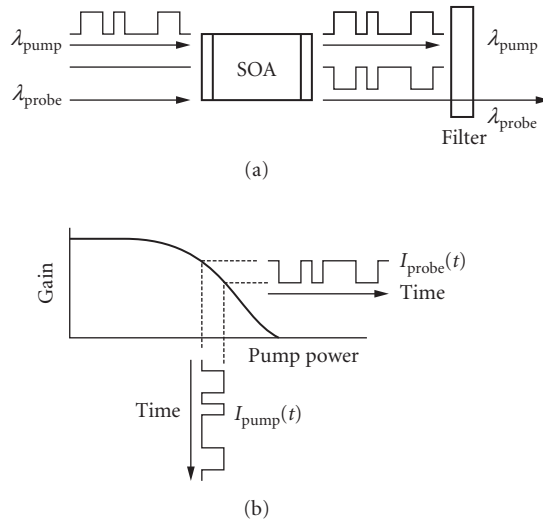


FIGURE 32 Principle of cross-gain modulation in a SOA. (a) In basic operation, a strong, modulated pump beam and a CW probe beam are input to the SOA. (b) The pump power is intense enough to compress the gain of the SOA, which creates an inverse of its intensity modulation on the probe. In this colinear arrangement, the outputs are separated by a filter (a).

Thus, the gain is compressed more for shorter wavelengths than for longer wavelengths, and so the extinction ratio (which is the gain compression) for the converted signal will be larger for conversion from longer to shorter wavelength than vice versa. At different wavelengths, the power required for gain compression will scale with the saturation power, which increases with longer wavelength (Fig. 11). To achieve an extinction ratio of 10 dB or larger, the pump output power must be several times the saturation power, as can be seen from Fig. 10.

The polarization sensitivity for wavelength conversion by XGM depends on the polarization dependence of the SOA gain. For SOAs with minimal polarization dependence, there is minimal polarization sensitivity for XGM. Dynamical processes that are faster than carrier recombination can lead to anisotropies even in nominally polarization-insensitive SOAs,⁶⁰ but these processes are fast and are not significant in cross-gain modulation wavelength conversion for multi-gigabit/s signals.

The temporal response of a wavelength converter based on XGM depends on the gain recovery time, which is dominated by the carrier recovery time. For SOAs, which operate with carrier densities in the range of 5×10^{18} to $10 \times 10^{18} \text{ cm}^{-3}$, the gain recovery is dominated by Auger recombination and is typically in the range of 100 ps (which is consistent with the data of Fig. 14). This would seem to support an operating bit rate of up to 10 Gb/s. To enable operation at higher bit-rates requires reducing the gain recovery time, which has been accomplished in several ways. Adding another significant recombination process speeds up gain recovery and can be accomplished by using stimulated emission. If the input signal powers are high enough, stimulated emission can become strong enough to exceed the Auger recombination rate and reduce the gain recovery time to enable operation at 40 Gb/s.⁶¹ Use of a long amplifier also helps, as the overall gain increases the stimulated emission rate and relatively enhances the gain for high-frequency signal components.^{62,63} In practice, stimulated emission can be increased by using a relatively high-power probe beam⁶⁴ or using a third beam, the control beam, to set the rate of stimulated emission.⁶⁵ Different materials can also affect the gain recovery time by carrier recombination. Recently, quantum dot material has shown promise for high-speed operation, because of its short-gain recovery time.⁶⁶ SOAs based on p-type modulation-doped MQW active regions have shown a fast intrinsic carrier recombination time, which can be exploited for high-speed nonlinear devices.⁶⁷ Another method for higher-speed operation of XGM involves filtering, which will be described next.

The carrier density affects not only the gain, but also the refractive index of the inverted semiconductor material. Thus, the temporal variation of the gain will impose a temporal variation on the refractive index within the SOA, which will lead to chirp on the wavelength-converted signal.⁶⁸ For large gain compression, the refractive index change can be several times π radians. The phase change corresponding to a gain change can be given approximately by¹⁷

$$\frac{\Delta\phi}{\Delta G} = -\alpha/8.68 \quad (15)$$

where ΔG is the gain change in decibels and α is the linewidth enhancement factor for semiconductor lasers and amplifiers.⁶⁹ For a value of $\alpha \approx 8$, a phase-change of π requires about 3.4 dB of gain compression.

Because the wavelength-converted output at λ_2 is chirped, filtering can reduce both the excess spectral content of the output and its temporal response. Thus, narrow-band filtering at the output of the SOA can be used to effectively speed up the response of the wavelength converter.⁷⁰ This approach has been used to achieve 100-Gb/s operation⁷¹ and up to 320-Gb/s operation of a wavelength converter based on XGM.⁷²

Cross-Phase Modulation

The refractive index changes with carrier density (and gain) led to chirp for signals created by XGM. The phase change caused by the refractive index change is the basis for another set of optical-optical gates and wavelength converters based on cross-phase modulation. To use XPM with intensity-modulated signals requires converting phase to amplitude and thus requires an interferometric structure.

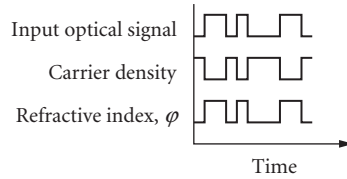


FIGURE 33 Principle underlying cross-phase modulation. An intense input optical signal (top) passes through the SOA. This signal reduces the carrier density by stimulated emission (middle). The refractive index of the active region depends on the free carrier density, increasing when the carrier density decreases (bottom). The phase change experienced by the probe propagating through the SOA is used for cross-phase modulation switching. Note that the temporal response is assumed instantaneous for this figure.

The operating principle is shown in Fig. 33. An input intensity-modulated signal will reduce the carrier density within the SOA by stimulated emission and reduces the gain as well, resulting in XGM. The refractive index within the active region depends on the free carrier density and increases as the free carrier density decreases. This produces a phase change $\Delta\phi = \Delta nL$, for a device of length L (assuming uniform Δn), as shown in the bottom of Fig. 33. This phase change can be read out by a second optical signal, if the SOA is placed in some kind of interferometric structure through which the second signal propagates.

The Mach-Zehnder structure, with an SOA in each arm, is used for many applications based on XPM.⁷³ Fig. 34a shows a Mach-Zehnder interferometer with an SOA in each arm and three input waveguides. A CW probe at λ_2 is input from the left. The relative phases in each arm can be adjusted by controlling the carrier density in each SOA by its bias current. If the phases of the two arms are

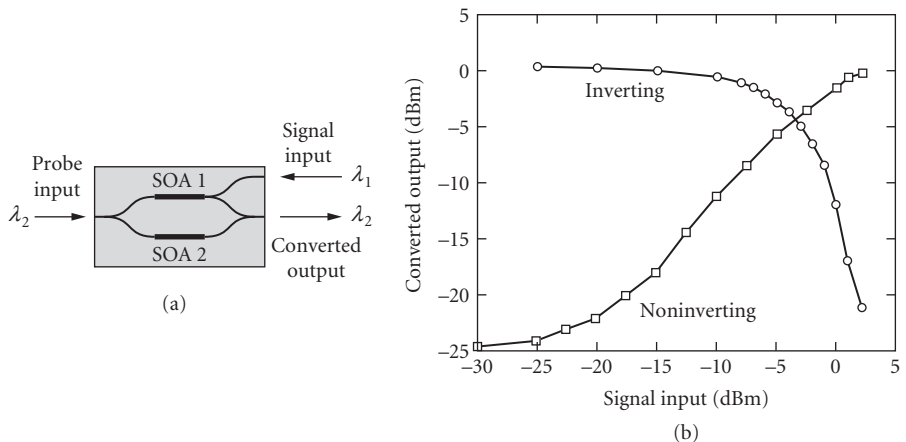


FIGURE 34 Mach-Zehnder interferometer used for cross-phase modulation wavelength conversion. (a) The arrangement shows the input (pump) signal and the probe signal counter-propagating. The input signal modulates the phase in SOA 1, switching the output state of the interferometer for the probe signal. (b) The output signal for inverting and noninverting operation, under CW conditions. Because of the nonlinear shape of the switch curves, it is possible to increase the extinction ratio from input to converted output.

equal, the two arms constructively interfere at the output port, producing a maximum for the probe at the output port. Note that each SOA also has gain, so that the output probe has greater intensity than the input. If the two arms have a phase difference of π , the probe experiences destructive interference at the output port and a minimum intensity. The values of the maxima and minima for points of constructive and destructive interference also depend on the gain in each arm. The greatest static extinction ratio occurs when the gains of the arms are equal. When an intensity-modulated signal at λ_1 is input into one arm of the interferometer (Fig. 34a), it changes the phase in that arm. If the phase is changed by π , the output of the probe signal can be switched from on to off, for the case of initial constructive interference, and from off to on, for the case of initial destructive interference. These cases correspond to inverting and noninverting operation, respectively. Figure 34b shows switching curves for inverting and noninverting operation. The extinction ratio for such static switching can be in excess of 25 dB (sometimes as large as 40 dB).

The free carrier density dependence of the refractive index has a very weak dependence on wavelength. Therefore, for a given gain compression, the extinction ratio and wavelength conversion penalties are almost independent of the direction and magnitude of wavelength shift (as long as it remains within the gain bandwidth of the SOA). In operation, the dynamic extinction ratio is not as large as the static extinction ratio, because of carrier dynamics which, at multi-gigabit/s bit rates, do not permit the device to reach a steady state. Nevertheless, dynamic extinction ratios of 15 dB are generally achieved.

The carrier density is varying during the pump intensity transitions, so chirp is imposed on the converted signal transitions. For the case of noninverting operation, the sign of the chirp leads to pulse compression in standard single-mode fibers in the 1550-nm regime.⁶² For inverting operation, the sign of the chirp leads to immediate pulse broadening in standard SMF.

The polarization dependence of XPM in the Mach-Zehnder structure depends on the polarization-dependence of its components. For polarization-independent operation, the SOAs and the waveguides must be made polarization-insensitive. It is possible to optimize the Mach-Zehnder structures in many other ways, to permit counter-propagation of pump and probe signals, to separate pump and probe signals without need of an external filter, and to maximize the overlaps of optical fields inside the SOA active regions.^{74,75} Other interferometer structures, such as Michelson and Sagnac, have also been used for nonlinear applications of XPM.

The gain recovery time in the SOA also affects the phase recovery time, which limits the response speed for XPM. To increase the operating speed of the optical-optical gate based on XPM in a Mach-Zehnder interferometer, it is possible to operate the device in a differential mode,⁷⁶ as illustrated in Fig. 35. By supplying pump inputs into the SOA in each arm, but with a delay τ between them, it is possible to carve a gating window of duration τ . Figure 35b shows that the phase dynamics of the SOA in the upper arm can be balanced by the phase dynamics in the lower arm. The output phase is determined by the difference in phase between the two arms, which creates the gating window. Of course, the relative amplitudes of the pumps in each arm must also be properly balanced so that full cancellation of the time-varying phase in each arm can be achieved, to create a gate with high extinction. Figure 35c shows the gating windows that can be carved by varying the relative delays between pump inputs to the upper and lower SOAs. Using this device as an optical demultiplexer, a 10.5-Gb/s signal has been demultiplexed from a 168-Gb/s data stream.⁷⁶

XPM in interferometric structures that require only a single SOA have also been demonstrated. The ultrafast nonlinear interferometer (UNI) device⁷⁷ operates by creating an interferometer based on orthogonal polarization states propagating through the SOA. By proper timing of the pump pulse with respect to temporally delayed probe pulses in the orthogonal polarizations, the pump can affect the index of only one polarization state. Thus, when the two orthogonal polarizations are recombined, switching is effected. In a different arrangement, a single SOA in which a pump and probe copropagate is combined with a 1-bit delay interferometer to convert XPM to intensity modulation, effectively copying the signal from the pump to the probe.⁷⁸ This can be fast and the SOA and delay interferometer can be integrated monolithically, but it does operate for a fixed bit rate, set by the delay interferometer.

In all of the above discussions, intensity modulated signals were required for both XGM and XPM. Wavelength conversion of phase-modulated signals has also been demonstrated, in a sophisticated two-stage device based on XPM, using a delay interferometer and a Mach-Zehnder wavelength converter.⁷⁹

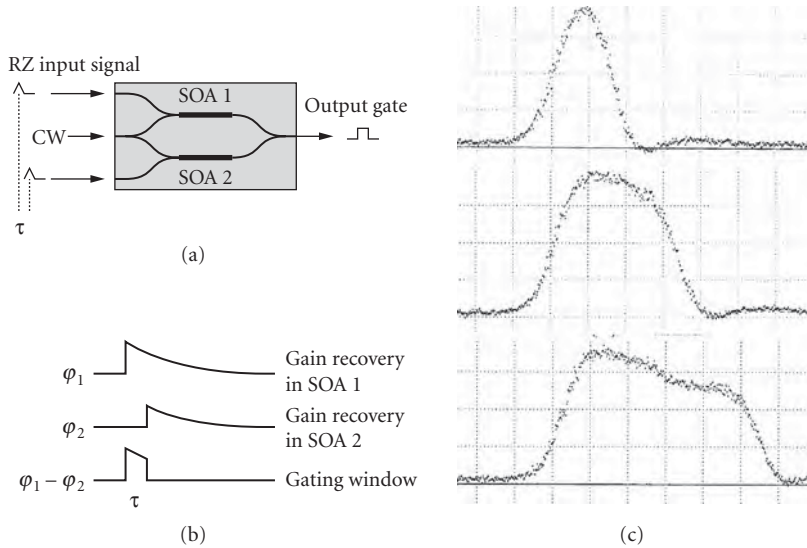


FIGURE 35 Differential operation of the Mach-Zehnder wavelength converter. (a) The input pulses, separated by a time delay τ , are sent into SOA 1 and SOA 2. The first pulse, to SOA 1, creates a phase change in the top arm, which switches the interferometer output for the CW probe input. The second pulse, to SOA 2, creates a phase change in the bottom arm, which cancels the remaining phase change in the top arm, closing the interferometer output. (b) Schematic of the gain recovery and the creation of a gating window. (c) Examples of tuning the switching window by varying the separation of pulses. The horizontal scale is 20 ps/division.

Four-Wave Mixing

Four-wave mixing is the fastest nonlinearity in SOAs, but it is very sensitive to polarization and to the wavelength separation between inputs. This is a $\chi^{(3)}$ nonlinearity, which depends on the optical field rather than optical intensity.⁸⁰ Figure 36 shows a schematic for four-wave mixing in a SOA with two input wavelengths. As explained previously, in four-wave mixing the two input fields interfere in the SOA to produce phase and index gratings. Each of the input fields then scatters off these gratings, producing longer and shorter wavelength sidebands, as shown in Fig. 16. If the two

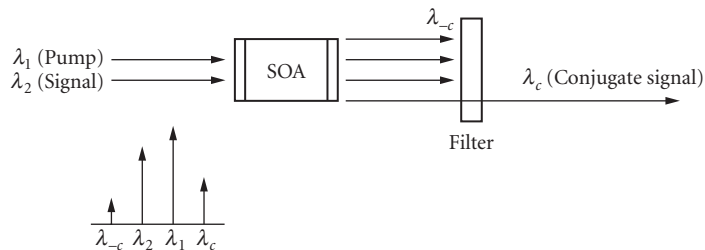


FIGURE 36 Four-wave mixing in a SOA. A strong pump and a signal are input to the SOA. Nonlinear mixing creates sidebands separated from the input pair by the difference in frequency between inputs (bottom). The desired signal, the conjugate signal, is separated from the inputs and the other sideband by a filter (top).

input wavelengths are λ_p and λ_s , for pump and signal, respectively, conservation of energy produces a converted signal λ_c at

$$\frac{1}{\lambda_c} = \frac{2}{\lambda_p} - \frac{1}{\lambda_s} \quad (16)$$

If the pump signal is CW, the converted signal will carry both the amplitude and the phase-conjugate information of the original signal. Thus, FWM is an optical-optical gate that, when used for wavelength conversion, straightforwardly works for both amplitude and phase-modulated signals. The converted signal described by Eq. (16) is created using one photon from the input signal and two from the input pump. The other sideband at λ_{-c} , which is created from two signal photons and one pump photon ($1/\lambda_{-c} = 2/\lambda_s - 1/\lambda_p$), has a more complicated amplitude and phase relation to the original signal.

An example of the dependence of four-wave mixing efficiency as a function of wavelength separation between the signal and pump beams is shown in Fig. 37. The conversion efficiency is weak, depends strongly on wavelength separation, and is asymmetric to longer or shorter wavelength. The shape and asymmetry arises because the nonlinear gratings formed within the SOA arise from several dynamical processes.⁸¹ Carrier density modulation, which has a characteristic lifetime of ≈ 100 ps, is responsible for the highest efficiency mixing process, but only to frequency shifts of ≈ 10 GHz (≈ 0.1 nm at 1550 nm). Carrier heating and cooling processes, which have a time constant ≈ 1 ps, are responsible for frequency shifts to ≈ 1 THz (≈ 8 nm), while the interband Kerr effect, with a characteristic time constant of ≈ 1 fs, is responsible for frequency shifts beyond 1 THz. Each dynamical process has its own amplitude-phase coupling constant and the overall four-wave mixing process is the coherent addition of nonlinear interactions from each process. Because of the different amplitude-phase coupling constants, the conversion efficiency is asymmetric with wavelength. Four-wave mixing requires phase-matching as well as energy conservation [Eq. (16)]. For typical SOA lengths below ≈ 1 mm

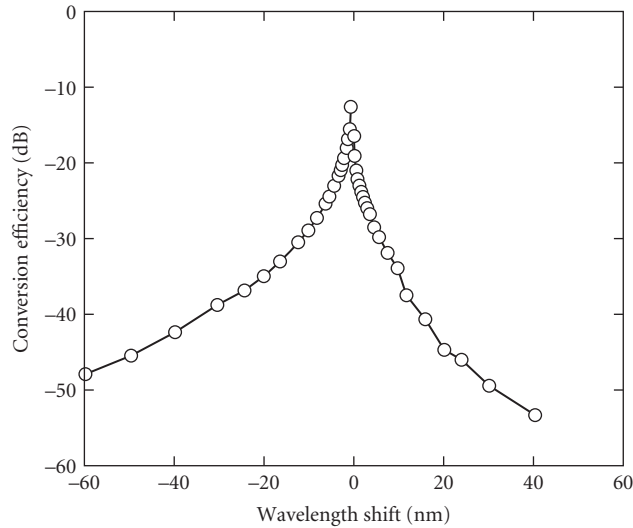


FIGURE 37 Four-wave mixing efficiency as a function of wavelength separation between pump and signal inputs, for a SOA operating near 1550 nm. For this experiment, the two inputs have equal intensity. Note the asymmetry and that conversion efficiency is larger to the shorter wavelength (higher frequency) side.

and for frequency shifts of less than several terahertz, phase-matching requirements are satisfied. For wavelength shifts greater than ≈ 100 GHz, the dynamical processes responsible are on the 1-ps or shorter time scale, so the response speed of FWM is very fast.

The FWM process is capable of shifting multiple signal wavelengths simultaneously, with a single pump wavelength. Each signal produces a converted output as described by Eq. (16). Thus, a band of wavelengths may be shifted simultaneously (but the wavelengths are not shifted independently).

The conversion process is weak: typically for input signals of a few dBm, conversion efficiency varies from -10 to -40 dB, as the wavelength shift increases. Because the SOA also produces ASE noise, the signal-to-noise ratio of the converted output can be affected. For conditions of small saturation, in addition to the wavelength shift, the converted power depends on $g^3 p_p^2 p_s$, where g , p_p , and p_s are the gain, pump power, and signal power, respectively, while the ASE noise depends on g .⁸² For saturated conditions, the converted power can increase more rapidly with pump power.⁸³ The efficiency of FWM can be enhanced by using long SOAs, up to 2 mm in length.^{84,85} This effect can be understood by noting that in a long SOA, the later portion is saturated and provides little gain. However, the nonlinearity is active in this region and the FWM fields grow coherently and quadratic in the SOA length, while the ASE adds incoherently and grows linearly in SOA length.⁸⁶ Therefore, the efficiency and the signal-to-noise ratio for FWM are enhanced in a long SOA.

FWM is very polarization sensitive—the results in Fig. 37 are obtained with all polarizations parallel. More complicated arrangements, involving two nondegenerate pump wavelengths, have been used to make the FWM process polarization-insensitive and to flatten the wavelength dependence of the conversion efficiency.^{87,88}

FWM in SOAs has been used for a variety of optical-optical gate functions, including wavelength conversion,⁸⁹ optical sampling,⁹⁰ optical logic, and the like. Because the converted signal preserves the phase information of the input signal, but is its phase-conjugate, FWM in SOAs has also been applied to dispersion compensation in fiber transmission systems.^{91,92}

Further Comments for Nonlinear Applications

The optimum length of the SOA depends on the application.⁸⁶ For linear applications, high-gain, high saturation power, and low noise figure are desirable properties. For short amplifiers with high gain, the carrier density (and therefore pumping current density) is high. This will lead to a large inversion factor and therefore low noise figure. Also, the carrier lifetime will be short because of the large carrier density, and the differential gain coefficient may be small because of band-filling, resulting in a good saturation power. However, if the SOA is too short, the required pumping current density to produce the desired gain may be unsustainable without damaging the SOA. Alternatively, large saturation power for a given gain can be achieved in a long amplifier with a lower gain coefficient. This can be achieved using a thin active region, which increases the saturation power by decreasing the mode confinement factor, Γ , increasing the carrier density, which reduces the carrier lifetime, and again decreasing the differential gain coefficient because of band-filling. In long SOAs, however, the ends of the SOA can be saturated by the amplifier's own ASE, which results in decreased inversion and higher noise figure. Additionally, the effects of internal loss in the SOA are more significant when the gain coefficient is low (because of Γ), which also effects the carrier inversion and noise figure. SOAs of various lengths have therefore been used for linear applications.

For nonlinear applications, it is desirable to have amplifiers with lower saturation input powers (the input power that would produce 3-dB gain compression of the amplifier gain), to enable XGM and XPM with smaller pump power requirements. This is achieved in high-gain amplifiers, with relatively longer lengths. Because the probe signal is input with high power, the ASE noise is of less significance for XGM and XPM than for linear systems. Also, for longer SOAs the high rate of stimulated emission can shorten the carrier lifetime, leading to shorter gain-recovery times. To fully understand the temporal response of an SOA for XGM or XPM, however, requires modeling the SOA as a distributed amplifier, which shows that the modulation frequency response varies along the length of the amplifier and depends on the magnitude of the internal loss.^{62,63,93} The final, output section of the SOA, where the gain is compressed significantly, has a gain response with larger bandwidth than

the preceding sections, which leads to high-speed operation. In general, the temporal response of the SOA for XGM or XPM is enhanced for a long amplifier and enhanced by large Γ , large differential gain coefficient, high current (for large carrier density), and large input power. For FWM applications, a long amplifier is also beneficial. The front portion of the SOA can amplify the input signals, while the latter, saturated section can build up the nonlinear interactions, as discussed in “Four-Wave Mixing” in Sec. 19.8.

The nonlinear applications for SOAs were illustrated above mostly for the function of all-optical wavelength conversion. But in general, the application is that of an optical-optical gate, which can be used for various all-optical logic functions, such as optical demultiplexers (as illustrated for a Mach-Zehnder based device, above,⁷⁶ XOR and other logic gates, optical header processors, optical sampling, and all-optical regeneration. The optical transfer function for an interferometer, such as the Mach-Zehnder wavelength converter in Fig. 34, is decidedly nonlinear, so it can be used as a 2R (reamplifying and reshaping) regenerator. When it is combined with optical retiming, it becomes a key element in a 3R optical regenerator (reamplify, reshape, retime). There is a significant body of work on all-optical regeneration, which often uses nonlinear SOAs as a fundamental gating and shaping mechanism.⁹⁴

19.9 FINAL REMARKS

This chapter has described the properties, technology, and applications for SOAs. It is not meant to be a complete literature review on the subject, but rather to describe the principles underlying the devices and their basic applications. The technology of semiconductor optical amplifiers has reached a level of maturity where devices with well-defined properties are supplied and characterized. They are capable of supplying linear gain for a number of applications and systems of low to moderate complexity. As stand-alone gain blocks, the fact that, unlike fiber amplifiers, their gain spectrum can be tuned by their stoichiometry, permits them to be used in systems with many wavelength ranges or broad wavelength ranges, such as CWDM.

Increasingly, SOAs are key elements in photonic integrated circuits, both for applications of their nonlinear functionality as well as for gain blocks. PICs of moderate size are being developed and can incorporate tens of active elements, including SOAs. For example, SOAs have been used as power amplifiers for ten-channel 40 Gb/s per channel transmitter-PICs.⁹⁵ SOAs are used as gain blocks and also the basic nonlinear elements in medium-size PICs for functionality such as optical packet forwarding,⁹⁶ wavelength selection,⁵⁰ and wavelength conversion.⁹⁷

19.10 REFERENCES

1. J. Crowe and W. Ahearn, “9B8—Semiconductor Laser Amplifier,” *IEEE J. Quantum Electron.* **2**(8):283–289, 1966.
2. T. Mukai and Y. Yamamoto, “Gain, Frequency Bandwidth, and Saturation Output Power of AlGaAs DH Laser Amplifiers,” *IEEE J. Quantum Electron.* **17**(6):1028–1034, 1981.
3. G. Eisenstein, R. M. Jopson, R. A. Linke, C. A. Burrus, U. Koren, M. S. Whalen, and K. L. Hall, “Gain Measurements of InGaAsP 1.5 μ m Optical Amplifiers,” *Electron. Lett.* **21**(23):1076–1077, 1985.
4. M. J. O’Mahony, “Semiconductor Laser Optical Amplifiers for Use in Future Fiber Systems,” *J. Lightwave Technol.* **6**(4):531–544, 1988.
5. T. Saitoh and T. Mukai, “Recent Progress in Semiconductor Laser Amplifiers,” *J. Lightwave Technol.* **6**(11):1656–1664, 1988.
6. P. Doussi  re, P. Garabedian, C. Graver, D. Bonnevie, T. Fillion, E. Derouin, M. Monnot, J. G. Provost, D. Leclerc, and M. Klenk, “1.55 μ m Polarisation Independent Semiconductor Optical Amplifier with 25 dB Fiber to Fiber Gain,” *IEEE Photon. Technol. Lett.* **6**(2):170–172, 1994.
7. P. J. A. Thijs, L. F. Tiemeijer, P. I. Kuindersma, J. J. M. Binsma, and T. van Dongen, “High-Performance 1.5 μ m Wavelength InGaAs-InGaAsP Strained Quantum Well Lasers and Amplifiers,” *IEEE J. Quantum Electron.* **27**(6):1426–1439, 1991.

8. J. Y. Emery, T. Ducellier, M. Bachmann, P. Doussi re, F. Pommereau, R. Ngo, F. Gaborit, L. Goldstein, G. Laube, and J. Barrau, "High Performance 1.55 μm Polarisation-Insensitive Semiconductor Optical Amplifier Based on Low-Tensile Strained Bulk GaInAsP," *Electron. Lett.* **33**(12):1083–1084, 1997.
9. K. Magari, M. Okamoto, and Y. Noguchi, "1.55 μm Polarization-Insensitive High Gain Tensile-Strained-Barrier MQW Optical Amplifier," *IEEE Photon. Technol. Lett.* **3**(11):998–1000, 1991.
10. L. F. Tiemeijer, P. J. A. Thijs, T. van Dongen, R. W. M. Slootweg, J. M. M. van der Heijden, J. J. M. Binsma, and M. P. C. M. Krijn, "Polarization Insensitive Multiple Quantum Well Laser Amplifiers for the 1300 nm Window," *Appl. Phys. Lett.* **62**(8):826–828, 1993.
11. M. A. Newkirk, B. I. Miller, U. Koren, M. G. Young, M. Chien, R. M. Jopson, and C. A. Burrus, "1.5 μm Multiquantum-Well Semiconductor Optical Amplifier with Tensile and Compressively Strained Wells for Polarization-Independent Gain," *IEEE Photon. Technol. Lett.* **5**(4):406–408, 1993.
12. A. E. Kelly, I. F. Lealman, L. J. Rivers, S. D. Perrin, and M. Silver, "Polarisation Insensitive, 25 dB Gain Semiconductor Laser Amplifier without Antireflection Coatings," *Electron. Lett.* **32**(19):1835–1836, 1996.
13. L. F. Tiemeijer, P. J. A. Thijs, T. van Dongen, J. J. M. Binsma, and E. J. Jansen, "Polarization Resolved, Complete Characterization of 1310 nm Fiber Pigtailed Multiple-Quantum-Well Optical Amplifiers," *J. Lightwave Technol.* **14**(6):1524–1533, 1996.
14. N. A. Olsson, "Lightwave Systems with Optical Amplifiers," *J. Lightwave Technol.* **7**(7):1071–1082, 1989.
15. A. E. Siegman, *Lasers*, Section 7.7, pages 297–303. University Science Books, Mill Valley, 1986 (from Eq. 82).
16. K. Morito, M. Ekawa, T. Watanabe, and Y. Kotaki, "High-Output-Power Polarization-Insensitive Semiconductor Optical Amplifier," *J. Lightwave Technol.* **21**(1):176–181, 2003.
17. J. M. Wiesenfeld, "Gain Dynamics and Associated Nonlinearities in Semiconductor Optical Amplifiers," *Int. J. High-Speed Electron. Sys.* **7**(1):179–222, 1996.
18. M. Bachmann, P. Doussi re, J. Y. Emery, R. N'Go, F. Pommereau, L. Goldstein, G. Soulage, and A. Jourdan, "Polarisation-Insensitive Clamped-Gain SOA with Integrated Spot-Size Converter and DBR Gratings for WDM Applications at 1.55 μm Wavelength," *Electron. Lett.* **32**(22):2076–2078, 1996.
19. L. F. Tiemeijer, G. N. van den Hoven, P. J. A. Thijs, T. van Dongen, J. J. M. Binsma, and E. J. Jansen, "1310-nm DBR-type MQW Gain-Clamped Semiconductor Optical Amplifiers with AM-CATV-Grade Linearity," *IEEE Photon. Technol. Lett.* **8**(11):1453–1455, 1996.
20. D. A. Francis, S. P. DiJaili, and J. D. Walker, "A Single-Chip Linear Optical Amplifier," In *Proc. Optical Fiber Communication Conference*, Anaheim, California, 2001. Paper PD13.
21. L. H. Spiekman, J. M. Wiesenfeld, A. H. Gnauck, L. D. Garrett, G. N. van den Hoven, T. van Dongen, M. J. H. Sander-Jochem, and J. J. M. Binsma, "Transmission of 8 DWDM Channels at 20 Gb/s over 160 km of Standard Fiber Using a Cascade of Semiconductor Optical Amplifiers," *IEEE Photon. Technol. Lett.* **12**(6):717–719, 2000.
22. V. G. Mutalik, G. van den Hoven, and L. Tiemeijer, "Analog Performance of 1310-nm Gain-Clamped Semiconductor Optical Amplifiers," In *Proc. Optical Fiber Communication Conference*, pages 266–267, Dallas, Texas, 1997.
23. B. Nyman, D. Favin, and G. Wolter, "Automated System for Measuring Polarization Dependent Loss," In *Proc. Optical Fiber Communication Conference*, pages 230–231, San Jose, California, 1994.
24. K. Morito, M. Ekawa, T. Watanabe, and Y. Kotaki, "High-Output-Power Polarization-Insensitive Semiconductor Optical Amplifier," *J. Lightwave Technol.* **21**(1):176–181, 2003.
25. S. Tanaka, S. Tomabechei, A. Uetake, M. Ekawa, and K. Morito, "Record High Saturation Output Power (+20 dBm) and Low NF (6.0 dB) Polarisation-Insensitive MQW-SOA Module," *Electron. Lett.* **42**(18):1059–1060, 2006.
26. K. Morito, S. Tanaka, S. Tomabechei, and A. Kuramata, "A Broad-Band MQW Semiconductor Optical Amplifier with High Saturation Output Power and Low Noise Figure," *IEEE Photon. Technol. Lett.* **17**(5):974–976, 2005.
27. K. Morito and S. Tanaka, "Record High Saturation Power (+22 dBm) and Low Noise Figure (5.7 dB) Polarization-Insensitive SOA Module," *IEEE Photon. Technol. Lett.* **17**(6):1298–1300, 2005.
28. P. W. Juodawlkis, J. J. Plant, L. J. Missaggia, K. E. Jensen, and F. J. O'Donnell, "Advances in 1.5- μm InGaAsP/InP Slab-Coupled Optical Waveguide Amplifiers (SCOWAs)," In *LEOS 2007. The 20th Annual Meeting of the IEEE*, pages 309–310, Oct. 2007.
29. W. Loh, J. J. Plant, F. J. O'Donnell, and P. W. Juodawlkis, "Noise Figure of a Packaged, High-Power Slab-Coupled Optical Waveguide Amplifier (SCOWA)," In *LEOS 2008. The 21st Annual Meeting of the IEEE*, pages 852–853, Nov. 2008.

30. T. Akiyama, M. Ekawa, M. Sugawara, K. Kawaguchi, H. Sudo, A. Kuramata, H. Ebe, and Y. Arakawa, "An Ultrawide-Band Semiconductor Optical Amplifier Having an Extremely High Penalty-Free Output Power of 23 dBm Achieved with Quantum Dots," *IEEE Photon. Technol. Lett.* **17**(8):1614–1616, 2005.
31. N. Yasuoka, K. Kawaguchi, H. Ebe, T. Akiyama, M. Ekawa, K. Morito, M. Sugawara, and Y. Arakawa, "Quantum-Dot Semiconductor Optical Amplifiers with Polarization-Independent Gains in 1.5- μ m Wavelength Bands," *IEEE Photon. Technol. Lett.* **20**(23):1908–1910, 2008.
32. P. L. Kuindersma, G. P. J. M. Cuijpers, J. G. L. Jennen, J. J. E. Reid, L. F. Teimeijer, H. de Waardt, and A. J. Boot, "10 Gbit/s RZ Transmission at 1309 nm Over 420 km Using a Chain of Multiple Quantum Well Semiconductor Optical Amplifier Modules at 38 km Intervals," In *Proc. European Conference on Optical Communications* 1996. Paper TuD.2.1.
33. K. D. LaViolette, "CTB Performance of Cascaded Externally Modulated and Directly Modulated CATV Transmitters," *IEEE Photon. Technol. Lett.* **8**(2):281–283, 1996.
34. Y. Awaji, J. Inoue, H. Sotobayashi, F. Kubota, and T. Ozeki, "Nonlinear Interchannel Cross Talk of Linear Optical Amplifier (LOA) in DWDM Applications," In *Optical Fiber Communications Conference*, 2003. OFC 2003, pages 441–443 vol. 2, Mar. 2003.
35. L. H. Spiekman, J. M. Wiesenfeld, A. H. Gnauck, L. D. Garrett, G. N. van den Hoven, T. van Dongen, M. J. H. Sander-Jochem, and J. J. M. Binsma, "8 $\times$ 10 Gb/s DWDM Transmission over 240 km of Standard Fiber Using a Cascade of Semiconductor Optical Amplifiers," *IEEE Photon. Technol. Lett.* **12**(8):1082–1084, 2000.
36. L. H. Spiekman, A. H. Gnauck, J. M. Wiesenfeld, and L. D. Garrett, "DWDM Transmission of Thirty Two 10 Gbit/s Channels through 160 km Link Using Semiconductor Optical Amplifiers," *Electron. Lett.* **36**(12):1046–1047, 2000.
37. Y. Sun, A. K. Srivastava, S. Banerjee, J. W. Sulhoff, R. Pan, K. Kantor, R. M. Jopson, and A. R. Chraplyvy, "Error-Free Transmission of 32 $\times$ 2.5 Gbit/s DWDM Channels over 125 km Using Cascaded In-Line Semiconductor Optical Amplifiers," *Electron. Lett.* **35**(21):1863–1865, 1999.
38. Y. Awaji, H. Sotobayashi, and F. Kubota, "Transmission of 80 Gb/s $\times$ 6 WDM over 100 km Using Linear Optical Amplifiers," *IEEE Photon. Technol. Lett.* **17**(3):699–701, 2005.
39. H. K. Kim and S. Chandrasekhar, "Reduction of Cross-Gain Modulation in the Semiconductor Optical Amplifier by Using Wavelength Modulated Signal," *IEEE Photon. Technol. Lett.* **12**(10):1412–1414, 2000.
40. A. K. Srivastava, S. Banerjee, B. R. Eichenbaum, C. Wolf, Y. Sun, J. W. Sulhoff, and A. R. Chraplyvy, "A Polarization Multiplexing Technique to Mitigate WDM Crosstalk in SOAs," *IEEE Photon. Technol. Lett.* **12**(10):1415–1416, 2000.
41. N. S. Bergano, F. W. Kerfoot, and C. R. Davidson, "Margin Measurements in Optical Amplifier Systems," *IEEE Photon. Technol. Lett.* **5**(3):304–306, 1993.
42. Z. Li, Y. Dong, J. Mo, Y. Wang, and C. Lu, "1050-km WDM Transmission of 8 $\times$ 10.709 Gb/s DPSK Signal Using Cascaded In-Line Semiconductor Optical Amplifier," *IEEE Photon. Technol. Lett.* **16**(7):1760–1762, 2004.
43. P. P. Iannone, K. C. Reichmann, and L. Spiekman, "In-Service Upgrade of an Amplified 130-km Metro CWDM Transmission System Using a Single LOA with 140-nm Bandwidth," In *Opt. Fiber Commun. Conf.*, vol. 2, pages 548–549. OSA, 2003. Paper ThQ3.
44. P. P. Iannone, H. H. Lee, K. C. Reichmann, X. Zhou, M. Du, B. Pálsdóttir, K. Feder, P. Westbrook, K. Brar, J. Mann, and L. Spiekman, "Four Extended-Reach TDM PONs Sharing a Bidirectional Hybrid CWDM Amplifier," *J. Lightwave Technol.* **26**(1):138–143, 2008.
45. T. Ismail, C. P. Liu, J. E. Mitchell, A. J. Seeds, X. Qian, A. Wonfor, R. V. Pent, and I. H. White, "Transmission of 37.6-GHz QPSK Wireless Data over 12.8-km Fiber with Remote Millimeter-Wave Local Oscillator Delivery Using a Bi-Directional SOA in a Full-Duplex System with 2.2-km CWDM Fiber Ring Architecture," *IEEE Photon. Technol. Lett.* **17**(9):1989–1991, 2005.
46. E. Almstrom, C. P. Larsen, L. Gillner, W. H. van Berlo, M. Gustavsson, and E. Berglind, "Experimental and Analytical Evaluation of Packaged 4 $\times$ 4 InGaAsP/InP Semiconductor Optical Amplifier Gate Switch Matrices for Optical Networks," *J. Lightwave Technol.* **14**(6):996–1004, 1996.
47. G. A. Fish, B. Mason, L. A. Coldren, and S. P. DenBaars, "Compact, 4 $\times$ 4 InGaAsP-InP Optical Crossconnect with a Scalable Architecture," *IEEE Photon. Technol. Lett.* **10**(9):1256–1258, 1998.
48. Y. Maeno, Y. Suemura, A. Tajima, and N. Henmi, "A 2.56-Tb/s Multiwavelength and Scalable Switch-Fabric for Fast Packet-Switching Networks," *IEEE Photon. Technol. Lett.* **10**(8):1180–1182, 1998.
49. A. Tajima, N. Kitamura, S. Takahashi, S. Kitamura, Y. Maeno, Y. Suemura, and N. Henmi, "10-Gb/s/Port Gated Divider Passive Combiner Optical Switch with Single-Mode to Multimode Combiner," *IEEE Photon. Technol. Lett.* **10**(1):162–164, 1998.

50. N. Kikuchi, Y. Shibata, H. Okamoto, Y. Kawaguchi, S. Oku, H. Ishii, Y. Yoshikuni, and Y. Tohmori, "Error-Free Signal Selection and High-Speed Channel Switching by Monolithically Integrated 64-channel WDM Channel Selector," *Electron. Lett.* **38**(15):823–824, 2002.
51. T. Sakamoto, A. Okada, M. Hirayama, Y. Sakai, O. Moriwaki, I. Ogawa, R. Sato, K. Noguchi, and M. Matsuoka, "Optical Packet Synchronizer Using Wavelength and Space Switching," *IEEE Photon. Technol. Lett.* **14**(9):1360–1362, 2002.
52. D. Chiaroni, "Packet Switching Matrix: A Key Element for the Backbone and the Metro," *IEEE J. Sel. Areas Commun.* **21**(7):1018–1025, 2003.
53. L. Dittmann, C. Devellder, D. Chiaroni, F. Neri, F. Callegati, W. Koerber, A. Stavdas, et al., "The European IST Project DAVID: A Viable Approach toward Optical Packet Switching," *IEEE J. Sel. Areas Commun.*, **21**(7):1026–1040, 2003.
54. U. Koren, B. I. Miller, M. G. Young, T. L. Koch, R. M. Jopson, A. H. Gnauck, J. D. Evankow, and M. Chien, "High-Frequency Modulation of Strained Layer Multiple Quantum Well Optical Amplifiers," *Electron. Lett.* **27**(1):62–64, 1991.
55. M. D. Feuer, J. M. Wiesenfeld, J. S. Perino, C. A. Burrus, G. Raybon, S. C. Shunk, and N. K. Dutta, "Single-Port Laser-Amplifier Modulators for Local Access," *IEEE Photon. Technol. Lett.* **8**(9):1175–1177, 1996.
56. K. Y. Cho, Y. Takuchima, and Y. C. Chung, "10-Gb/s Operation of RSOA for WDM PON," *IEEE Photon. Technol. Lett.* **20**(18):1533–1535, 2008.
57. M. Koga, N. Tokura, and K. Nawata, "Gain-Controlled All-Optical Inverter Switch in a Semiconductor Laser Amplifier," *Appl. Opt.* **27**(19):3964–3965, 1988.
58. B. Glance, J. M. Wiesenfeld, U. Koren, A. H. Gnauck, H. M. Presby, and A. Jourdan, "High-Performance Optical Wavelength Shifter," *Electron. Lett.* **28**(18):1714–1715, 1992.
59. J. M. Wiesenfeld and B. Glance, "Cascadability and Fanout of Semiconductor Optical Amplifier Wavelength Shifter," *IEEE Photon. Technol. Lett.* **4**(10):1168–1171, 1992.
60. G. Lenz, E. P. Ippen, J. M. Wiesenfeld, M. A. Newkirk, and U. Koren, "Femtosecond Dynamics of the Nonlinear Anisotropy in Polarization Insensitive Semiconductor Optical Amplifiers," *Appl. Phys. Lett.* **68**:2933–2935, 1996.
61. S. L. Danielsen, C. Joergensen, M. Vaa, B. Mikkelsen, K. E. Stubkjaer, P. Doussiere, F. Pommerau, L. Goldstein, R. Ngo, and M. Goix, "Bit Error Rate Assessment of 40 Gbit/s All Optical Polarization Independent Wavelength Converter," *Electron. Lett.* **32**(18):1688–1690, 1996.
62. T. Durhuus, B. Mikkelsen, C. Joergensen, S. L. Danielsen, and K. E. Stubkjaer, "All-Optical Wavelength Conversion by Semiconductor Optical Amplifiers," *J. Lightwave Technol.* **14**(6):942–954, 1996.
63. D. Marcenac and A. Mecozzi, "Switches and Frequency Converters Based on Crossgain Modulation in Semiconductor Optical Amplifiers," *IEEE Photon. Technol. Lett.* **9**(6):749–751, 1997.
64. J. M. Wiesenfeld, B. Glance, J. S. Perino, and A. H. Gnauck, "Wavelength Conversion at 10 Gb/s Using a Semiconductor Optical Amplifier," *IEEE Photon. Technol. Lett.* **5**(11):1300–1303, 1993.
65. R. J. Manning and D. A. O. Davies, "Three-Wavelength Device for All-Optical Signal Processing," *Opt. Lett.* **19**(12):889–891, 1994.
66. T. Akiyama, N. Hatori, Y. Nakata, H. Ebe, and M. Sugawara, "Pattern-Effect-Free Semiconductor Optical Amplifier Achieved Using Quantum Dots," *Electron. Lett.* **38**(19):1139–1140, 2002.
67. L. Zhang, I. Kang, A. Bhardwaj, N. Sauer, S. Cabot, J. Jaques, and D. T. Neilson, "Reduced Recovery Time Semiconductor Optical Amplifier Using p-Type-Doped Multiple Quantum Wells," *IEEE Photon. Technol. Lett.* **18**(22):2323–2325, 2006.
68. J. S. Perino, J. M. Wiesenfeld, and B. Glance, "Fiber Transmission of 10 Gbit/s Signals Following Wavelength Conversion Using a Travelling-Wave Semiconductor Optical Amplifier," *Electron. Lett.* **30**(3):256–258, 1994.
69. G. P. Agrawal and N. K. Dutta, *Long-Wavelength Semiconductor Lasers*. Van Nostrand Reinhold, New York, 1986.
70. J. Leuthold, D. M. Marom, S. Cabot, J. J. Jaques, R. Ryf, and C. R. Giles, "All-Optical Wavelength Conversion Using a Pulse Reformatting Optical Filter," *J. Lightwave Technol.* **22**(1):186–192, 2004.
71. A. D. Ellis, A. E. Kelly, D. Nessel, D. Pitcher, D. G. Moodie, and R. Kashyup, "Error Free 100 Gbit/s Wavelength Conversion Using Grating Assisted Cross Gain Modulation in 2 mm Long Semiconductor Amplifier," *Electron. Lett.* **34**(20):1958–1959, 1998.

72. Y. Liu, E. Tangdiongga, Z. Li, H. de Waardt, A. M. J. Koonen, G. D. Khoe, X. Shu, I. Bennion, and H. J. S. Dorren, "Error-Free 320-Gb/s All-Optical Wavelength Conversion Using a Single Semiconductor Optical Amplifier," *J. Lightwave Technol.* **25**(1):103–108, 2007.
73. T. Durhuus, C. Joergensen, B. Mikkelsen, R. J. S. Pedersen, and K. E. Stubkjaer, "All Optical Wavelength Conversion by SOAs in a Mach-Zehnder Configuration," *IEEE Photon. Technol. Lett.* **6**(1):53–55, 1994.
74. J. Leuthold, P. A. Besse, E. Gamper, M. Dulk, St. Fischer, and H. Melchior, "Cascadable Dual-Order Mode All-Optical Switch with Integrated Data- and Control Signal Separators," *Electron. Lett.* **34**(16):1598–1600, 1998.
75. B. Dagens, C. Janz, D. Leclerc, V. Verdrager, F. Poingt, I. Guillemot, F. Gaborit, and D. Ottenwalder, "Design Optimization of All-Active Mach-Zehnder Wavelength Converters," *IEEE Photon. Technol. Lett.* **11**(4):424–426, 1999.
76. S. Nakamura, Y. Ueno, K. Tajima, J. Sasaki, T. Sugimoto, T. Kato, T. Shimoda, M. Itoh, H. Hatakeyama, T. Tamanuki, and T. Sasaki, "Demultiplexing of 168-Gb/s Data Pulses with a Hybrid-Integrated Symmetric Mach-Zehnder All-Optical Switch," *IEEE Photon. Technol. Lett.* **12**(4):425–427, 2000.
77. N. S. Patel, K. L. Hall, and K. A. Rauschenbach, "40-Gbit/s Cascadable All-Optical Logic with an Ultrafast Nonlinear Interferometer," *Opt. Lett.* **21**(18):1466–1468, 1996.
78. J. Leuthold, C. H. Joyner, B. Mikkelsen, G. Raybon, J. L. Pleumeekers, B. I. Miller, K. Dreyer, and C. A. Burrus, "100 Gbit/s All-Optical Wavelength Conversion with Integrated SOA Delayed-Interference Configuration," *Electron. Lett.* **36**(13):1129–1130, 2000.
79. I. Kang, C. Dorner, L. Zhang, M. Rasras, L. Buhl, A. Bhardway, S. Cabot, et al., "Regenerative All Optical Wavelength Conversion of 40-Gb/s DPSK Signals Using a Semiconductor Optical Amplifier Mach-Zehnder Interferometer," In *Proc. European Conference on Optical Communications* **6**:29–30, 2005. Paper Th4.3.3.
80. G. P. Agrawal, "Population Pulsations and Nondegenerate Four-Wave Mixing in Semiconductor Lasers and Amplifiers," *J. Opt. Soc. Am. B* **5**(1):147–159, 1988.
81. J. Zhou, N. Park, J. W. Dawson, K. J. Vahala, M. A. Newkirk, and B. I. Miller, "Terahertz Four-Wave Mixing Spectroscopy for Study of Ultrafast Dynamics in a Semiconductor Optical Amplifier," *Appl. Phys. Lett.* **63**(9):1179–1181, 1993.
82. J. Zhou, N. Park, J. W. Dawson, K. J. Vahala, M. A. Newkirk, and B. I. Miller, "Efficiency of Broadband Four-Wave Mixing Wavelength Conversion Using Semiconductor Traveling-Wave Amplifiers," *IEEE Photon. Technol. Lett.* **6**(1):50–52, 1994.
83. A. Mecozzi, "Analytical Theory of Four-Wave Mixing in Semiconductor Amplifiers," *Opt. Lett.* **19**(12):892–894, 1994.
84. F. Girardin, J. Eckner, G. Guekos, R. Dall'Ara, A. Mecozzi, A. D'Ottavi, F. Martelli, S. Scotti, and P. Spano, "Low-Noise and Very High-Efficiency Fourwave Mixing in 1.5-mm-Long Semiconductor Optical Amplifiers," *IEEE Photon. Technol. Lett.* **9**(6):746–748, 1997.
85. A. E. Kelly, D. D. Marcenac, and D. Nasset, "40 Gbit/s Wavelength Conversion over 24.6 nm Using FWM in a Semiconductor Optical Amplifier with an Optimised MQW Active Region," *Electron. Lett.* **33**(25):2123–2124, 1997.
86. A. Mecozzi and J. M. Wiesenfeld, "The Roles of Semiconductor Optical Amplifiers in Optical Networks," *Opt. Photonics News* **12**(3):36–42, 2001.
87. R. M. Jopson and R. E. Tench, "Polarisation-Independent Phase Conjugation of Lightwave Signals," *Electron. Lett.* **29**(25):2216–2217, 1993.
88. G. Contestabile, A. D'Ottavi, F. Martelli, A. Mecozzi, P. Spano, and A. Tersigni, "Polarization-Independent Four-Wave Mixing in a Bidirectional Traveling wave Semiconductor Optical Amplifier," *Appl. Phys. Lett.* **75**(25):3914–3916, 1999.
89. R. Ludwig and G. Raybon, "BER Measurements of Frequency Converted Signals Using Four-Wave Mixing in a Semiconductor Laser Amplifier at 1, 2.5, 5 and 10 Gbit/s," *Electron. Lett.* **30**(4):338–339, 1994.
90. S. Diez, R. Ludwig, C. Schmidt, U. Feiste, and H. G. Weber, "160-Gb/s Optical Sampling by Gain-Transparent Four-Wave Mixing in a Semiconductor Optical Amplifier," *IEEE Photon. Technol. Lett.* **11**(11):1402–1404, 1999.
91. W. Pieper, C. Kurtzke, R. Schnabel, D. Breuer, R. Ludwig, K. Petermann, and H. G. Weber, "Nonlinearity-Insensitive Standard-Fibre Transmission Based on Optical-Phase Conjugation in a Semiconductor-Laser Amplifier," *Electron. Lett.* **30**(9):724–726, 1994.
92. D. D. Marcenac, D. Nasset, A. E. Kelly, M. Brierley, A. D. Ellis, D. G. Moodie, and C. W. Ford, "40 Gbit/s Transmission over 406 km of NDSF Using Mid-Span Spectral Inversion by Four-Wave-Mixing in a 2 mm Long Semiconductor Optical Amplifier," *Electron. Lett.* **33**(10):879–880, 1997.

93. M. L. Nielsen, D. J. Blumenthal, and J. Mork, "A Transfer Function Approach to the Small-Signal Response of Saturated Semiconductor Optical Amplifiers," *J. Lightwave Technol.* **18**(12):2151–2157, 2000.
94. O. Leclerc, B. Lavigne, E. Balmeffre, P. Brindel, L. Pierre, D. Rouvillain, and F. Segueineau, "Optical Regeneration at 40 Gb/s and Beyond," *J. Lightwave Technol.* **21**(11):2779–2790, 2003.
95. R. Nagarajan, M. Kato, V. G. Dominic, C. H. Joyner, Jr. R. P. Schneider, A. G. Dentai, T. Desikan, et al., "400 Gbit/s (10 Channel $\times$ 40 Gbit/s) DWDM Photonic Integrated Circuits," *Electron. Lett.* **41**(6):347–349, 2005.
96. V. Lal, M. Masanovic, D. Wolfson, G. Fish, C. Coldren, and D. J. Blumenthal, "Monolithic Widely Tunable Optical Packet Forwarding Chip in InP for All-Optical Label Switching with 40 Gbps Payloads and 10 Gbps Labels," In *Proc. European Conference on Optical Communications*, 6:25–26, 2005. Paper Th4.3.1.
97. P. Bernasconi, L. Zhang, W. Yang, N. Sauer, L. L. Buhl, J. H. Sinsky, I. Kang, S. Chandrasekhar, and D. T. Neilson, "Monolithically Integrated 40-Gb/s Switchable Wavelength Converter," *J. Lightwave Technol.* **24**(1):71–76, 2006.

This page intentionally left blank

OPTICAL TIME-DIVISION MULTIPLEXED COMMUNICATION NETWORKS

Peter J. Delfyett

*CREOL, The College of Optics and Photonics
University of Central Florida
Orlando, Florida*

20.1 GLOSSARY

Definitions

Bandwidth. A measure of the frequency spread of a signal, system, or information-carrying capacity.

Chirping. The time dependence of the instantaneous frequency of a signal.

Commutator/decommutator. Devices that assist in the sampling, multiplexing, and demultiplexing of time domain signals.

Homogeneous broadening. Physical mechanism that broadens the linewidth of a laser transition. The amount of broadening is exactly the same for all excited states.

Kerr effect. Dependence of a material's index of refraction on the square of an applied electric field.

Mode partition noise. Noise associated with mode competition in a multimode laser.

Multiplexing/demultiplexing. Process of combining and separating several independent signals that share a common communication channel.

Passband. Range of frequencies allowed to pass in a linear system.

Picosecond. One trillionth of a second.

Pockel's effect. Dependence of a material's index of refraction on an applied electric field.

Photon lifetime. Time associated with the decay in light intensity within an optical resonator.

***p-n* junction.** Region that joins two materials of opposite doping. This occurs when n-type and p-type materials are joined to form a continuous crystal.

Quantum-confined Stark effect (QCSE). Optical absorption induced by an applied electric field across a semiconductor quantum well.

Quantum well. Thin semiconductor layer sandwiched between material with a larger bandgap. The relevant dimension of the layer is on the order of 10 nm.

Sampling. The process of acquiring discrete values of a continuous signal.

Spontaneous emission. Energy decay mechanism to reduce the energy of excited states by the emission of light.

Spatial hole burning. The resultant nonuniform spatial distribution of optical gain in a material, owing to standing waves in an optical resonator.

Stimulated emission. Energy decay mechanism that is induced by the presence of light in matter to reduce the energy of excited states by the emission of light.

Terabit. Information equivalent to one trillion bits of information.

Abbreviations

ADC	analog to digital converter
APD	avalanche photodetector
CMI	code mark inversion
DFB	distributed feedback
DBR	distributed Bragg reflector
DS	digital signal
EDFA	erbium-doped fiber amplifier
FP	fabry-Perot
FDM	frequency division multiplexing
LED	light-emitting diode
NRZ	non-return-to-zero
OOK	on-off keying
OC-N	optical carrier (Nth level)
PCM	pulse code modulation
PLM	pulse length modulation
PAM	pulse amplitude modulation
PPM	pulse position modulation
RZ	return-to-zero
SONET	synchronous optical network
SDH	synchronous digital hierarchy
STS	synchronous transmission signal
SPE	synchronous payload envelope
TOAD	terahertz optical asymmetric demultiplexer
SLALOM	semiconductor laser amplifier loop optical mirror
UNI	unbalanced nonlinear interferometer
TDM	time-division multiplexing
TDMA	time domain multiple access
WDM	wavelength division multiplexing
VCO	voltage-controlled oscillator

Symbols

$W(\text{Hz})$	bandwidth of a signal in units of hertz
$x_s(t)$	sampled version of a continuous function of time
$x(t)$	continuous analog signal
T	period
f_s	sampling frequency
p_T	periodic sampling pulse train

δ	delta function
n	index of refraction
n	integer
$X(\omega)$	frequency spectrum of the signal $x(t)$
ω	angular frequency (rad/s)
N	number of levels in an analog-to-digital converter
B	number of bits representing N levels in an analog-to-digital converter
Λ	grating period
λ	wavelength
φ	phase shift
τ_D or τ_P	photon decay time or photon lifetime
τ_{RT}	round trip propagation time of an optical cavity
$R_{1,2}$	mirror reflectivities

20.2 INTRODUCTION

Information and data services, such as voice, data, video, and Internet, are integral parts of our everyday personal and business lives. In 2001, the worldwide telecommunication traffic began being dominated by the fast-growing Internet traffic, as compared to the slow-growing voice traffic that dominated networks in years prior. In 2008, the amount of data traffic being transported over fiber optic networks is about 100 Tb/s, up from 1 Tb/s in 2001. The growth of and demand for bandwidth and communication services is growing steadily at about 80 percent per year globally, and high-speed photonic technologies must evolve to meet this demand. Currently deployed high-speed optical data transmission links are based on electrical time domain multiplexed (ETDM) transmission, where only electronic-signal-processing is performed at the transmitter and receiver to multiplex and demultiplex multiple users to and from the communication channel, respectively. In contrast, optical time domain multiplexing (OTDM) uses optical components and subsystems to multiplex and demultiplex user information at the transmitter and receiver.

The general organization of this chapter is to initially provide the reader with a brief review of digital signals and sampling. Following this introduction, time-division multiplexing (TDM) is introduced. These two sections provide the reader with a firm understanding of the overall system perspective as to how these networks are designed. To provide an understanding of the current state of the art, a review of selected high-speed optical and optoelectronic device technologies is given. Before a final summary and outlook toward future directions, a specific ultrahigh-speed optical time-division optical link is discussed, to coalesce the concepts with the discussed device technology.

20.3 MULTIPLEXING AND DEMULTIPLEXING

Fundamental Concepts

Multiplexing is a technique used to combine the information of multiple communication sites or users over a common communication medium and sending the information over a communication channel where the bandwidth, or information carrying capacity, is shared between each user. In the case where the shared medium is time, a communication link is created by combining the information from several independent sources and transmitting the information from each source simultaneously. This is done by temporally interleaving the bits of each source of information so that each user sends data for a very short period of time over the communication channel. The user waits

until all other users transmit their data before the first user can transmit another bit of information. At the receiver end, the data is demultiplexed for an intended user, while the rest of the information continues to its destination.

Sampling

An important concept in time-division multiplexing is being able to have a simple and effective method for converting real-world information into a form that is suitable for transmission by light. This process of transforming real signals into a form that is suitable for reliable transmission requires one to sample the analog signal to be sent and digitize and convert the analog signal to a stream of 1s and 0s. This process is usually performed by a sample and hold circuit, followed by an analog to digital converter (ADC). In the following section the concepts of signal sampling and digitization is reviewed, with the motivation to convey the idea of the robustness of digital communications.

Sampling Theorem

Time-division multiplexing relies on the fact that an analog bandwidth-limited signal may be exactly specified by taking samples of the signal, if the samples are taken sufficiently frequently. Time multiplexing is achieved by interleaving the samples of the individual signals. To see that any signal can be exactly represented by a sequence of samples, an understanding of the sampling theorem is needed. The theorem states that a real-valued bandwidth-limited signal that has no spectral components above a frequency of W hertz is determined uniquely by its value at uniform intervals spaced no greater than $1/(2W)$ s apart. This means that an analog signal can be completely reconstructed from a set of uniformly spaced discrete samples in time. The signal samples $x_s(t)$ are usually obtained by multiplying the signal $x(t)$ by a train of narrow pulses $p_T(t)$, with a time period $T = 1/f_s \leq 1/2W$. The process of sampling can be mathematically represented as

$$\begin{aligned} x_s(t) &= x(t) \cdot p_T(t) \\ &= x(t) \cdot \sum_{n=-\infty}^{+\infty} \delta(t - nT) \\ &= \sum_{n=-\infty}^{+\infty} x(nT) \delta(t - nT) \end{aligned} \quad (1)$$

where it is assumed that the sampling pulses are ideal impulses and n is an integer. Defining the Fourier transform and its inverse as

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) \exp(-j\omega t) dt \quad (2)$$

and

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega) \exp(+j\omega t) d\omega \quad (3)$$

one can show that the spectrum $X_s(\omega)$ of the signal $x_s(t)$ is given by

$$\begin{aligned} X_s(\omega) &= \frac{1}{T} \sum P\left(\frac{2\pi n}{T}\right) \cdot X\left(\omega - \frac{2\pi n}{T}\right) \\ &= \frac{1}{T} P(\omega) \cdot \sum X\left(\omega - \frac{2\pi n}{T}\right) \end{aligned} \quad (4)$$

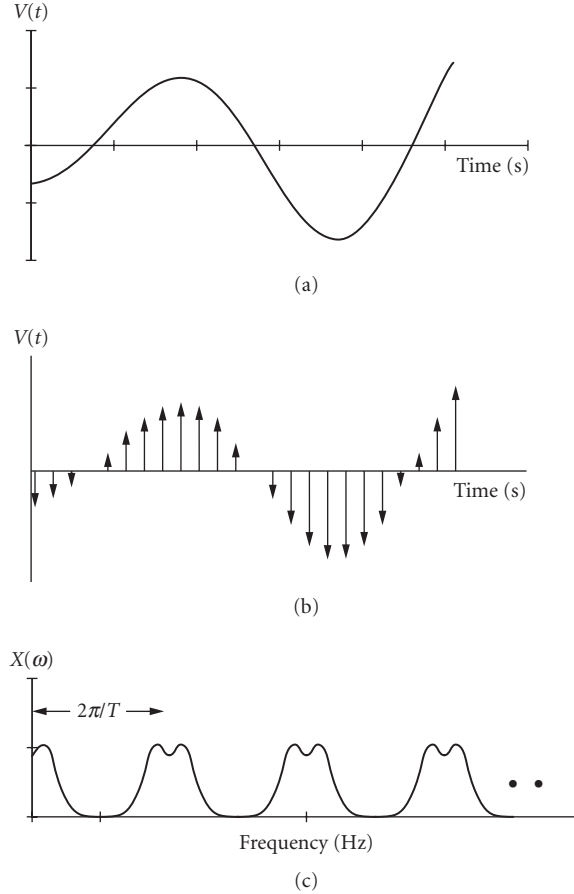


FIGURE 1 An analog bandwidth-limited signal (a), along with its sampled counterpart, sampled at a rate of ~ 8 times the Nyquist rate (b), and (c) frequency spectrum of a band-limited signal that has been sampled at a rate $T = 1/2W$, where W is the bandwidth of the signal.

In the case of the sampling pulses, p , being perfect delta functions, and given that the Fourier transform of $\delta(t)$ is 1, the signal spectrum is given by

$$X_s = \sum X\left(\omega - \frac{2\pi n}{T}\right) \quad (5)$$

This is represented pictorially in Fig. 1a to c. In Fig. 1a and b an analog signal and its sampled version, where the sample intervals is ~ 8 times the nominal sample rate of $1/(2W)$ are represented. From Fig. 1c it is clear that the spectrum of the signal is repeated in frequency every $2\pi/T$ Hz, if the sample rate T is $1/(2W)$. By employing (passing the signal through) an ideal rectangular low-pass filter, that is, a uniform (constant) passband with a sharp cutoff, centered at dc with a bandwidth of $2\pi/T$, the signal can be completely recovered. This filter characteristic implies an impulse response of

$$h(t) = 2W \frac{\sin(2\pi Wt)}{2\pi Wt} \quad (6)$$

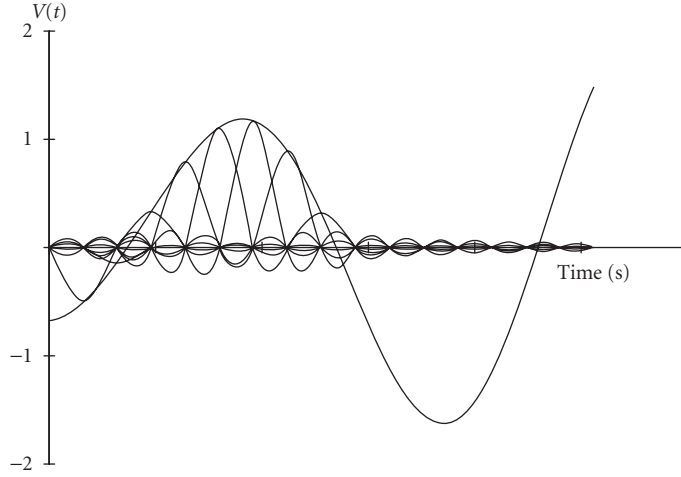


FIGURE 2 Temporal reconstruction of the sampled signal after passing the samples through a rectangular filter.

The reconstructed signal can now be given as

$$\begin{aligned}
 x(t) &= 2W \sum_{n=-\infty}^{+\infty} x(nT) \cdot \frac{\sin[2\pi W(t - nT)]}{2\pi W(t - nT)} \\
 &= x(t)/T \quad T = \frac{1}{2W}
 \end{aligned} \tag{7}$$

This reconstruction is shown in Fig. 2. It should be noted that the oscillating nature of the impulse response $h(t)$ interferes destructively with other sample responses, for times away from the centroid of each reconstructed sample.

Interleaving

The sampling principle can be exploited in time-division multiplexing by considering the ideal case of a single point-to-point link connecting N users to N other users over a single communication channel (see Fig. 3). As the rotary arm of the switch swings around, it samples each signal sequentially. The rotary switch at the receiving end is in synchronism with the switch at the sending end. The two switches make contact simultaneously at a similar number of contacts. With each revolution of the switch one sample is taken of each input signal and presented to the correspondingly numbered contact of the receiving end switch. The train of samples at Receiver 1, pass through a low-pass filter and at the filter output the original signal $m(t)$ appears reconstructed.

When the signals to be multiplexed vary rapidly in time, electronic switching systems are employed, as opposed to simple mechanical switches. The transmitter commutator samples and combines samples, while the receiver decommutator separates or demultiplexes samples belonging to individual signals so that these signals may be reconstructed.

The interleaving of the samples that allow multiplexing is shown in Fig. 4. For illustrative purposes, only two analog signals are considered. Both signals are repetitively sampled at a sample rate of T ; however, the instant at which the samples of each signal are taken are different. The input signal to Receiver 1 in Fig. 3, is the train of samples from Transmitter 1 and the input signal to Receiver 2 is

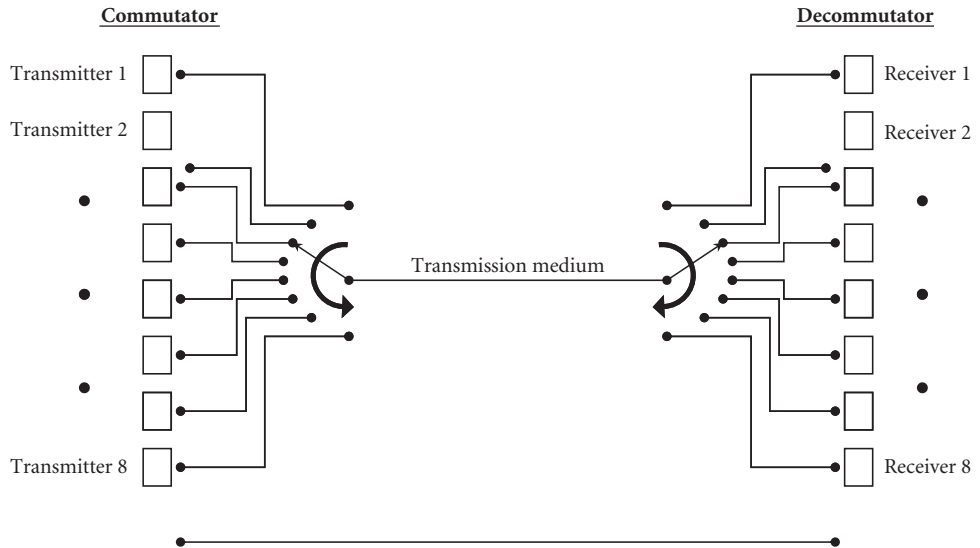


FIGURE 3 Illustration of a time multiplexer/demultiplexer based on simple mechanical switches, called commutators and decommutators.

the train of samples from Transmitter 2. The relative timing of the sampled signals of Transmitter 1 has been drawn to be exactly between the samples of Transmitter 2 for clarity; however, in practice, these samples would be separated by a smaller timing interval to accommodate additional temporally multiplexed signals.

Demultiplexing—Synchronization of Transmitter and Receiver

In any type of time-division multiplexing system, it is required that the sampling at the transmitter end and the demultiplexing at the receiver end be synchronized to each other. As an example, consider the diagram of the commutator of Fig. 3. When the transmitting multiplexer is set in a position that samples and transmits information from Transmitter 1, the receiving demultiplexer must be in a position to demultiplex and receive information that is intended for Receiver 1. To accomplish this timing synchronization, the receiver has a local clock signal that controls the timing of the commutator to switch from one time slot to the next. The clock signal may be a narrow band sinusoidal signal from which an appropriate clocking signal, with sufficiently fast rising edges of the appropriate signal strength, can be derived. The repetition rate of the clock in a simple configuration would then be equal to the sample rate of an individual channel times the number of channels being multiplexed, thereby assigning one time slot per clock cycle.

At the receiver end, the clock signal is required to keep the decommutator synchronized to the commutator, that is, running at the same rate. In addition, there must be additional timing information to provide agreement as to the relative positions, or phase of the commutator-decommutator pair, which ensures that information from Transmitter 1 is guaranteed to be received at the desired destination of Receiver 1. The time interval from the beginning of the time slot allocated to a particular channel, until the next recurrence of that particular time slot is commonly referred to as a *frame*. As a result, timing information is required at both the bit (time slot) and frame levels. A common arrangement in time-division-multiplexed systems is to allow for one or more time slots per frame to provide timing information, depending on the temporal duration of the transmitted frame. It should

be noted that there are a variety of methods for providing timing information, such as directly using a portion of the allocated bandwidth, as mentioned above, or alternatively, recovering a clock signal by deriving timing information directly from the transmitted data.

Digital Signals—Analog to Digital Conversion

The sampled signal, as shown in Fig. 4, represent the actual values of the analog signal at the sampling instants. In a practical communication system or in a realistic measurement setup, the received or measured values can never be absolutely correct, because of the noise introduced by the transmission channel or small inaccuracies impressed on the received data owing to the detection or measurement process. It turns out that it is sufficient to transmit and receive only the quantized values of the signal samples. The quantized values of sampled signals represented to the nearest digit, may be represented in a binary form or in any coded form using only 1s and 0s. For example, sampled values of a signal between 2.5 and 3.4 would be represented by the quantized value of 3, and could be represented as 11, using 2 bits (in base 2 arithmetic). This method of representing a sampled analog signal is known as *pulse code modulation*. An error is introduced on the signal by this quantization process. The magnitude of this error is given by

$$\varepsilon = \frac{0.4}{N} \quad (8)$$

where N is the number of levels determined by $N = 2^B$, and B is the B-bit binary code, for example, $B = 8$ for 8-bit words representing 256 levels. Thus one can minimize the error by increasing the number of levels, which is achieved by reducing the step size in the quantization process. It is interesting to note that using only 4 bit (16 levels), a maximum error of 2.5 percent is achieved, while increasing the number of bits to 8 (256 levels) gives a maximum error of 0.15 percent.

Optical Representation of Binary Digits and Line Coding

The binary digits can be represented and transmitted on an optical beam and passed through an optical fiber or transmitted in free space. The optical beam is modulated to form pulses to represent the sampled and digitized information. A family of four such representations is shown in Fig. 5. There are two particular forms of data transmission that are quite common in optical communications owing to the fact that their modulation formats occur naturally in both direct and externally modulated optical sources. These two formats are referred to as “non-return-to-zero” (NRZ), or “return-to-zero” (RZ). In addition to NRZ and RZ data formats, pulse-code-modulated data signals are transmitted in other codes that are designed to optimize the link performance, owing to channel constraints.

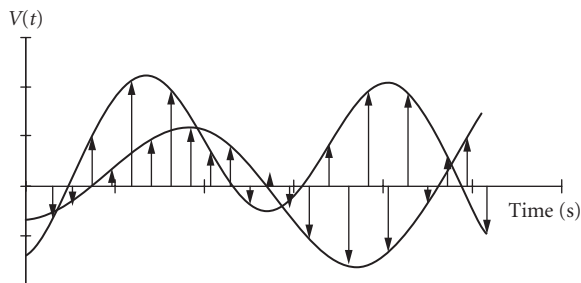


FIGURE 4 Two band-limited analog signals and their respective samples occurring at a rate of approximately six times the highest frequency, or three times the Nyquist rate.

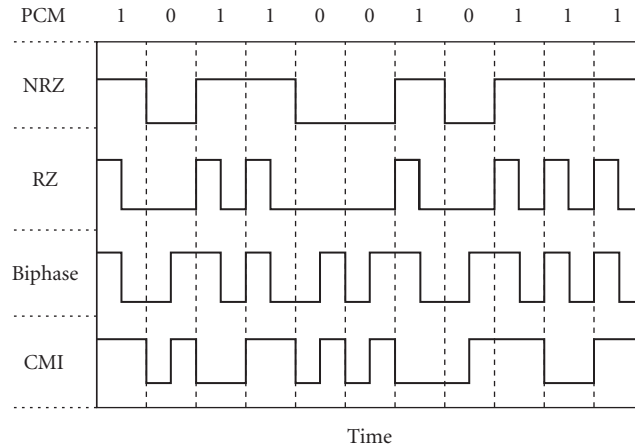


FIGURE 5 Line-coded representations of the pulse-code-modulated logic signal “10110010111.” NRZ: non-return to zero; RZ: return to zero; biphase, also commonly referred to as Manchester coding; CMI: code mark inversion.

Some important data transmission formats for optical time-division-multiplexed networks are code mark inversion (CMI), and Manchester coding or biphase coding. In CMI, the coded data has no transitions for logical 1 levels. Instead, the logic level alternates between a high and low level. For logical 0, on the other hand, there is always a transition from low to high at the middle of the bit interval. This transition for every logical 0 bit ensures proper timing recovery. For Manchester coding, logic 1 is represented by a return-to-zero pulse with a 50 percent duty cycle over the bit period (a half-cycle square wave), and logic 0 is represented by a similar return-to-zero waveform of “opposite phase,” hence the name biphase. The salient feature of both biphase and CMI coding is that their power spectra have significant energy at the bit rate, owing to the guarantee of a significant number of transitions from logic 1 to 0. This should be compared to the power spectra of RZ and NRZ data, which are shown in Fig. 6. The NRZ has no energy at the bit rate, while the RZ power spectrum does have energy at the bit rate, but the spectrum is also broad, having a width twice as large as NRZ.

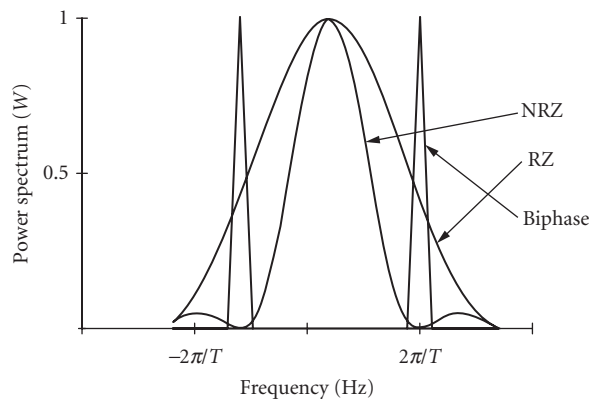


FIGURE 6 Power spectra of NRZ, RZ, and biphase line-coded data. Note the relative power at the bit rate.

The received data power spectra is important for TDM transmission links, where at the receiver end, a clock or synchronization signal is required to demultiplex the data. It is useful to be able to recover a clock or synchronization signal derived from the transmitted data, instead of using a portion of the channel bandwidth to send a clock signal. Therefore by choosing a transmission format with a large power spectral component at the transmitted bit rate, provides an easy method to recover a clock signal.

Timing Recovery

Time-division multiplexing and time-division multiple access networks inherently require timing signals to assist in demultiplexing individual signals from their multiplexed counterparts. One possible method is to utilize a portion of the communication bandwidth to transmit a timing signal. Technically, this is feasible, however, (1) this approach requires hardware dedicated to timing functions distributed at each network node that performs multiplexing and demultiplexing functions, and (2) network planners want to optimize the channel bandwidth without resorting to dedicating a portion of the channel bandwidth to timing functions. The desired approach is to derive a timing signal directly from the transmitted data. This allows the production of the required timing signals for multiplexing and demultiplexing without the need of using valuable channel bandwidth.

As suggested by Fig. 7, a simple method for recovering a timing signal from transmitted return-to-zero data is to use a bandpass filter to pass a portion of the power spectrum of the transmitted data. The filtered output from the tank circuit is a pure sinusoid that provides the timing information.

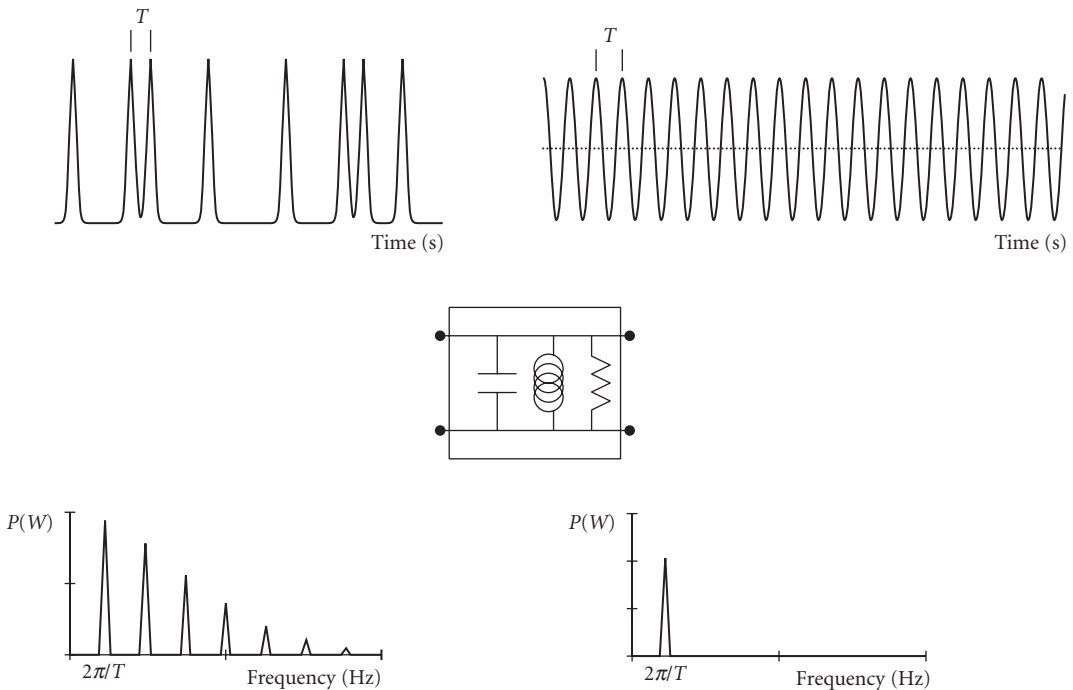


FIGURE 7 Principle of clock recovery using line filtering. Upper left: Input RZ data stream. Lower left: Power spectrum of a periodic RZ sequence. Center: Schematic of an electrical tank circuit for realizing a bandpass filter. Lower right: Power spectrum of the filtered signal. Upper right: Filtered time domain clock signal.

An important parameter to consider in line filtering is the quality factor, designated as the filter Q . Generally, the Q factor is defined as

$$Q = \frac{\omega_o}{\Delta\omega} \quad (9)$$

where ω_o is the resonant frequency and $\Delta\omega$ is the bandwidth of the filter. It should also be noted that the Q is a measure of the amount of energy stored in the bandpass filter, such that the output from the filter decays exponentially at a rate directly proportional to Q . In addition, for bandpass filters based on passive electrical circuits, the output peak signal is directly proportional to Q . These two important physical features of passive line filtering imply that the filter output will provide a large and stable timing signal if the Q factor is large. However, since Q is inversely proportional to the filter bandwidth, a large Q typically implies a small filter bandwidth. As a result, if the transmitter bit rate and the resonant frequency of the tank circuit do not coincide, the clock output could be zero. In addition the clock output is very sensitive to the frequency offset between the transmitter and resonant frequency. Therefore, line filtering can provide a large and stable clock signal for large filter Q , but the same filter will not perform well when the bit rate of the received signal has a large frequency variation. In TDM bit timing recovery, the ability to recover the clock of an input signal over a wide frequency range is called frequency acquisition or locking range, and the ability to tolerate timing jitter and a long interval of zero transitions is called frequency tracking or hold over time. Therefore, the trade-off exists between the locking range (low Q) and hold over time (large Q) in line filtering.

A second general scheme to realize timing recovery and overcome the drawbacks of line filtering using passive linear components is the use of a phase-locked loop in conjunction with a voltage-controlled oscillator (VCO) (see Fig. 8a). In this case, two signals are fed into the mixer. One signal is derived from the data, for example, from a line-filtered signal possessing energy at the bit rate, while the second signal is a sinusoid generated from the VCO. The mixer is used as a phase detector and produces a DC voltage that is applied to the VCO to adjust its frequency of operation. The overall design of the PLL is to adjust the voltage of the PLL to track the frequency and phase of the input data signal. Owing to the active components in the PLL, this approach for timing recovery can realize a broad locking range, low insertion loss, and good phase-tracking capabilities. It should be noted that

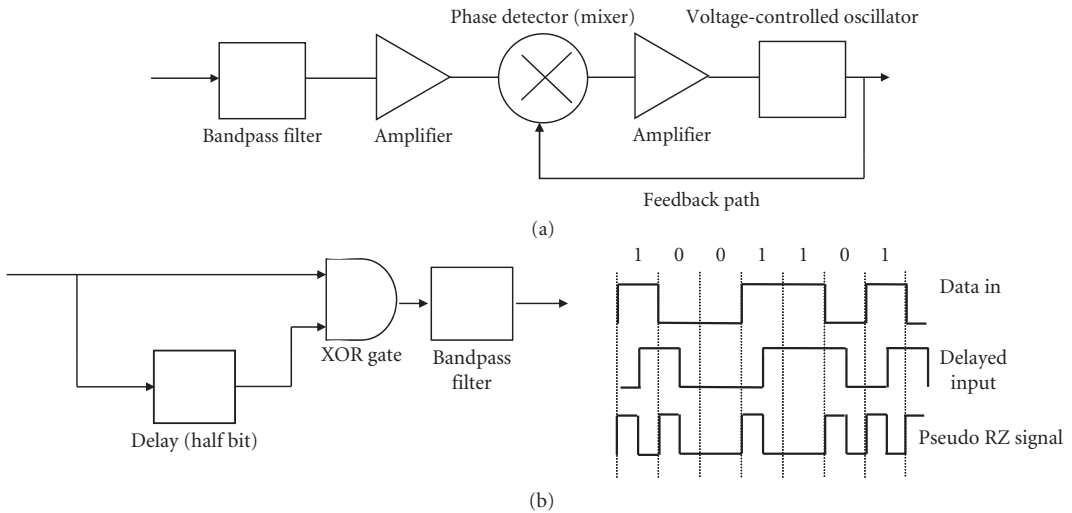


FIGURE 8 (a) Schematic diagram of a phase-locked loop using a mixer as a phase detector and a voltage-controlled oscillator to provide the clock signal that can track phase wander in the data stream. (b) Data format conversion between input NRZ data to RZ output data using an electronic logic gate. The subsequent RZ output is then suitable for use in a clock recovery device.

while the concepts for timing recovery described in this section were illustrated using techniques that are not directly applicable to ultrahigh-speed optical networking, the underlying principles will still hold for high-speed all-optical techniques. These approaches are discussed in more detail later in the section on device technology.

While both these techniques require the input data to be in the return-to-zero format, many data transmission links use nonreturn-to-zero line coding owing to its bandwidth efficiency. Unfortunately, in NRZ format there is no component in the power spectrum at the bit rate. As a result, some preprocessing of the input data signal is required before clock recovery can be performed. A simple method for achieving this is illustrated in Fig. 8b. The general concept is to present the data signal with a delayed version of the data at the input ports of a logic gate that performs the exclusive OR operation. The temporal delay, in this case, should be equal to 1/2 bit. The output of the XOR gate is a pseudo RZ data stream that can then be line filtered for clock recovery.

20.4 INTRODUCTION TO DEVICE TECHNOLOGY

Thus far, a general description of the concepts of digital communications and the salient features of TDM and TDMA have been presented. Next we address specific device technology that is employed in OTDM networks, for example, sources, modulators, receivers, clock recovery oscillators, demultiplexers, to provide an understanding of how and why specific device technology may be employed in a system to either optimize the network performance, to minimize cost, or to provide the maximum flexibility in supporting a wide variety of user applications.

Optical Time division Multiplexing—Serial versus Parallel

Optical time division multiplexing can generally be achieved by two main methods. The first method is referred as *parallel multiplexing*, while the second method classified as *serial multiplexing*. These two approaches are schematically illustrated in Fig. 9. The advantage of the parallel type of multiplexer is that it employs simple, linear passive optical components, not including the intensity modulator, and that the limitation in the transmission speed is not limited by the modulator or any other high-speed switching element. The drawback, however, is that the relative temporal delays between each channel must be accurately controlled and stabilized, which increases the complexity of this approach. Alternatively, the serial approach to multiplexing is simple to configure. In this approach a high-speed optical clock pulse train and modulation signal pulses are combined and introduced into an all-optical switch to create a modulated channel on the high bit-rate clock signal. Cascading this process allows all the channels to be independently modulated, with the requirement that the relative delay between each channel must be appropriately adjusted.

Device Technology—Transmitters

For advanced lightwave systems and networks, it is the semiconductor laser that dominates as the primary optical source that is used to generate the light that is modulated and transmitted as information. The reason for their dominance is that these devices are very small, typically a few hundred micrometers on a side, have excellent efficiency in converting electrons to photons, and are low cost. In addition, semiconductor diode lasers can generate optical signals at wavelengths of 1.3 and 1.55 μm . These wavelengths are important because they correspond to the spectral regions where optical signals experience minimal dispersion (spreading of the optical data bits) and minimal loss.

These devices initially evolved from simple light-emitting diodes, comprising a simple p - n junction, to Fabry-Perot (FP) semiconductor lasers, to distributed feedback (DFB) lasers and distributed Bragg reflector (DBR) lasers, and finally to mod-locked semiconductor diode lasers and optical fiber lasers. Below, a simple description of each of these devices is given, along with advantages and disadvantages that influence how these optical transmitter are deployed in current optical systems and networks.

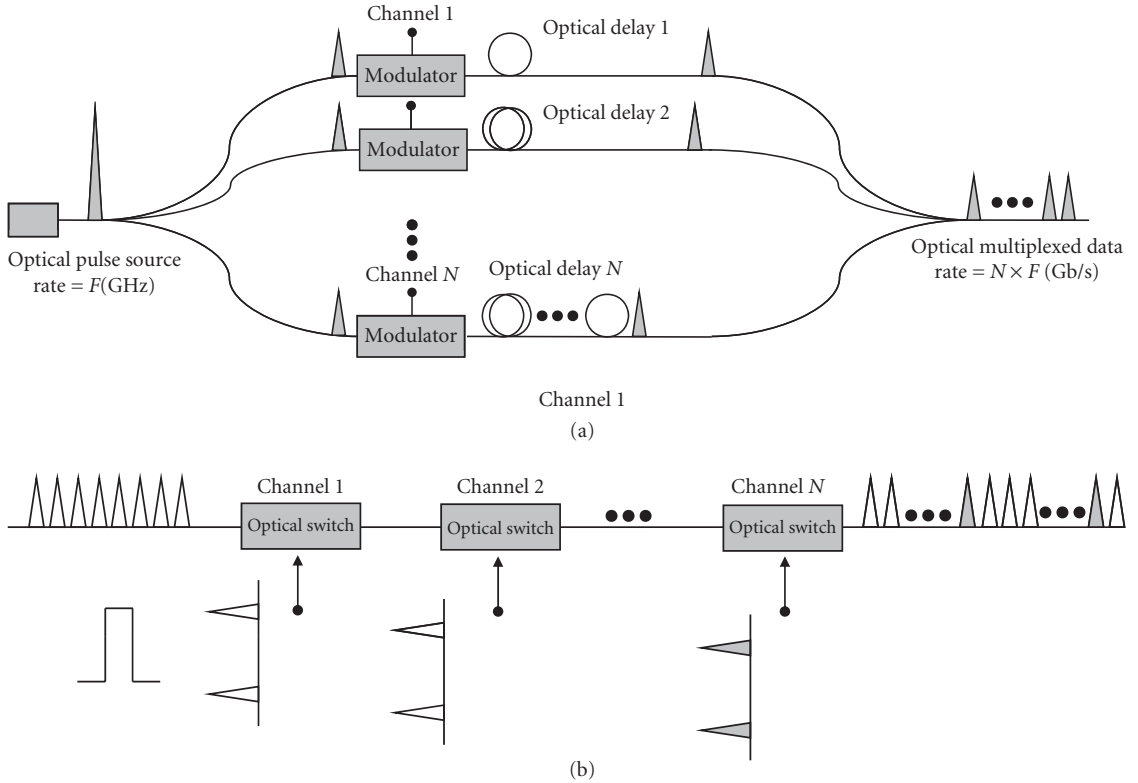


FIGURE 9 Schematic of optical time-division multiplexing for interleaving high-speed RZ optical pulses: (a) Parallel implementation and (b) serial implementation.

Fabry-Perot Semiconductor Lasers The Fabry-Perot semiconductor laser diode comprises a semiconductor p - n junction that is heavily doped and fabricated from a direct-gap semiconductor material. The injected current is sufficiently large to provide optical gain. The optical feedback is provided by mirrors, which are usually obtained by cleaving the semiconductor material along its crystal planes. The large refractive index difference between the crystal and the surrounding air causes the cleaved surfaces to act as reflectors. As a result, the semiconductor crystal acts as both the gain medium and as an optical resonator, or cavity (see Fig. 10). Provided that the gain coefficient is sufficiently large, the feedback transforms the device into an optical oscillator or laser diode.

Considering that the physical dimensions of the semiconductor diode laser are quite small, the short length of the diode forces the longitudinal mode spacing $c/2nL$ to be quite large. Here, c is the speed of light, L is the length of the diode chip, and n is the refractive index. Nevertheless, many of these modes can generally fit within the broad gain bandwidth allowed in a semiconductor diode laser. As an example, consider a FP laser diode operating at $1.3 \mu\text{m}$, fabricated from the InGaAsP material system. If $n = 3.5$ and $L = 400 \mu\text{m}$, the modes are spaced by 107 GHz, and corresponds to a wavelength spacing of 0.6 nm. In this device, the gain bandwidth can be 1.2 THz, corresponding to a wavelength spread of 7 nm, and as many as 11 modes can oscillate. Given that the mode spacing can be modified by cleaving the device so that only one axial mode exists within the gain bandwidth, the resulting device length would be approximately $36 \mu\text{m}$, which is difficult to achieve. It should be noted that if the bias current is increased well above threshold, the device can tend to oscillate on a single longitudinal mode. However for telecommunications, it is very desirable to directly modulate

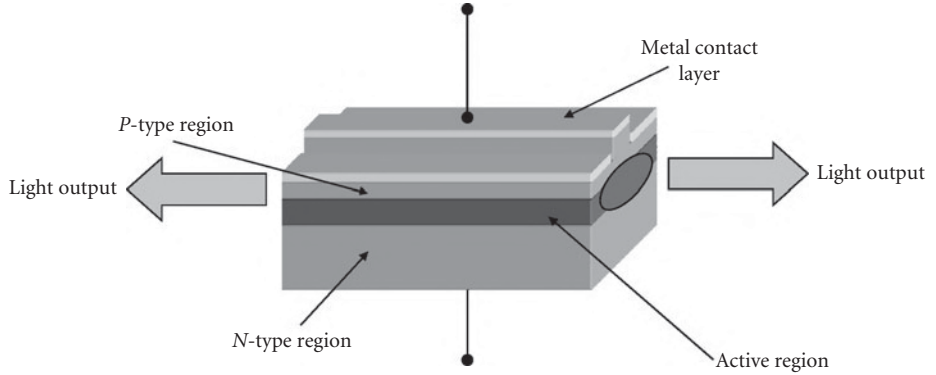


FIGURE 10 Schematic illustration of a simple Fabry-Perot semiconductor diode laser.

the laser, thus avoiding the cost of an external modulator. However, in the case of direct modulation, the output emission spectrum will be multimode, and as a result, effects of dispersion will broaden the optical data bits, and force the data rate to be reduced to avoid intersymbol interference. Given this effect, Fabry-Perot lasers tend to have a more limited use in longer optical links.

Distributed Feedback Lasers The effects of dispersion and the broad spectral emission from semiconductor LEDs and semiconductor Fabry-Perot laser diodes tend to reduce the overall optical data transmission rate. Thus, methods have been developed to design novel semiconductor laser structures that will only operate on a single longitudinal mode. This then will allow these devices to be directly modulated and allow for longer transmission paths since the overall spectral width is narrowed, and the effect of dispersion is minimized.

The preferred method of achieving single frequency operation from semiconductor diode lasers is to incorporate frequency-selective reflectors at both end of the diode chip or, alternately, fabricate the grating directly adjacent to the active layer. These two approaches result in devices referred to as distributed Bragg reflector lasers and distributed feedback lasers, respectively. In practice, it is easier to fabricate a single grating structure above the active layer, as opposed to two separate gratings at each end and as a result, the DFB laser has become the laser of choice for telecommunication applications. These devices operate with spectral widths on the order of a few megahertz, and have modulation bandwidths over 10 GHz. Clearly, the high modulation bandwidth and low spectral width make these devices well suited for direct modulation, or on-off keyed (OOK) optical networks. It should be noted that the narrow linewidth of a few megahertz is for the device operating in a continuous wave mode, while modulating the device will necessarily broaden the spectral width.

In DFB lasers, Bragg reflection gratings are employed along the longitudinal direction of the laser cavity and are used to suppress the lasing of additional longitudinal modes. As shown in Fig. 11a, a periodic structure, similar to a corrugated washboard, is fabricated over the active layer, where the periodic spacing is denoted as Λ . Owing to this periodic structure, both forward and backward traveling waves must interfere constructively with each other. In order to achieve this constructive interference between the forward and backward waves, the round trip phase change over one period should be $2\pi m$, where m is an integer and is called the order of the Bragg diffraction. With $m = 1$, the first-order Bragg wavelength, λ_B , is

$$2\pi = 2\Lambda(2\pi n/\lambda_B) \quad (10)$$

or

$$\lambda_B = 2\Lambda n \quad (11)$$

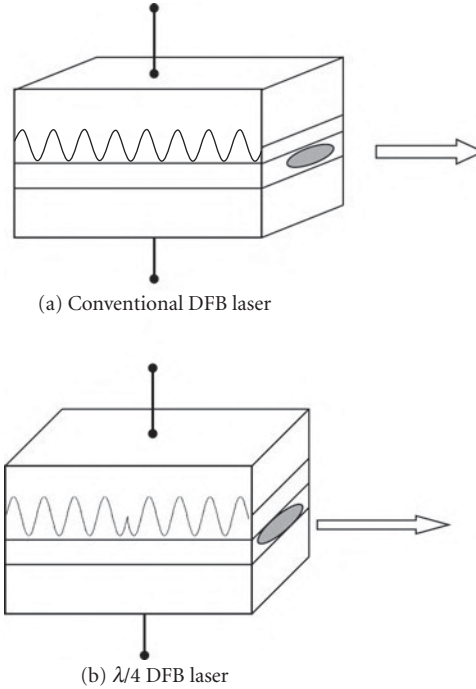


FIGURE 11 Schematic illustrations of distributed feedback lasers (DFB): (a) Conventional DFB and (b) Quarter-wave DFB, showing the discontinuity of the Bragg grating structure to achieve single wavelength operation.

where n is the refractive index of the semiconductor. Therefore, the period of the periodic structure determines the wavelength for the single mode output. In reality, a periodic DFB structure generates two main modes, symmetrically placed on either side of the Bragg wavelength λ_B . In order to suppress this dual frequency emission, and generate only one mode at the Bragg wavelength, a phase shift of $\lambda/4$ can be used to remove the symmetry. As shown in Fig. 11b, the periodic structure has a phase discontinuity of $\pi/2$ at the middle, which gives an equivalent $\lambda/4$ phase shift. Owing to the ability of the $\lambda/4$ DFB structure to generate a single frequency, narrow spectral linewidth, these devices are the preferred device for present telecommunications.

Mode-Locked Lasers Mode locking is a technique for obtaining very short bursts of light from lasers, and can be easily achieved employing both semiconductor and fiber gain media. As a result of mode locking, the light that is produced is automatically in a pulsed form that produces RZ data if passed through an external modulator being electrically driven with NRZ data. More importantly, *the temporal duration of the optical bits produced by mode locking is much shorter than the period of the driving signal*. To contrast this, consider a DFB laser whose light is externally modulated. In this case, the temporal duration of the optical bits will be equal to the temporal duration of the electrical pulses driving the external modulator. As a result, the maximum possible data transmission rate achievable from the DFB will be limited to the speed of the electronic driving signal. With mode locking; however, a low frequency electrical drive signal can be used to generate ultrashort optical bits. By following the light production with external modulation and optical bit interleaving, one can realize the ultimate in OTDM transmission rates. To show the difference between a mode-locked

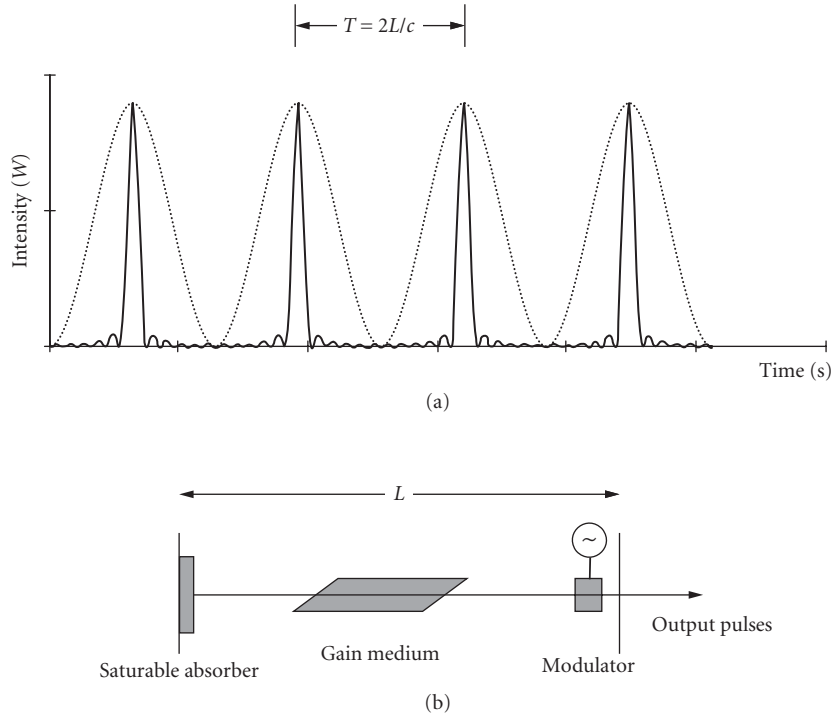


FIGURE 12 Optical intensity distribution of five coherent, phase-locked modes of a laser (a), and a schematic diagram of an external cavity mode-locked laser (b). Superimposed on the optical pulse train is a typical sinusoid that could be used to mode-lock the laser, showing that much shorter optical pulses can be obtained from a low-frequency signal.

pulse train and its drive, Fig. 12 plots a sinusoid and a mode-locked pulse train consisting of five locked optical modes.

To understand the process of mode locking, it should be recalled that a laser can oscillate on many longitudinal modes that are equally spaced by the longitudinal mode spacing, $c/(2nL)$. Normally, these modes oscillate independently; however, techniques can be employed to couple and lock their relative phases together. The modes can then be regarded as the components of a Fourier-series expansion of a periodic function of time of period $T = (2nL)/c$, which represents a periodic train of optical pulses. Consider, for example, a laser with multiple longitudinal modes separated by $c/2nL$. The output intensity of a perfectly mode-locked laser, as a function of time t , and axial position z , with M locked longitudinal modes, each with equal intensity, is given by

$$I(t, z) = M^2 |A|^2 \frac{\sin^2 c^2 [M(t - z/c)/T]}{\sin^2 [(t - z/c)T]} \quad (12)$$

where T is the periodicity of the optical pulses, and $\sin c(x)$ is $\sin(x)/x$. In practice, there are several methods to generate optical pulse trains by mode locking and generally fall into two categories: (1) active mode locking and (2) passive mode locking. In both cases, to lock the longitudinal modes in phase, the gain of the laser is allowed to increase above its threshold for a short duration, by opening and closing a shutter that is placed within the optical cavity. This allows a pulse of light to form. By allowing the light to propagate around the cavity and continually reopening and closing the shutter

at a rate inversely proportional to the round-trip time forms a stable, well-defined optical pulse is formed. If the shutter is realized by using an external modulator, the technique is referred to as *active mode locking*, where as if the shutter is realized by a device or material that is activated by the light intensity itself, the process is called *passive mode locking*. Both techniques can be used simultaneously and is referred to as *hybrid mode locking* (see Fig. 12b).

Direct and Indirect Modulation

To transmit information in OTDM networks, the light output of the laser source must be modulated in intensity. Depending on whether the output light is modulated by directly modulating the current source to the laser or whether the light is modulated external to, or after the light has been generated, the process of modulation can be classified as either (1) direct modulation or (2) indirect or external modulation (see Fig. 13a and b). With direct modulation, the light is directly modulated inside the light source, while external modulation, uses a separate external modulator placed after the laser source.

Direct modulation is used in many optical communication systems owing to its simple and cost-effective implementation. However, owing to the physics of laser action and the finite response of populating the lasing levels owing to current injection, the light output under direct modulation cannot respond to the input electrical signal instantaneously. Instead, there are turn-on delays and oscillations that occur when the modulating signal, which is used as the pumping current, has large and fast changes. As a result, direct modulation has several undesirable effects, such as frequency chirping and linewidth broadening. In frequency chirping, the spectrum of the output-generated light is time varying, that is, the wavelength and spectrum changes in time. This is owing to the fact

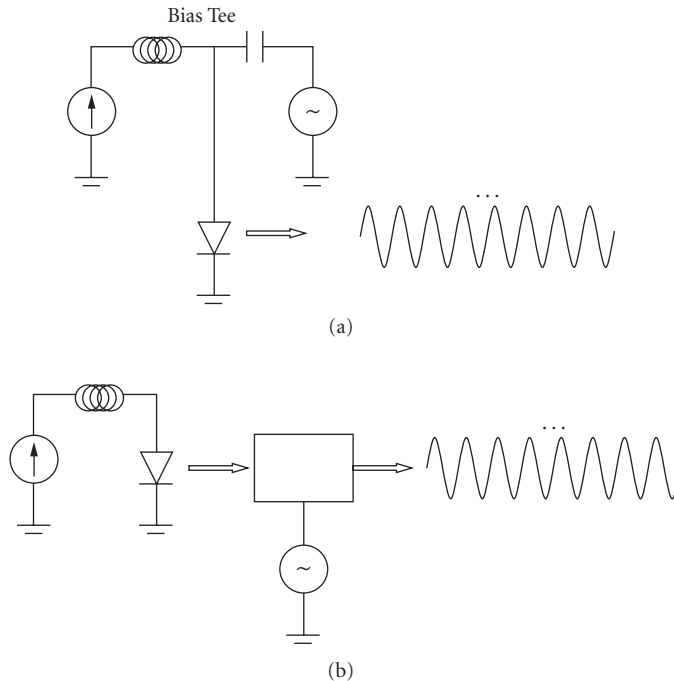


FIGURE 13 Illustrative example of direct modulation (a) and external modulation (b) of a laser diode.

that as the laser is turned on and off, the gain is changed from a very low value to a high value. Since the index of refraction of the laser diode is closely related to the optical gain of the device, as the gain changes, so does its index. It is this time varying refractive index that leads to frequency chirping, and is sometimes referred to as phase modulation.

External Modulation

To avoid the undesirable frequency chirping effects in DFB lasers and mode partition noise in FP lasers associated with direct modulation, external modulation provides an alternative approach to achieve light modulation with the added benefit of avoiding these detrimental effects. A typical external modulator consists of an optical waveguide in which the incident light propagates through and the refractive index or absorption of the medium is modulated by a signal that represents the data to be transmitted. Depending on the specific device, one can realize three basic types of external modulators: (1) electro-optic, (2) acousto-optic, and (3) electroabsorption. Generally, acousto-optic modulators respond slowly, on the order of several nanoseconds, and as a result are not used for external modulators in telecommunication applications. Electroabsorption (EA) modulators rely on the fact that the band edge of a semiconductor can be frequency shifted to realize an intensity modulation for a well-defined wavelength that is close to the band edge of the modulator. Linear frequency responses up to 50 GHz are possible; however, the fact that both the wavelength of the laser and the modulator must be accurately matched makes this approach more difficult to implement with individual devices. It should be noted, however, that EA modulators and semiconductor lasers can be integrated in the same device, helping to remove restrictions on matching the transmitter's and modulator's wavelength.

The typical desirable properties of an external modulator, from a communications perspective, are that these devices should possess a large modulation bandwidth, a large depth of modulation, a small insertion loss or loss of the signal light passing through the device, and a low electrical drive power. In addition, for some types of communication TDM links, a high degree of linearity between the drive signal and modulated light signal is required (typical for analog links), and an independence of input polarization (polarization diversity) is desired. Finally, low cost and small size of these devices are extremely useful for cost-effective and wide-area deployment.

Electro-Optic Modulators An electro-optic modulator can be a simple optical channel or waveguide that the light to be modulated propagates. The material that is chosen to realize the electro-optic modulator must possess an optical birefringence that can be controlled or adjusted by an external electrical field that is applied along or transverse to the direction of propagation of the light to be modulated. This birefringence means that the index of refraction is different for light that propagates in different directions in the crystal. If the input light has a well-defined polarization state, this light can be made to see, or experience, a different refractive index for different input polarization states. By adjusting the applied voltage to the electro-optic modulator, the polarization can be made to rotate, or the speed of the light can be slightly varied. This modification of the input light property can be used to realize a change of the output light intensity, by using a crossed polarizer or by interfering the modulated light with an exact copy of the unmodulated light. This can easily be achieved by using a waveguide interferometer, such as a Mach-Zehnder interferometer. If the refractive index is directly proportional to the applied electric field, the effect is referred to as the *Pockel's effect*.

Generally, for high-speed telecommunication applications, device designers employ the use of the electro-optic effect as a phase modulator in conjunction with an integrated Mach-Zehnder interferometer or an integrated directional coupler. Phase modulation (or delay/retardation modulation) does not effect the intensity of the input light beam. However, by incorporating a phase modulator in one branch of an interferometer, the resultant output light from the interferometer will be intensity modulated. Consider an integrated Mach-Zehnder interferometer in Fig. 14. If the waveguide divides the input optical power equally, the transmitted intensity is related to the output intensity by the well-known interferometer equation $I_o = I_i \cos^2(\varphi/2)$, where φ is the phase difference between the two light beams, and the transmittance function is defined as $I_o/I_i = \cos^2(\varphi/2)$.

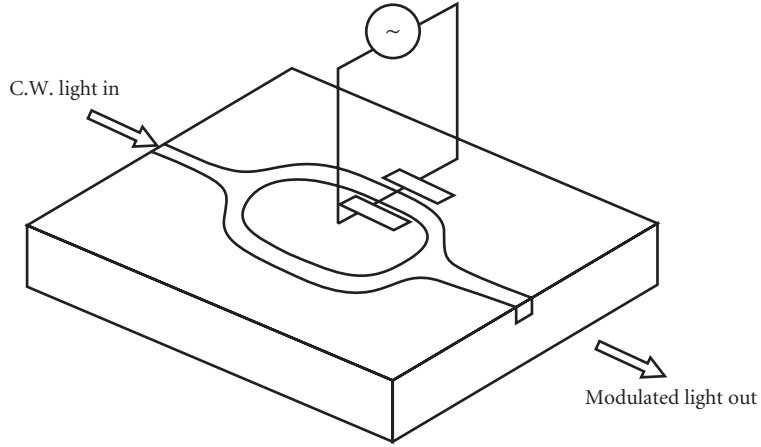


FIGURE 14 Illustration of an integrated lithium niobate Mach-Zehnder modulator.

Owing to the presence of the phase modulator in one of the interferometer arms, and with the phase being controlled by the applied voltage in accordance with a linear relation for the Pockel's effect, for example, $\varphi = \varphi_o - \pi V/V\pi$. In this equation, φ_o is determined by the optical path difference between the two beams and $V\pi$ is the voltage required to achieve a π phase shift between the two beams. The transmittance of the device, therefore becomes a function of the applied voltage V ,

$$T(V) = \cos^2(\varphi/2 - \pi V/2V\pi) \quad (13)$$

This function is plotted in Fig. 15 for an arbitrary value of φ_o . Commercially available integrated devices operate at speeds up to 40 GHz and are quite suitable for OTDM applications such as modulation and demultiplexing.

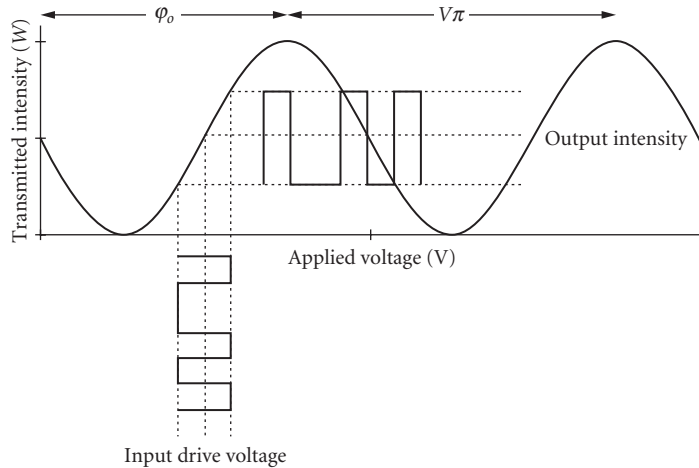


FIGURE 15 Input-output relations of an external modulator based on the Pockel's effect. Superimposed on the transfer function is a modulated drive signal and the resultant output intensity from the modulator.

Electroabsorption Modulators Electroabsorption modulators are intensity modulators that rely on the quantum confined Stark effect. In this device, thin layers of semiconductor material are grown on a semiconductor substrate to generate a multiplicity of semiconductor quantum wells or multiple quantum wells (MQW). For telecommunication applications, the semiconductor material family that is generally used is InGaAsP/InP. The number of quantum wells can vary, but is typically on the order of 10, with an overall device length of a few hundred micrometers. Owing to the dimensions of the thin layers, typically 100 Å or less, the electron and holes bind to form excitons. These excitons have sharp, and well-defined, optical absorption peaks that occur near the bandgap of the semiconductor material. By applying an electric field, or bias voltage, in a direction perpendicular to the quantum well layers, the relative position of the exciton absorption peak can be made to shift to longer wavelengths. As a result, an optical field that passes through these wells can be preferentially absorbed, if the polarization of the light field is parallel to the quantum well layers. Therefore, by modulating the bias voltage across the MQWs, the input light can be modulated. These devices can theoretically possess modulation speeds as high as 50 GHz, with contrasts approaching 50 dB. A typical device schematic and absorption curve is shown in Fig. 16a and b.

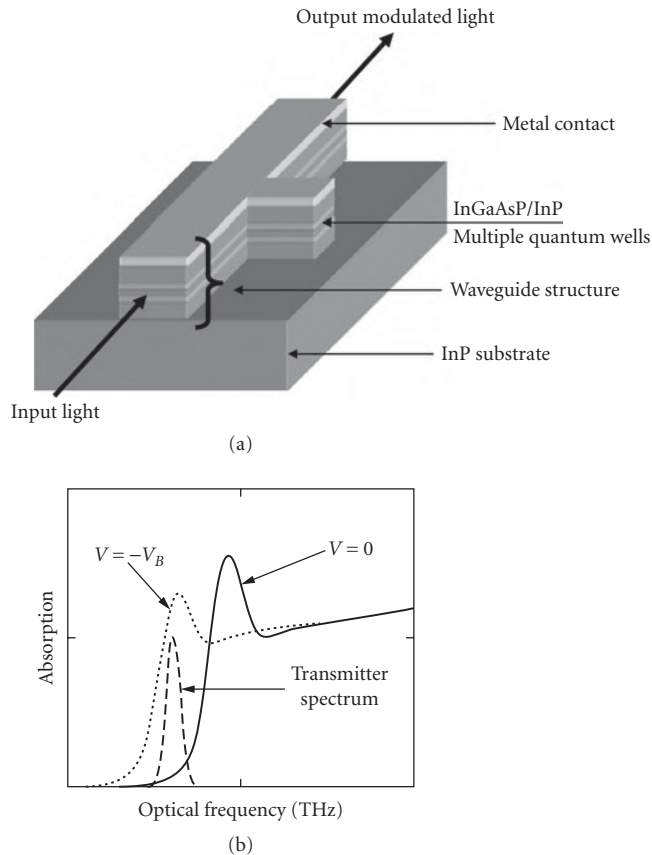


FIGURE 16 (a) Schematic diagram of an electro-absorption modulator. Light propagation occurs along the fabricated waveguide structure, in the plane of the semiconductor multiple quantum wells. (b) Typical absorption spectrum of a multiple quantum well stack under reverse bias and zero bias. Superimposed is a spectrum of a laser transmitter, showing how the shift in the absorption edge can either allow passage or attenuate the transmitted light.

Optical Clock Recovery

In time-division-multiplexed and multiple-access networks, it is necessary to regenerate a timing signal to be used for demultiplexing. Above, a general discussion of clock extraction was given, and in this section, an extension to those concepts are outlined for clock recovery in the optical domain. As in the conventional approaches to clock recovery, optical clock extraction has three general approaches: (1) the optical tank circuit, (2) high-speed phase-locked loops, and (3) injection locking of pulsed optical oscillators. The optical tank circuit can be easily realized by using a simple Fabry-Perot cavity. For clock extraction, the length L of the cavity must be related to the optical transmission bit rate. For example, if the input optical bit rate is 10 Gb/s, the effective length of the optical tank cavity is 15 mm. While the concept of the optical tank circuit is intuitively pleasing, since it has many of the same features as electrical tank circuits, that is, a cavity Q and its associated decay time. In the case of a simple Fabry-Perot cavity as the optical tank circuit, the optical decay time, or photon lifetime, is given by

$$\tau_D = \frac{\tau_{RT}}{1 - R_1 R_2} \quad (14)$$

where τ_{RT} is the round-trip time given as $2L/c$, and R_1 and R_2 are the reflection coefficients of the cavity mirrors. One major difference between the optical tank circuit and its electrical counterpart is that the output of the optical tank circuit never exceeds the input optical intensity (see Fig. 17a).

A second technique that builds on the concept of the optical tank is optical injection seeding or injection locking. In this technique, the optical data bits are injected into a nonlinear device such as a passively mode-locked semiconductor laser diode (see Fig. 17b). The key difference between this approach and the optical tank circuit approach is that the injection-locking technique has internal gain to compensate for the finite photon lifetime, or decay, of the empty cavity. In addition to the gain, the cavity also contains a nonlinear element, for example, a saturable absorber to initiate and sustain pulsed operation. Another important characteristic of the injection-locking technique, using passively mode-locked laser diodes is that clock extraction can be prescaled, that is, a clock signal can

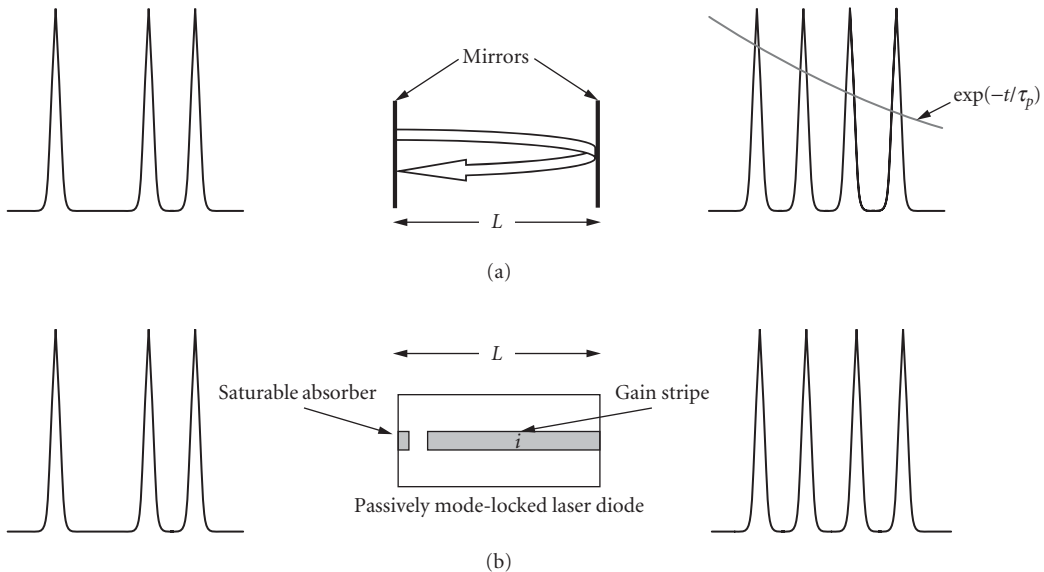


FIGURE 17 All optical clock recovery based on optical injection of (a) an optical tank circuit (Fabry-Perot cavity), and (b) a mode-locked semiconductor diode laser.

be obtained at bit rates exactly equal to the input data bit rate or at harmonics or subharmonics of the input bit rate. In this case of generating a prescaled clock signal at a subharmonic of the input data stream, the resultant signal can directly be used for demultiplexing, without any addition signal processing.

The operation of the injection-seeded optical clock is as follows. The passively mode-locked laser produces optical pulses at its natural rate, which is proportional to the longitudinal mode spacing of the device cavity, $c/(2L)$. Optical data bits from the transmitter are injected into the mode-locked laser, where the data transmission rate is generally a harmonic of the clock rate. This criteria immediately provides the prescaling required for demultiplexing. The injected optical bits serve as a seeding mechanism to allow the clock to build up pulses from the injected optical bits. As the injected optical bits and the internal clock pulse compete for gain, the continuous injection of optical bits force the internal clock pulse to evolve and shift in time to produce pulses that are synchronized with the input data. It should be noted that it is not necessary for the input optical bit rate to be equal to or greater than the nominal pulse rate of the clock, for example, the input data rate can be lower than the nominal bit rate of the clock. This is analogous to the transmitter sending data with primarily 0s, with logic 1 pulses occurring infrequently. The physical operating mechanism can also be understood by examining the operation in the frequency domain. From a frequency domain perspective, since the injected optical data bits are injected at a well-defined bit rate, the optical spectrum has a series of discrete line spectra centered around the laser emission wavelength, and separated in frequency by the bit rate. Since the optical clock emits a periodic train of optical pulses, its optical spectra is also a series of discrete line spectra, separated by the clock-repetition frequency. If the line spectra of the injected data bits fall within optical gain bandwidth of the optical clock, the injected line spectra will serve as seeding signals to force the optical clock to emit line spectra similar to the injected signals. Since the injected data bits are repetitively pulsed, the relative phase relation between the discrete line spectra have the proper phase relation to force the clock to emit synchronously with the injected data.

All-Optical Switching for Demultiplexing

In an all-optical switch, light controls light with the aid of a nonlinear optical material. It should be noted here that all materials will exhibit a nonlinear optical response, but the strength of the response will vary widely depending on the specific material. One important effect to realize an all-optical switch is the optical Kerr effect, where the refractive index of a medium is proportional to the square of the incident electric field. Since light is inducing the nonlinearity, or in other words, providing the incident electric field, the refractive index becomes proportional to the light intensity. Since the intensity of a light beam can change the refractive index, the speed of a second weaker beam can be modified owing to the presence of the intense beam. This effect is used extensively in combination with an optical interferometer to realize all-optical switching (see section on electro-optic modulation using a Mach-Zehnder interferometer). Consider, for example, a Mach-Zehnder interferometer that includes a nonlinear optical material that possess the optical Kerr effect (see Fig. 18). If data to be demultiplexed is injected into the interferometer, the relative phase delay in each arm can be adjusted so that the entire injected data signal is present only at one output port. If an intense optical control beam is injected into the nonlinear optical medium, synchronized with a single data bit passing through the nonlinear medium, that bit can be slowed down, such that destructive interference occurs at the original output port and constructive interference occurs at the secondary output port. In this case, the single bit has been switched out of the interferometer, while all other bits are transmitted.

Optical switches have been realized using optical fiber in the form of a Sagnac interferometer, and the fiber itself is used as the nonlinear medium. These devices are usually referred to as nonlinear optical loop mirrors (NOLM). Other versions of all-optical switches may use semiconductor optical amplifiers as the nonlinear optical element. In this case, it is the change in gain induced by the control pulse that changes the refractive index owing to the Kramers-Kronig relations. Devices such as these are referred to as a terahertz asymmetric optical demultiplexers (TOAD), semiconductor laser amplifier loop optical mirrors (SLALOM), and unbalanced nonlinear interferometers (UNI).

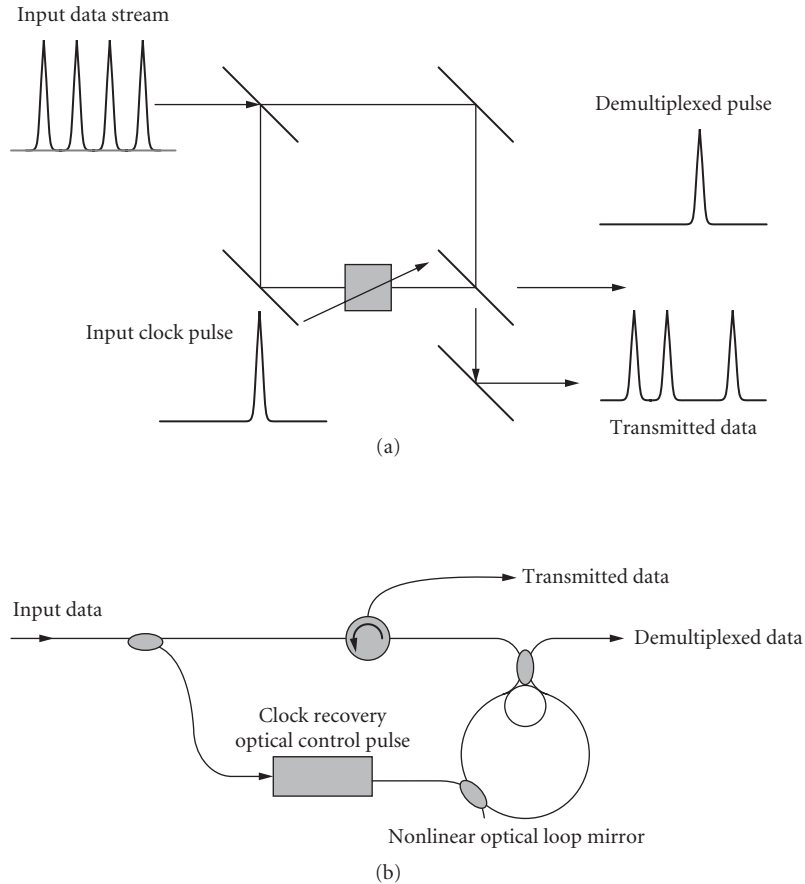


FIGURE 18 Schematic diagram of an all optical switch. (a) Simple configuration based on a Mach-Zehnder interferometer and a separate nonlinear material activated by an independent control pulse. (b) An optical fiber implementation of an all optical switch. This implementation relies on the inherent nonlinearity of the fiber that is induced by an independent control pulse.

Ultrahigh-Speed Optical Time-Division-Multiplexed Optical Link—A Tutorial Example

To demonstrate how the ultrafast device technology can realize state-of-the-art performance in optical time domain multiplexed systems, the following is an example that incorporates the device technology discussed in this chapter. Figure 19 shows a schematic illustration of an ultrahigh-speed OTDM transmission experiment that was used by a collaborative research team of scientists at the Heinrich Hertz Institute in Germany and Fujitsu Laboratories in Japan, to demonstrate 2.56-Tb/s transmission over an optical fiber span of 160 km. The transmitter comprised a 10-GHz mode-locked laser operating at 1550-nm wavelength, generating pulses of 0.42 ps. The pulse train was modulated and temporally multiplexed in two multiplexing stages to realize a maximum data transmission rate of 2.56 Tb/s. This signal is transmitted through 160 km of dispersion-managed optical fiber. At the receiver, the data generates a timing reference signal that drives a mode-locked fiber

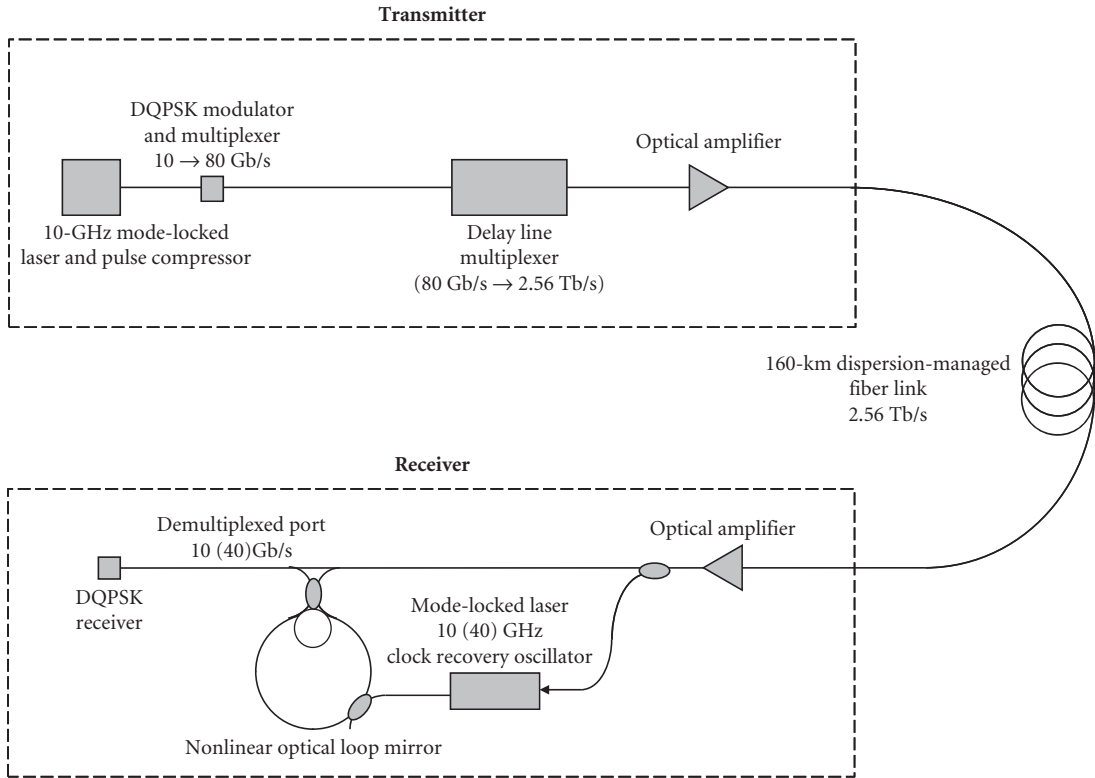


FIGURE 19 Schematic diagram of a 640-Gb/s optical fiber link, using all the critical components described in this chapter, for example, mode-locked laser, high-speed modulator, temporal interleaver, optical clock recovery, all optical switching for demultiplexing and an optical photoreceiver.

laser that is used as an optical gate in a nonlinear optical loop mirror. The resulting demultiplexed data is then detected using a DQPSK photodetection scheme. The robustness of this experiment is quantified by the ratio of the number of errors received to the data pulse period. Error-free operation is defined as less than 10^{-9} . The system performed error free with a received optical power of less than 0.1 mW.

20.5 SUMMARY AND FUTURE OUTLOOK

This chapter reviewed the fundamental basics of optical time-division-multiplexed communication networks, starting from an elementary perspective of digital sampling. Given this as an underlying background, specific device technology was introduced to show how the system functionality can be realized using ultrahigh-speed optics and photonic technologies. Finally, as an example of how these system and device technologies are incorporated into a functioning ultrahigh-speed optical time-division-multiplexed system, a 2.56-Tb/s link was discussed.

In the introduction, the difference between ETDM and OTDM was described. To have an idea of what the future may provide, it should be noted that 100-Gb/s ETDM are now being currently tested in laboratories, while the same data rates were investigated using OTDM techniques nearly

a decade earlier. As electronic technology continues to improve in speed, OTDM techniques are replaced by their electronic counterparts in commercially deployed systems. As a result, we have seen OTDM transmission technology pushing the boundaries of ultrahigh-speed data transmission in optical fibers, thus providing a roadmap for the development of commercial ETDM transmission over optical fiber. As we have shown, OTDM techniques are demonstrating transmission rates in excess of 2.5 Tb/s today, and perhaps suggest similar data rates for ETDM fiber optic transmission in the next decade.

20.6 BIBLIOGRAPHY

- Das, J., "Fundamentals of Digital Communication," in Bishnu P. Pal, (ed.), Section 18 *Fundamentals of Fiber Optics in Telecommunication and Sensor Systems*, Wiley Eastern, New Delhi, 1992, pp. 7-415-7-451.
- Kawanishi, S., "Ultrahigh-Speed Optical Time-Division-Multiplexed Transmission Technology Based on Optical Signal Processing," *J. Quant. Electron. IEEE* **34**(11):2064-2078 (1998).
- Saruwatari, M., "All Optical Time Division Multiplexing Technology," in N. Grote and H. Venghaus, (eds.), *Fiber Optic Communication Devices*, Springer-Verlag, Berlin 2001.
- Schuh, K. and E. Lach, E., "High-Bit-Rate ETDM Transmission Systems," in I. P. Kaminow, T. Li, and A. E. Willner (eds.), *Optical Fiber Telecommunications VB*, Elsevier-Academic Press, New York, 2008, pp. 179-200.
- Weber, H. G., Ferber, S., Kroh, M., Schmidt-Langhorst, C., Ludwig, R., Marembert, V., Boerner, C., Futami, F., Watanabe, S., and C. Schubert, C., "Single Channel 1.28 Tbit/s and 2.56 Tbit/s DQPSK Transmission," *Electron. Lett.* **42**:178-179 (2006).
- Weber, H. G. and Ludwig, R., "Ultra-High-Speed OTDM Transmission Technology," in I. P. Kaminow, T. Li, and A. E. Willner (eds.), *Optical Fiber Telecommunications VB*, Elsevier-Academic Press, New York, 2008, pp. 201-232.
- Weber, H. G. and Nakazawa, N., (eds.), *Optical and Fiber Communication Reports 3*, Springer Science + Business Media, LLC, 2007.

This page intentionally left blank

WDM FIBER-OPTIC COMMUNICATION NETWORKS

Alan E. Willner

*University of Southern California
Los Angeles, California*

Changyuan Yu

*National University of Singapore, and
A *STAR Institute for Infocomm Research
Singapore*

Zhongqi Pan

*University of Louisiana at Lafayette
Lafayette, Louisiana*

Yong Xie

*Texas Instruments Inc.
Dallas, Texas*

21.1 INTRODUCTION

The progress in optical communications over the past 30 years has been astounding. It has experienced many revolutionary changes since the days of short-distance multimode transmission at $0.8\text{ }\mu\text{m}$.¹ In 1980, AT&T could transmit 672 two-way conversations along a pair of optical fibers.² In 1994, an AT&T network connecting Florida with the Virgin Islands was able to carry 320,000 two-way conversations along two pairs of optical fibers. The major explosion came after the maturity of fiber amplifiers and wavelength-division multiplexing (WDM) technologies. By 2003, the transoceanic system of Tyco Telecommunications is able to transmit 128 wavelengths per fiber pair at 10 Gb/s/wavelength with the total capacity 10 Tb (eight fiber pairs)—a capability of transmitting more than 100 million simultaneous voice circuits on a eight-fiber pair cable. In experiments, a recent notable report demonstrated a record of 25.6-Tb/s WDM transmission using 160 channels within 8000 GHz fiber bandwidth.³

High-speed single-channel systems may offer compact optical systems with a minimized footprint and a maximized cost-per-bit efficiency.⁴ Transmission of single-channel 1.28 Tb/s signal over 70 km was achieved by traditional return-to-zero (RZ) modulation format using optical time-division multiplexing (OTDM) and polarization multiplexing.⁵ Although single-channel results are quite impressive, they have

two disadvantages: (1) they make use of only a very small fraction of the enormous bandwidth available in an optical fiber, and (2) they connect two distinct end points, not allowing for a multiuser environment. Since the required rates of data transmission among many users have been increasing at an impressive pace for the past several years, it is a highly desirable objective to connect many users with a high-bandwidth optical communication system.

WDM-related technologies have been growing rapidly and clearly dominated the research field and the telecommunications market. For instance, the new fiber bands (S and L) are being opened up, new modulation schemes are being deployed, unprecedented bit-rate/wavelength (≥ 40 Gb/s) is being carried over all-optical distance. Over 100 WDM channels are being simultaneously transmitted over ultralong distances.⁶⁻⁹ It is difficult to overstate the impact of fiber amplifiers and WDM technologies in both generating and supporting the telecommunications revolution during last 10 years.

Due to its high capacity and performance, optical fiber communications have already replaced many conventional communication systems in point-to-point transmission and networks. Driven by the rising capacity demand and the need to reduce cost, in addition to the fixed WDM transmission links with point-to-point configuration, the next step of the network migration is to insert the reconfigurable optical add/drop multiplexers (OADMs) in the links and network nodes. An automatically switching network can support the flexibility of the transmission. Hence, the necessary bandwidth can be provided to the customers on demand.^{10,11}

This chapter intends to review the state-of-the-art WDM technologies and reflect the tremendous progress over the past few years.

Fiber Bandwidth

The driving factor for the use of multichannel optical systems is the abundant bandwidth available in the optical fiber. The attenuation curve as a function of optical carrier wavelength is shown in Fig. 1.¹² There are two low-loss windows: one near 1.3 μm and an even lower-loss one near 1.55 μm . Consider the window at 1.55 μm , which is approximately 25,000 GHz wide. (Note that due to the extremely desirable characteristics of the erbium-doped fiber amplifier (EDFA), which amplifies only near 1.55 μm , most systems would use EDFAs and therefore not use the dispersion-zero 1.3- μm band of the existing embedded conventional fiber base.) The high-bandwidth characteristic of the optical fiber implies that a single optical carrier at 1.55 μm can be baseband-modulated at approximately 25,000 Gb/s, occupying 25,000 GHz surrounding 1.55 μm , before transmission losses of the optical fiber would limit transmission. Obviously, this bit rate is impossible for present-day electrical and optical devices to achieve, given that even heroic lasers, external modulators, switches,

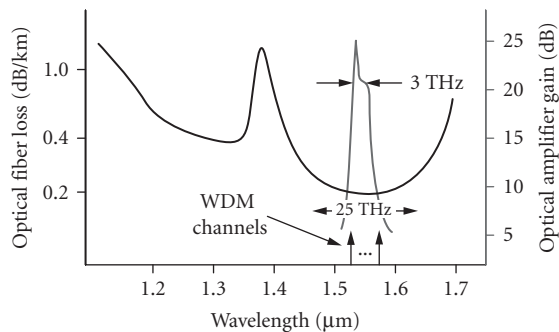


FIGURE 1 Fiber loss as a function of wavelength in conventional single-mode silica fiber. The gain spectrum of the EDFA is also shown.¹²

and detectors all have bandwidths less than or equal to 100 GHz. Practical data links today are significantly slower, perhaps no more than tens of gigabits per second per channel. Since the single high-speed channel makes use of an extremely small portion of the available fiber bandwidth, an efficient multiplexing method is needed to take full advantage of the huge bandwidth offered by optical fibers. As we will see in this chapter, WDM has been proven to be the most appropriate approach.

Introduction to WDM Technology

In real systems, even a single channel will probably be a combination of many lower-speed signals since few individual applications today utilize this high bandwidth. These lower-speed channels are multiplexed together in time to form a higher-speed channel. This time-division multiplexing (TDM) can be accomplished in either the electrical or optical domain. In TDM, each lower-speed channel transmits a bit (or a collection of bits known as a packet) in a given time slot and then waits its turn to transmit another bit (or packet) after all the other channels have had their opportunity to transmit. Until the late 1980s, fiber communication was mainly confined to transmitting a single optical channel using TDM technology. Due to fiber attenuation, this channel required periodic regeneration which included detection, electronic processing, and optical retransmission. Such regeneration causes a high-speed optoelectronic bottleneck, is bit-rate specific, and can only handle a single wavelength. The need for these single-channel regenerators (i.e., repeaters) was replaced when the EDFA was developed, enabling high-speed repeaterless single-channel transmission. We can think of this single approximate gigabits per second channel as a single high-speed “lane” in a highway in which the cars represent packets of optical data and the highway represents the optical fiber. It seems natural to dramatically increase the system capacity by transmitting several different independent wavelengths simultaneously down a fiber in order to more fully utilize this enormous fiber bandwidth.^{13,14} Therefore, the intent was to develop a multiple-lane highway, with each lane representing data traveling on a different wavelength.

In the most basic WDM arrangement as shown in Fig. 2, the desired number of lasers, each emitting a different wavelength, are multiplexed together by a wavelength multiplexer (or a combiner) into the same high-bandwidth fiber.^{13–16} Each of N different-wavelength lasers is operating at the slower gigabits per second speeds, but the aggregate system is transmitting at N times the individual

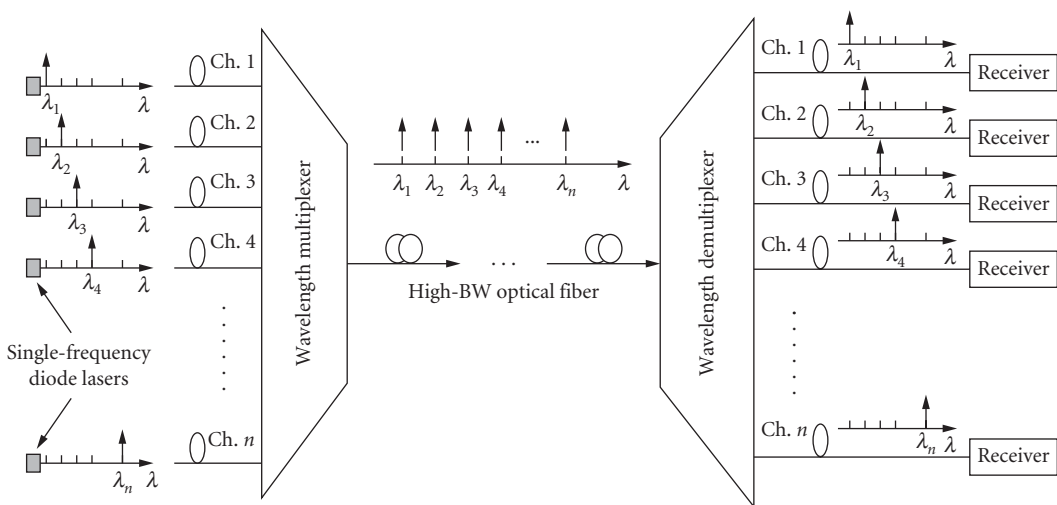


FIGURE 2 Diagram of a simple WDM system.

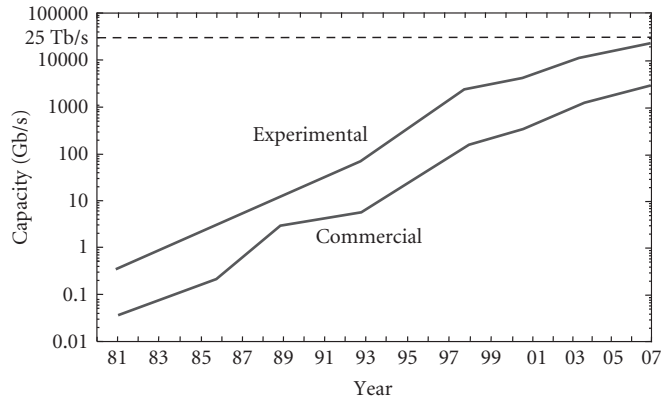


FIGURE 3 Continuous capacity growth in optical fiber transmission systems.

laser speed providing a significant capacity enhancement. After being transmitted through a high-bandwidth optical fiber, the combined optical signals must be demultiplexed by a wavelength demultiplexer at the receiving end by distributing the total optical power to each output port and then requiring that each receiver selectively recovers only one wavelength. Therefore, only one signal is allowed to pass and establish a connection between source and destination. WDM allows us to make use much of the available fiber bandwidth, although various device, system, and network issues will still limit utilization of the full fiber bandwidth.

Figure 3 shows the continuous capacity growth in optical fiber systems over the past 30 years. The highest capacity has been achieved using WDM. One interesting point about this trend predicted two decades ago by T. Li of AT&T Bell Labs is that the transmission capacity doubles every 2 years. WDM technology has provided the platform for this trend to continue, and there is no reason to assume that WDM won't continue to produce dramatic progress.

21.2 BASIC ARCHITECTURE OF WDM NETWORKS

We have explained how WDM enables the utilization of a significant portion of the available fiber bandwidth by allowing many independent signals to be transmitted simultaneously in one fiber. In fact, WDM technology also enables wavelength routing and switching of data paths in an optical network. By utilizing wavelength-selective components, each data channel's wavelength can be routed and detected independently through the network. The wavelength determines the communication path by acting as the signature address of the origin, destination, or routing. Data can then be presumed not traveling on optical fiber but on wavelength-specific "light-paths" from source to destination that can be arranged by a network controller to optimize throughput. Therefore, the basic system architecture that can take the full advantage of WDM technology is an important issue, and will be discussed in this section.

Point-to-Point Links

As shown in Fig. 2, in a simple point-to-point WDM system, several channels are multiplexed at one node, the combined signals are then transmitted across some distance of fiber, and the channels are demultiplexed at a destination node. This point-to-point WDM link facilitates the high-bandwidth fiber transmission without routing or switching in the optical data path.

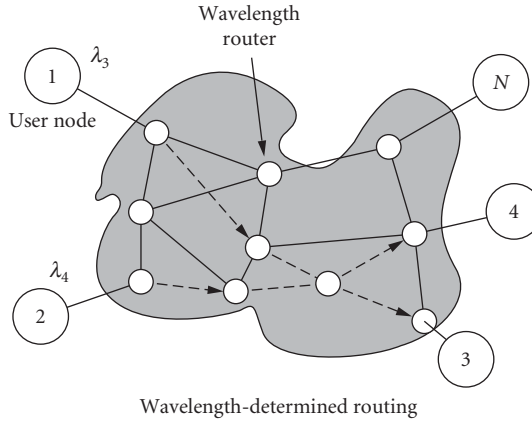


FIGURE 4 A generic multiuser network in which the communications links and routing paths are determined by the wavelengths used within the optical switching fabric.

Wavelength-Routed Networks

Figure 4 shows a more complex multiuser WDM network structure, where the wavelength is used as the signature address for either the transmitters or the receivers, and determines the routing path through an optical network. In order for each node to be able to communicate with any other node and facilitate proper link setup, the transmitters or the receivers must be wavelength tunable; we have arbitrarily chosen the transmitters to be tunable in this network example. Note that the wavelengths are routed passively in wavelength-routed networks.

WDM Stars, Rings, and Meshes

Three common WDM network topologies are star, ring, and mesh networks.^{17–19} In the star topology, each node has a transmitter and receiver, with the transmitter connected to one of the passive central star's inputs and the receiver connected to one of the star's outputs, as is shown in Fig. 5a. Rings, as shown in Fig. 5b, are also popular because: (1) many electrical networks use this topology, and (2) rings are easy to implement for any geographical network configuration. In this example,

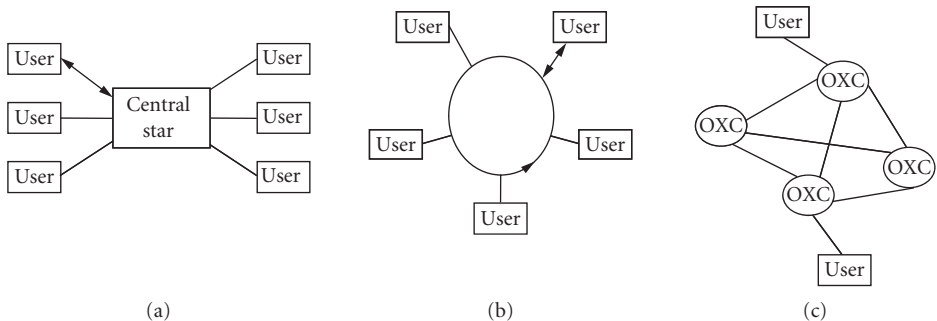


FIGURE 5 WDM: (a) stars; (b) rings; and (c) meshes.

each node in the unidirectional ring can transmit on a specific signature wavelength, and each node can recover any other node's wavelength signal by means of a wavelength-tunable receiver. Although not depicted in the figure, each node must recover a specific channel. This can be performed: (1) where a small portion of the combined traffic is tapped off by a passive optical coupler, thereby allowing a tunable filter to recover a specific channel, or (2) in which a channel-dropping filter completely removes only the desired signal and allows all other channels to continue propagating around the ring. Furthermore, a synchronous optical network (SONET) dual-ring architecture, with one ring providing service and the other protection, can provide automatic fault detection and protection switching.²⁰

In both the star and ring topologies, each node has a designated wavelength, and any two nodes can communicate with each other by transmitting and recovering that wavelength. This implies that N wavelengths are required to connect N nodes. The obvious advantage of this configuration, known as a single-hop network, is that data transfer occurs with an uninterrupted optical path between the origin and destination; the optical data starts at the originating node and reaches the destination node without stopping at any other intermediate node. A disadvantage of this single-hop WDM network is that the network and all its components must accommodate N wavelengths, which may be difficult (or impossible) to achieve in a large network, that is, the present fabrication technology cannot provide and the transmission capability cannot accommodate 1000 distinct wavelengths for a 1000-user network.

It is important to address that the reliability is a problem in fiber ring. If a station is disabled or if a fiber breaks, the whole network goes down. To address this problem, a double-ring optical network, also called a "self-healing" ring, is used to bypass the defective stations and loops back around a fiber break, as shown in Fig. 6. Each station has two inputs and two outputs connected to two rings that operate in opposite directions.

An alternative to require N wavelengths to accommodate N nodes is to have a multihop network (mesh network) in which two nodes can communicate with each other by sending data through a third node, with many such intermediate hops possible, shown in Fig. 5c. In the mesh network, the nodes are connected by reconfigurable optical crossconnects (OXC).²¹ The wavelength can be dynamically switched and routed by controlling the OXCs. Therefore, the required number of wavelengths and the tunable range of the components can be reduced in this topology. Moreover, the mesh topology can also provide multiple paths between two nodes to make the network protection and restoration easier to realize. If a failure occurs in one of the paths, the system can automatically find another path and restore communications between any two nodes. However, OXCs with large numbers of ports are extremely difficult to obtain, which limits the scalability of the mesh network.

In addition, there exist several other network topologies, such as tree network, which is widely used in broadcasting or distributing systems. At the "base" of the tree is the source transmitter from which emanates the signal to be broadcast throughout the network. From this base, the tree splits

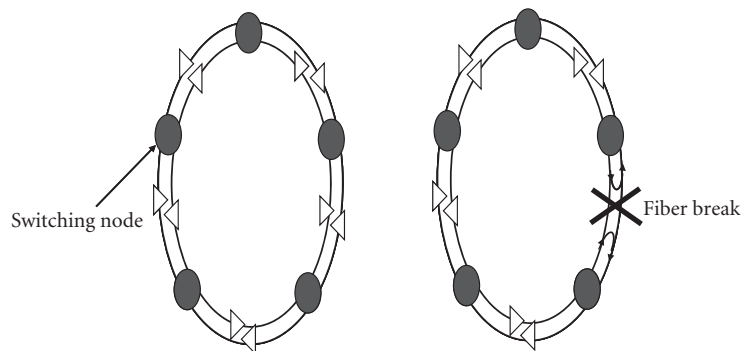


FIGURE 6 A self-healing ring network.

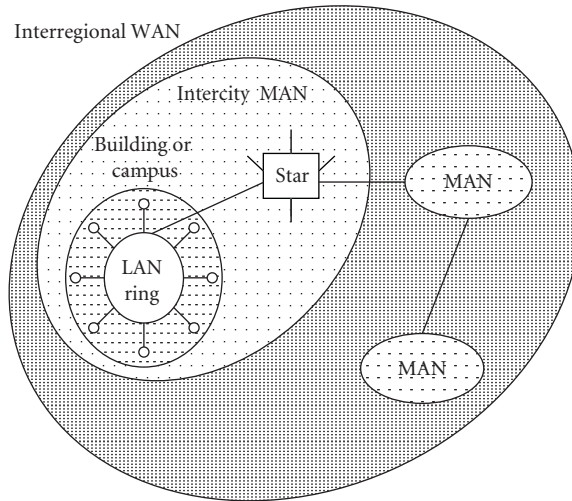


FIGURE 7 Hybrid network topologies and architectures woven together to form a large network.

many times into different “branches,” with each branch either having nodes connected to it or further dividing into subbranches. This continues until all the nodes in the network can access the base transmitter. Whereas the other topologies are intended to support bidirectional communication among the nodes, this topology is useful for distributing information unidirectionally from a central point to a multitude of users. This is a very straightforward topology and is in use in many systems, most notably cable television (CATV).

By introducing Fig. 7 in which a larger network is composed of smaller ones, we have also introduced the subject of the architecture of the network which depends on the network’s geographical extent. The three main architectural types are the local-, metropolitan-, and wide-area networks, denoted by LAN, MAN, and WAN, respectively.²² Although no rule exists, the generally accepted understanding is that a LAN interconnects a small number of users covering a few kilometers (i.e., intra- and interbuilding), a MAN interconnects users inside a city and its outlying regions, and a WAN interconnects significant portions of a country (100 of kilometers). Based on Fig. 7, the smaller networks represent LANs, the larger ones MANs, and the entire figure would represent a WAN. In other words, a WAN is composed of smaller MANs, and a MAN is composed of smaller LANs. Hybrid systems exist, and typically a wide-area network will consist of smaller local-area networks, with mixing and matching between the most practical topologies for a given system. For example, stars and rings may be desirable for LANs whereas buses may be the only practical solution for WANs. It is, at present, unclear which network topology and architecture will ultimately and most effectively take advantage of high-capacity optical systems.

Circuit and Packet Switching

The two fundamental types of underlying telecommunication network infrastructures, based on how traffic is multiplexed and switched, are *circuit-switched* and *packet switched*. A circuit-switched network provides circuit-switched connections to its customers. Once the connection is established, a guaranteed amount of bandwidth is allocated to each connection and become available to the connection anytime. The network is also transparent and the nodes seem to be directly connected. In addition, circuit switching requires a lower switching speed (more than milliseconds). Many types of communication links and distribution systems may satisfactorily be interconnected by

circuit switching which is relatively simple to operate. The problems with circuit switching include (1) it is not efficient at handling bursty data traffic (low utilization for traffic with changing intensity or short lived connections); (2) there is a delay before the connection can be used; (3) the resources are permanently allocated to a connection and cannot be used for any other users, and (4) circuit-switched networks are more sensitive to faults (e.g., if a part of the connection fails, the whole transfer fails).

Since optical circuit switching is the relatively mature technology today, the current deployed WDM wavelength-routing network is generally circuit-switched. The basic mechanism of communication in a wavelength-routed network is a lightpath (corresponding to a circuit), which is an all optical connection (communication channel) linking multiple optical segments from a source node to a destination node over a wavelength on each intermediate link. At each intermediate node of the network, the lightpath is routed and switched from one link to another link. A lightpath can use either the same wavelength throughout the whole link or a concatenation of different wavelengths after undergoing wavelength conversion at intermediate optical nodes. In the absence of any wavelength conversion device, a lightpath is required to be on the same wavelength channel throughout its path in the network; this requirement is referred to as the wavelength continuity property of the lightpath. Once the setup of a lightpath is completed, the whole lightpath is available during the connection. Note that different lightpaths can use the same wavelength as long as they do not share any common links (i.e., same wavelength can be reused spatially in different parts of the network).

As shown in Fig. 8, the key network elements in the wavelength-routing network are optical line terminal, OADM (see Fig. 9), and OXC (see Fig. 10). There are many different OADM structures such as parallel or serial, fixed or reconfigurable. In general, an ideal OADM would add/drop any channel and any number of channels, and would be remotely controlled and reconfigured without disturbance to unaffected channels. There will be more discussion on reconfigurable in “Network Reconfigurability” section. The other requirements for an OADM include low and fixed loss, independent of set of wavelengths dropped, and low cost. An OXC can switch the channels from input to output ports and input to output wavelengths. The functions in an OXC node include providing lightpath, rerouting (switching protection), restoring failed lightpath, monitoring performances, accessing to test signals, wavelength conversion, and multiplexing and grooming. An OXC can be either

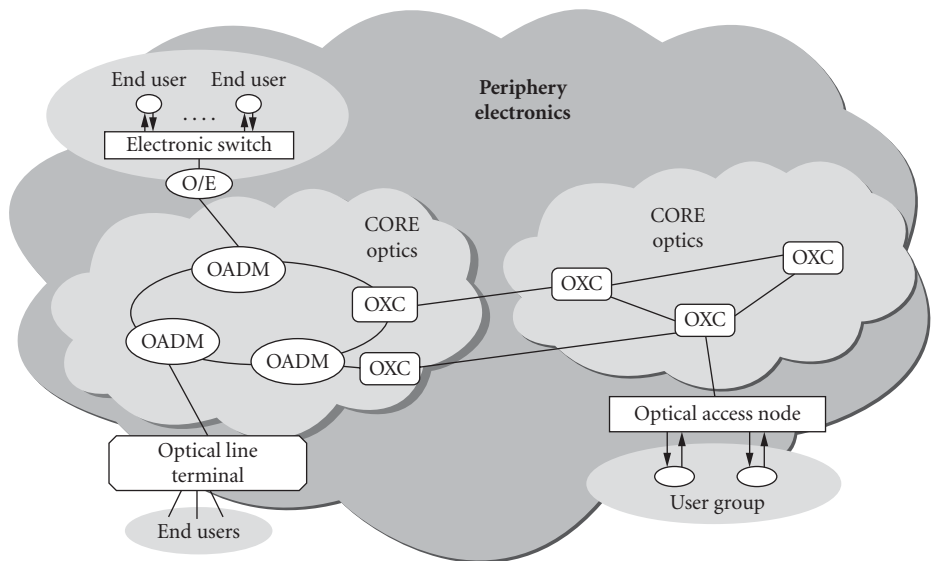


FIGURE 8 An optical network showing optical line terminal, OADM, and OXC.

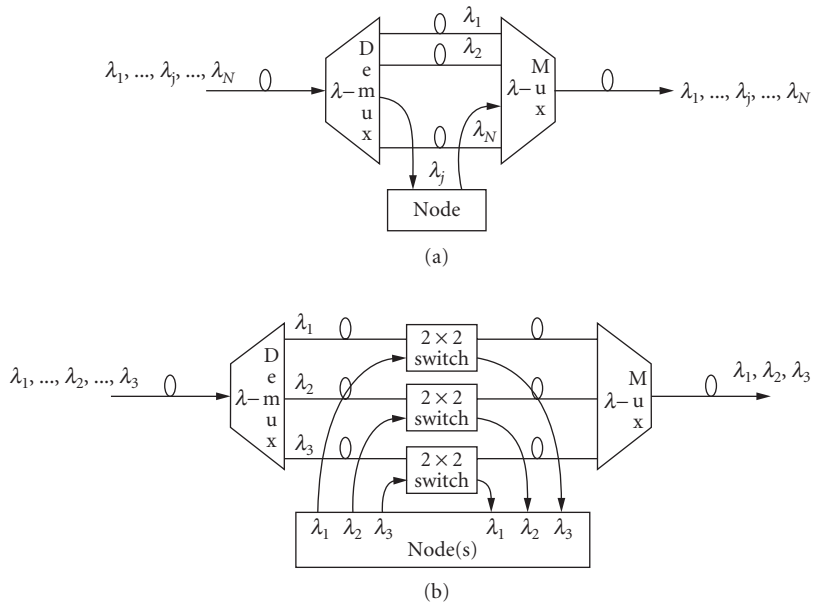


FIGURE 9 Optical add-drop multiplexing (OADM) systems: (a) fixed versus (b) reconfigurable.

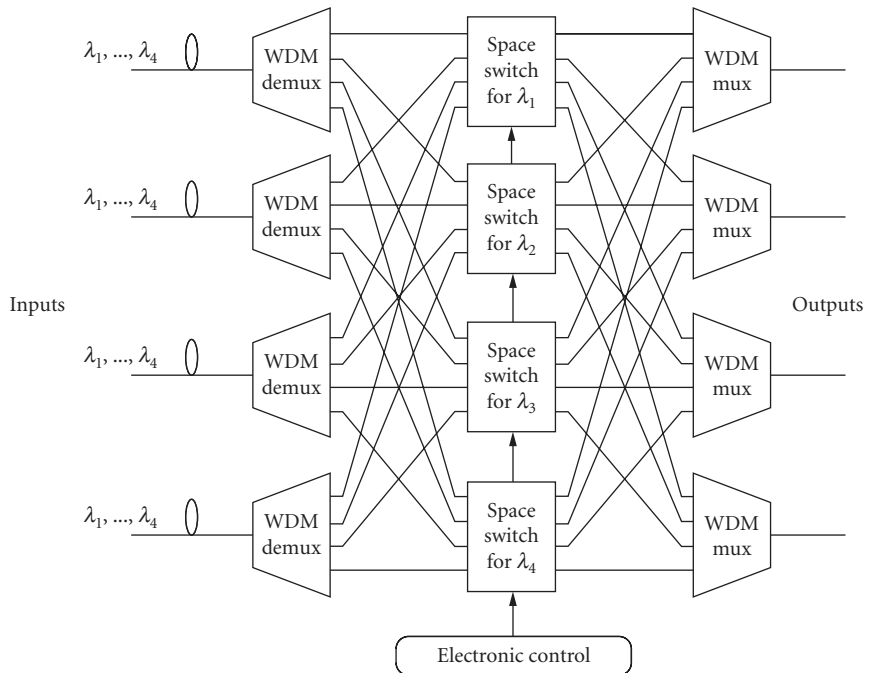


FIGURE 10 An optical crossconnect system in reconfigurable optical networks.

electrical (performing O-E-O conversion for each WDM channels) or optical, transparent or opaque. To accomplish all these functions, the OXC needs three building blocks: (1) fiber switching, to route all of the wavelengths on an incoming fiber to a different outgoing fiber (optical space switches); (2) wavelength switching, to switch specific wavelengths from an incoming fiber to multiple outgoing fibers (multiplexing/demultiplexing); and (3) wavelength conversion, to take incoming wavelengths and convert them to another optical frequency on the outgoing port. This is essential to achieve strictly nonblocking architectures when using wavelength switching.

In packet switched networks, the data stream is broken into small packets. These data packets are multiplexed together with packets from other data stream inside the network. The packets are switched inside the network based on their destination. To facilitate this switching, a packet header is added to the payload in each packet. The header carries addressing information. The switching nodes read the header and then determine where to switch the packet. At the receiver end, packets belonging to a particular stream are put back together. Therefore, packet switching requires switching speeds of microseconds or less. For high-speed optical transmission, packet switching holds the promise for more efficient data transfer, for which no long-distance handshaking is required. The high-bandwidth links are used more efficiently.

As shown in Fig. 11, an optical packet switching node is generally composed of three parts: control unit, switching unit, and input/output interfaces. The control unit retains information about network topology, the forwarding table, scheduling, and buffering. It decides the switching time and is in charge of resolving contentions at a node. The switching unit allows the data to remain in the optical domain during the routing process. It is especially important for optical packet-switched networks that the switching speed be fast enough to minimize overhead. The input/output interface is where optical technologies are utilized to deal with contention problems. Optical buffers and wavelength converters are the building blocks of time and wavelength domain contention resolution modules, respectively, and are housed in the interface units. In addition, other physical layer functionalities required for an optical switching node such as synchronization are realized at the interface.

Note that network packet switching can be accomplished in a conceptually straightforward manner by requiring a node to optoelectronically detect and retransmit each and every incoming optical data packet. The control and routing information is contained in the newly detected electronic packet, and all the switching functions can occur in the electrical domain prior to optical retransmission of the signal. Unfortunately, this approach implies that an optoelectronic speed bottleneck will eventually occur in this system. On the other hand, it is extremely difficult to accomplish the signal processing in the optical domain currently. Alternatively, much research is focused toward maintaining an all-optical data path and performing the switching functions all-optically with only some electronic control of the optical components. The reason is that the control unit detects and processes only the

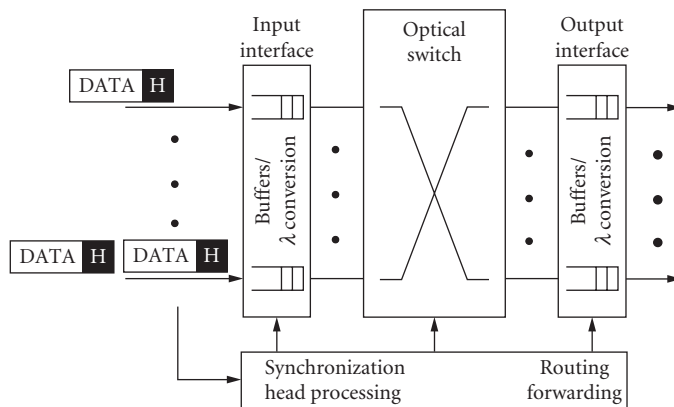


FIGURE 11 Schematic diagram of an optical packet-switching node.

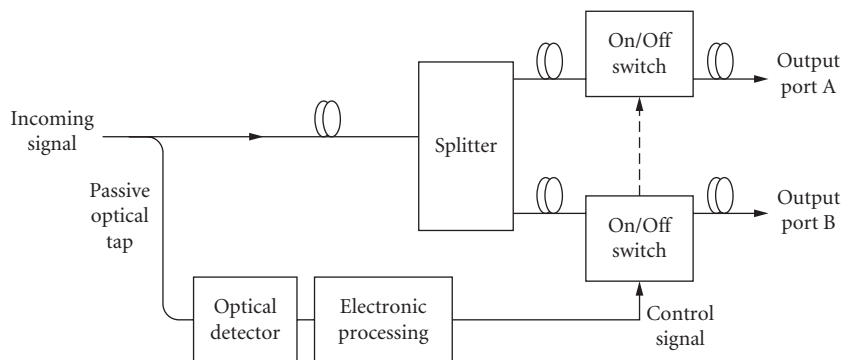


FIGURE 12 Passive optical tapping of an optical packet in order to determine routing information and allow a node to electronically control an optical switch.

header of a packet (not the payload). Therefore it is possible to transmit the header at a lower bit rate to facilitate the processing. As a result, the presence of electronics at the control unit does not necessarily pose a limitation on the data transmission rate. Figure 12 shows a generic solution which passively taps an incoming optical signal. Information about the signal is made known electrically to the node but the signal itself remains in the optical domain. The routing information may be contained in the packet header or in some other form (i.e., wavelength, etc.). Header information can be transmitted in several different ways. For example, the baseband header can be transmitted as a data field inside a packet either at the same bit rate as the data, or at a lower rate to relax the speed requirement on the electronics for header detection. It can also be located out-of-band either on a subcarrier on the same wavelength or on a different wavelength, altogether.

The various functions that may be performed in an optical switching node include (1) address/label recognition to determine the intended output port, (2) header updating or label swapping to prepare the packet header for the next node, (3) bit and/or packet synchronization to the local node to time the switching process, (4) routing-table caching as a reference for routing decisions, (5) output-port contention resolution via buffering and/or wavelength conversion, (6) signal monitoring to assess the signal quality, (7) signal regeneration to combat the accumulated distortion of the signal, and (8) optical switching to direct packets to the appropriate output ports.

It is very important to mention that in the past several years, data traffic has been growing at a much faster rate than voice traffic. Data traffic is “bursty” in nature, and reserving bandwidth for bursty traffic could be very inefficient. Circuit-switching networks are not optimized for this type of traffic. On the other hand, packet-switching networks require a complex processing unit and fast optical switches. Therefore, optical burst switching (OBS) was introduced to reduce the processing required for switching at each node and to avoid optical buffering. OBS is somewhere between packet switching and circuit switching with the switching period on the order of many packets. OBS takes advantage of time domain statistical multiplexing to utilize the large bandwidth of a single channel to transmit several lower-bandwidth bursty channels. At the network edge, packets are aggregated to generate bursts that are sent over the network core. In almost all OBS schemes, the header is sent separately from the payload with an offset time that is dependent on the scheme.^{23,24} The header is sent to the switches, and the path is then reserved for the payload that follows. The loose coupling between the control packet and the burst alleviates the need for optical buffering and relaxes the requirement for switching speed.

Many challenging optical switching issues require solutions,²⁵ such as routing control,^{26–28} contention resolution,^{29–40} and optical header processing.^{41–48} Most of these issues will relate to packet switching, although circuit switching will also require attention. Due to the limited space, we give only some references to facilitate further study by the reader.

Network Reconfigurability

In addition to high capacity, the reconfigurable WDM network could offer flexibility and availability. A reconfigurable network is highly desirable to meet the requirements of high bandwidth and bursty traffic in future networks. Through the reconfigurable network, service providers and network operators could respond quickly and cost-effectively to new revenue opportunities. As shown in Fig. 9, a fixed add/drop multiplexing node can only process the signal(s) at a given wavelength or a group of wavelengths. While in the dynamic reconfigurable node, operators could add or drop any number of wavelengths. This added flexibility would save operating and maintenance costs and improve network efficiency. In general, a network is reconfigurable if it can provide the following functionality for multichannel operations: (1) channel add/drop and (2) path reconfiguration for bandwidth allocation or restoration. It appears that a reconfigurable network is highly desirable to meet the requirements of high bandwidth and bursty traffic in future networks.

Since a reconfigurable network allows dynamic network optimization to accommodate changing traffic patterns, it provides more efficient use of network resources. Figure 13 shows blocking probability as a function of call arrival rate in a WDM ring network with 20 nodes.⁴⁹ A configurable topology can support 6 times the traffic of a fixed WDM topology for the same blocking probability.

Among many different solutions, the reconfigurable optical add/drop multiplexers (ROADMs) have emerged as key building blocks for the next-generation WDM systems with the goal of any wavelength anywhere. ROADMs add the ability to remotely switch traffic at the wavelength layer in WDM network, thus allowing individual data channels to be added or dropped without optical-electrical-optical (O-E-O) conversion of all WDM channels.^{50–55} Figure 14 shows a few ROADM architectures based on different switching technologies: discrete switches or switch matrix plus filters [mux/switch/demux, variable optical attenuators (VOA)]; wavelength blockers (WB); integrated planar lightwave circuits (iPLC); and wavelength-selective switches (WSS). With the exception of mux-switch-demux design, the devices are typically implemented in broadcast-and-select optical architectures with passive splitters in the pass-through path. A relevant attribute of ROADM technology is the integration of multiplexing/demultiplexing and switching into a single component. This integration can significantly lower pass-through losses when compared with multiple discrete components.⁵³

ROADM networks can deliver considerable operation benefits such as simplified planning and engineering, and improved network utilization. Its applications include optical multicast/broadcast, scalable colorless add/drop capacity, capacity/service upgrade without traffic disruption, cost-effective optical protection and restoration, and low cost of ownership.^{50,53}

The key component technologies enabling network reconfigurability include wavelength-tunable lasers and laser arrays, wavelength routers, optical switches, OXCs, OADMs, optical amplifiers, and tunable optical filters, and the like. Although huge benefits are possible with a reconfigurable topology,

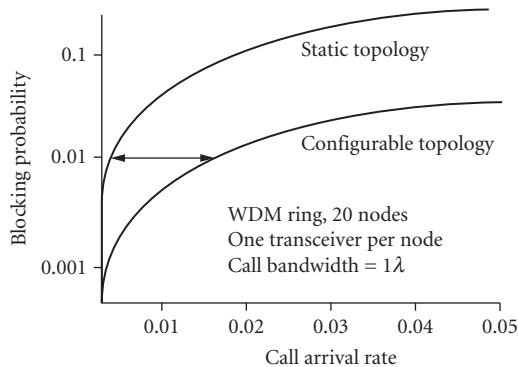


FIGURE 13 Blocking probability as a function of call arrival rate in a WDM ring.⁴⁹

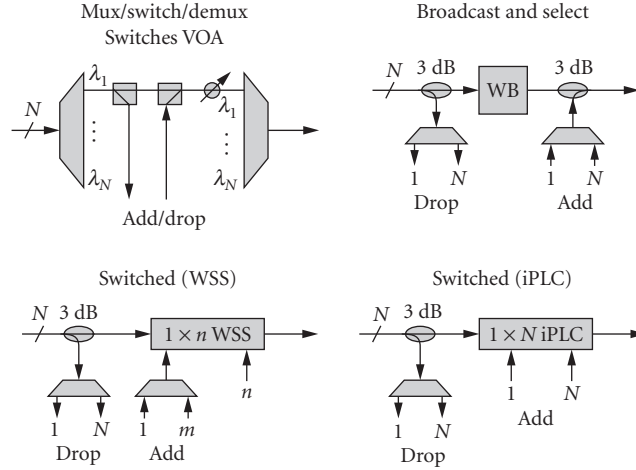


FIGURE 14 ROADMs architectures overview.⁵³

the path to reconfigurability is paved with various degrading effects. As shown in Fig. 4, the signal may pass through different lengths of fiber links due to the dynamic routing, causing some degrading effects in reconfigurable networks to be more critical than in static networks, such as nonstatic dispersion and nonlinearity accumulation due to reconfigurable paths, EDFA gain transients, channel power nonuniformity, cross-talk in optical switching and cross-connects, and wavelength drift of components. We will discuss some of these important effects in Sec. 21.3, followed by selected advanced technologies to dealing with these effects in WDM systems.

21.3 FIBER SYSTEM IMPAIRMENTS

One key benefit of reconfigurable WDM networks might be the transparency to bit rate, protocol and modulation format of all the various wavelength channels propagating in the system. However, key challenges exist when determining an optimum path through the network, since an optical wavelength might accumulate different physical impairments as it is switched through the network. These nonidealities will be imposed by both the transmission links and the optical switching nodes. Since system performance depends on many different optical impairments, a network-layer routing and wavelength assignment algorithm might rapidly provision a lightpath that cannot meet the signal-quality requirement.^{56–59} In this section, we will give a brief review of different physical-layer impairments, including fiber attenuation and power loss, fiber chromatic dispersion, fiber polarization mode dispersion, and fiber nonlinear effects. The management of both fiber dispersion and nonlinearities will also be discussed at the end of this section. Note that EDFA-related impairments, such as noise, fast power transients, and gain peaking in EDFA cascades will be described in Sec. 21.5.

Fiber Attenuation and Optical Power Loss

The most basic characteristic of a link is the power loss, which is caused by fiber attenuation and connections.⁶⁰ *Attenuation*, defined as the ratio of the input power to the output power, is the loss of optical power as light travels along the fiber. Attenuation in an optical fiber is caused by absorption, scattering, and bending losses. The fundamental physical limits imposed on the fiber attenuation are

due to scattering off the silica atoms at shorter wavelengths and the material absorption at longer wavelengths. There are two minima in the loss curve: one near $1.3\ \mu\text{m}$ and an even lower one near $1.55\ \mu\text{m}$ (see Fig. 1). Fiber bending can also induce power loss because radiation escapes through its bends. The bending loss is inversely proportional to the bend radius and is wavelength dependent.

Power loss is also present at fiber connections, such as connectors, splices, and couplers. Coupling of light into and out of a small-core fiber is much more difficult to achieve than coupling electrical signals in copper wires since: (1) photons are weakly confined to the waveguide whereas electrons are tightly bound to the wire, and (2) the core of a fiber is typically much smaller than the core of an electrical wire. First, light must be coupled into the fiber from a diverging laser beam, and two fibers must be connected to each other. Second, connecting two different fibers in a system must be performed with great care due to the small size of the cores. One wishes to achieve connections exhibiting: (1) low loss, (2) low back reflection, (3) repeatability, and (4) reliability. Two popular methods of connecting fibers are the permanent splice and the mechanical connector. The permanent “fusion” splice can be accomplished by placing two fiber ends near each other, generating a high-voltage electric arc which melts the fiber ends, and “fusing” the fibers together. Losses and back reflections tend to be extremely low, being less than 0.1 dB and less than $-60\ \text{dB}$, respectively. Disadvantages of these fusion splices include (1) the splice is delicate and must be protected, and (2) the splice is permanent. Alternatively, there are several types of mechanical connectors, such as ST and FC/PC. Losses and back reflections are still fairly good, and are typically less than 0.3 dB and less than $-45\ \text{dB}$, respectively.

Low loss is extremely important since a light pulse must contain a certain minimum amount of power in order to be detected, such that “0” or “1” data bit can be unambiguously detected. If not for dispersion, we would clearly prefer to operate with $1.55\ \mu\text{m}$ light due to its lower loss for long-distance systems.

Chromatic Dispersion

In fact, in any medium (other than vacuum) and in any waveguide structure (other than ideal infinite free space), different electromagnetic frequencies travel at different speeds. This is the essence of chromatic dispersion. As the real fiber-optic world is rather distant from the ideal concepts of both vacuum and infinite free space, dispersion will always be a concern when one is dealing with the propagation of electromagnetic radiation through fiber. The velocity in fiber of a single monochromatic wavelength is constant. However, data modulation causes a broadening of the spectrum of even the most monochromatic laser pulse. Thus, all modulated data has a nonzero spectral width which spans several wavelengths, and the different spectral components of modulated data travel at different speeds. In particular, for digital data intensity modulated on an optical carrier, chromatic dispersion leads to pulse broadening—which in turn leads to chromatic dispersion limiting the maximum data rate that can be transmitted through optical fiber (see Fig. 15).

Considering that the chromatic dispersion in optical fibers is due to the frequency-dependent nature of the propagation characteristics for both the material and the waveguide structure, the speed of light of a particular wavelength λ can be expressed as follows using a Taylor series expansion of the value of the refractive index as a function of the wavelength.

$$v(\lambda) = \frac{c_0}{n(\lambda)} = \frac{c_0}{n_0(\lambda_0) + \frac{\partial n}{\partial \lambda} \delta\lambda + \frac{\partial^2 n}{\partial \lambda^2} (\delta\lambda)^2} \quad (1)$$

where c_0 is the speed of light in vacuum, λ_0 is a reference wavelength, and the terms in $\partial n / \partial \lambda$ and $\partial^2 n / \partial \lambda^2$ are associated with the chromatic dispersion and the dispersion slope (i.e., the variation of the chromatic dispersion with wavelength), respectively. Transmission fiber has positive dispersion, that is, longer wavelengths see longer propagation delays. The units of chromatic dispersion are picoseconds per nanometer per kilometer, meaning that shorter time pulses, wider frequency spread due to data modulation, and longer fiber lengths will each contribute linearly to temporal dispersion. Figure 16 shows the dispersion coefficient, D (ps/nm·km), of a conventional single-mode fiber with the material and waveguide contributions plotted separately.⁶⁰ For a given system, a pulse will disperse

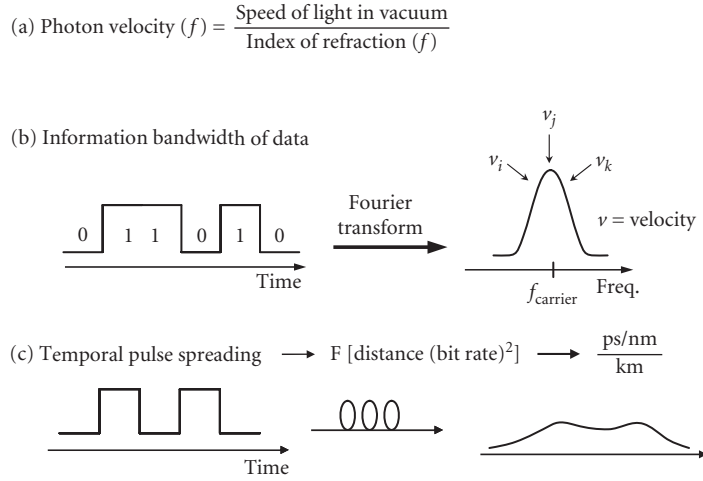


FIGURE 15 The origin of chromatic dispersion in data transmission. (a) Chromatic dispersion is caused by the frequency-dependent refractive index in fiber. (b) The nonzero spectral width due to data modulation. (c) Dispersion leads to pulse broadening, proportional to the transmission distance and the data rate.

more in time for a wider frequency distribution of the light and for a longer length of fiber. Higher data rates inherently have both shorter pulses and wider frequency spreads. Therefore, as network speed increases, the impact of chromatic dispersion rises precipitously as the square of the increase in data rate. The quadratic increase with the data rate is a result of two effects, each with a linear contribution. On one hand, a doubling of the data rate makes the spectrum twice as wide, doubling the effect of dispersion. On the other hand, the same doubling of the data rate makes the data pulses only half as long (hence twice as sensitive to dispersion). The combination of a wider signal spectrum and a shorter pulse width is what leads to the overall quadratic impact—when the bit rate increases by a factor of 4, the effects of chromatic dispersion increase by a whopping factor of 16!⁶¹

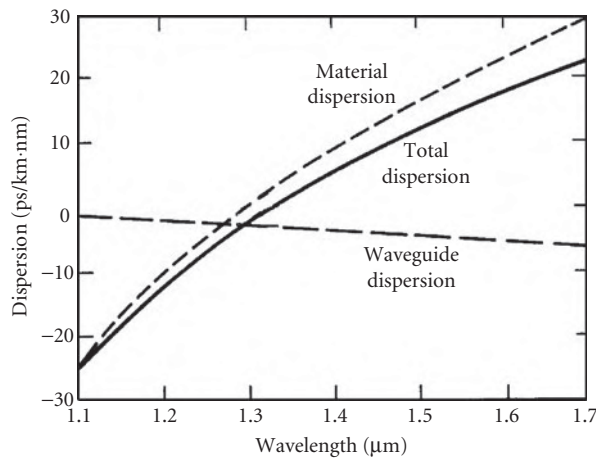


FIGURE 16 Dispersion coefficient, D , as a function of wavelength in the conventional silica single-mode fiber.⁶⁰

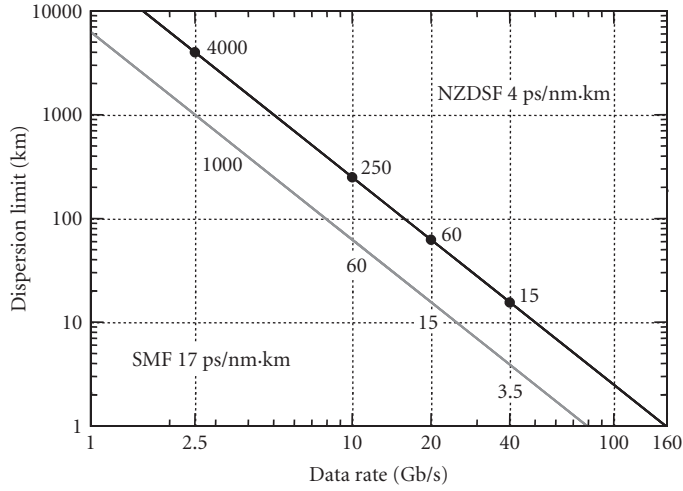


FIGURE 17 Transmission distance limitations due to uncompensated dispersion in SMF as a function of data rate for intensity modulated optical signals.⁶²

The data rate and the data-modulation format can significantly affect the sensitivity of a system to chromatic dispersion. For example, the common non-return-to-zero (NRZ) data format, in which the optical power stays high throughout the entire time slot of a “1” bit, is more robust to chromatic dispersion than is the return-to-zero (RZ) format, in which the optical power stays high in only part of the time slot of a “1” bit. This difference is due to the fact that RZ data has a much wider channel frequency spectrum compared to NRZ data, thus incurring more chromatic dispersion. However, in a real WDM system, the RZ format increases the maximum allowable transmission distance by virtue of its reduced duty cycle (compared to the NRZ format) making it less susceptible to fiber nonlinearities. We will discuss some robust modulation formats in Sec. 21.4 too.

A rule for the maximum distance over which data can be transmitted is to consider a broadening of the pulse equal to the bit period. For a bit period B , a dispersion value D and a spectral width $\Delta\lambda$, the dispersion-limited distance is given by

$$L_D = \frac{1}{D \cdot B \cdot \Delta\lambda} = \frac{1}{D \cdot B \cdot (cB)} \propto \frac{1}{B^2} \quad (2)$$

(see Fig. 17). For example, for single mode fiber, $D = 17$ ps/nm-km, so for 10 Gb/s data the distance is L_D equal to 52 km. In fact, a more exact calculation shows that for 60 km, the dispersion induced power penalty is less than 1 dB.⁶² The power penalty for uncompensated dispersion rises exponentially with transmission distance, and thus to maintain good signal quality, dispersion compensation is required.

Polarization-Mode Dispersion

Single-mode fibers actually support two perpendicular polarizations of the original transmitted signal (fundamental mode). In an ideal fiber (perfect) these two modes are indistinguishable, and have the same propagation constants owing to the cylindrical symmetry of the waveguide. However, the core of an optical fiber may not be perfectly circular, and the resultant ellipse has two orthogonal axes. The index-of-refraction of a waveguide, which determines the speed of light, depends on the shape of the waveguide as well as the glass material itself. Therefore, light polarized along one fiber axis travels at a

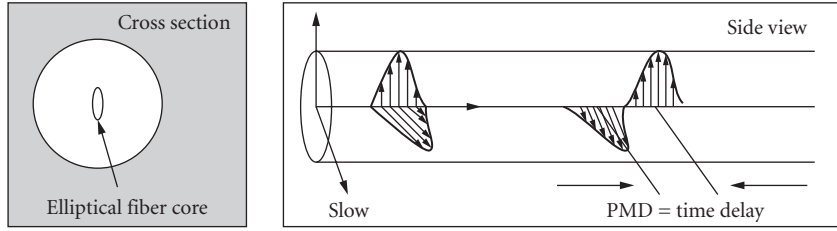


FIGURE 18 Illustration of polarization mode dispersion caused by imperfect round fiber core. An input optical pulse has its power transmitted on two orthogonal polarization modes, each arriving at different times.

different speed as does light polarized along the orthogonal fiber axis (see Fig. 18). This phenomenon is called polarization mode dispersion (PMD). Fiber asymmetry may be inherent in the fiber from the manufacturing process, or it may be a result of mechanical stress on the deployed fiber. The inherent asymmetries of the fiber are fairly constant over time, while the mechanical stress due to movement of the fiber can vary, resulting in a dynamic aspect to PMD. Since the light in the two orthogonal axes travel with different group velocities, to the first order, this differential light speed will cause a temporal spreading of signals, which is termed the differential group delay (DGD).

Because of random variations in the perturbations along a fiber span, PMD in long fiber spans accumulates in a random-walk-like process that leads to a square root of transmission-length dependence.⁶³ Moreover, PMD does not have a single value for a given span of fiber. Rather, it is described in terms of average DGD, and a fiber has a distribution of DGD values over time. The probability of the DGD of a fiber section being a certain value at any particular time follows a maxwellian distribution (see Fig. 19). The probability of DGD = $\Delta\tau$ given by

$$\text{prob}(\Delta\tau) = \sqrt{\frac{2}{\pi}} \frac{\Delta\tau^2}{\alpha^3} \exp\left(-\frac{\Delta\tau^2}{2\alpha^2}\right) \quad (3)$$

with mean value $\langle\Delta\tau\rangle = \sqrt{8/\pi}\alpha$. PMD is usually expressed in $\text{ps}/\text{km}^{1/2}$ in long fiber spans, and the typical PMD parameter (D_p) is 0.1 to 10 $\text{ps}/\text{km}^{1/2}$.^{64,65}

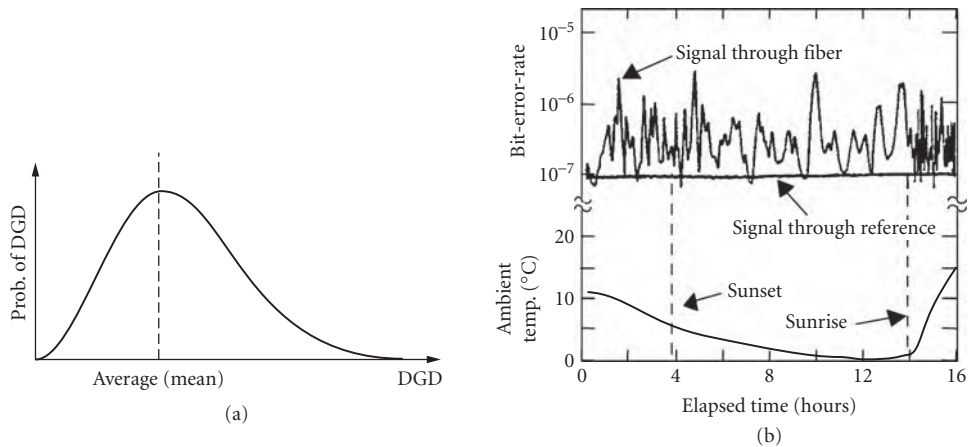


FIGURE 19 (a) Probability distribution of DGD in a typical fiber. (b) System performance (bit-error rate) fluctuations due to changes in temperature caused by PMD.⁶⁶

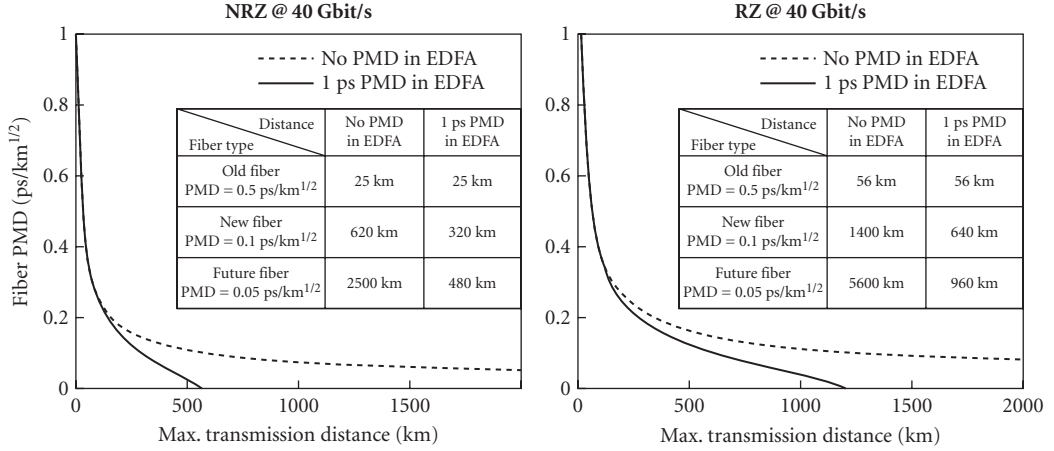


FIGURE 20 Limitations of transmission distances caused by fiber PMD.

Today's fiber has a very low PMD value and is well characterized. But there is still a small residual asymmetry in the fiber core. Moreover, slight polarization dependencies exist in discrete inline components such as isolators, couplers, filters, erbium-doped fiber, modulators, and multiplexers. Therefore, even under the best of circumstances, PMD still significantly limit the deployment of more than or equal to 40 Gb/s systems (see Fig. 20).⁶⁷

Other polarization-related impairments, such as polarization-dependent loss (PDL), and polarization-dependent gain (PDG), may also cause deleterious effects in a fiber transmission link.⁶⁸ Moreover, the interaction between PMD and PDL/PDG may lead to significant overall performance degradation, which dramatically surpasses the result of adding the degradations induced by the two impairments independently.^{69–73} The readers can find more in-depth discussion in the related literatures.

Fiber Nonlinearities

Most nonlinear effects originate from the nonlinear refractive index of fiber. The refractive index not only depends on the frequency of light but also on the intensity (optical power), and it is related to the optical power as⁷⁴

$$\tilde{n}(\omega, P) = n_o(\omega) + n_2 I = n_o(\omega) + n_2 \frac{P}{A_{\text{eff}}} \quad (4)$$

where $n_o(\omega)$ is the linear refractive index of silica, n_2 is the intensity-dependent refractive index coefficient, P is the optical power inside the fiber, and A_{eff} is the effective mode area of the fiber. The typical value of n_2 is $2.6 \times 10^{-20} \text{ m}^2/\text{W}$. This number takes into account the averaging of the polarization states of the light as it travels in the fiber. The intensity dependence of the refractive index gives rise to three major nonlinear effects.

Self-Phase Modulation A million photons “see” a different glass than does a single photon, and a photon traveling along with many other photons will slow down. Self-phase modulation (SPM) occurs because of the varying intensity profile of an optical pulse on a single WDM channel (see Fig. 21a). This intensity profile causes a refractive index profile and, thus, a photon speed differential. The resulting phase change for light propagating in an optical fiber is expressed as

$$\Phi_{\text{NL}} = \gamma P L_{\text{eff}} \quad (5)$$

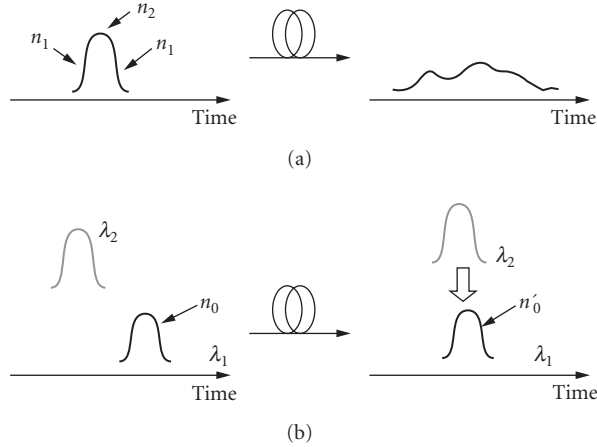


FIGURE 21 (a) Self-phase modulation: the photons in the pulse “see” different refractive index and (b) cross-phase modulation: the glass that a photon in the λ_2 pulse “sees” changes as other channels (with potentially varying power) move to coincide with the λ_2 pulse.

where the quantities γ and L_{eff} are defined as

$$\gamma = \frac{2\pi n_2}{\lambda A_{\text{eff}}} \quad \text{and} \quad L_{\text{eff}} = \frac{1 - e^{-\alpha L}}{\alpha} \quad (6)$$

where α is the fiber attenuation loss, L_{eff} is the effective nonlinear length of the fiber that accounts for fiber loss, and γ is the nonlinear coefficient measured in radians per kilometer per watt. A typical range of values for γ is about 10^{-30} rad/km · W. Although the nonlinear coefficient is small, the long transmission lengths and high optical powers that have been made possible by the use of optical amplifiers can cause a large enough nonlinear phase change to play a significant role in state-of-the-art lightwave systems.

Cross-Phase Modulation When considering many WDM channels copropagating in a fiber, photons from channels 2 through N can distort the index profile that is experienced by channel 1. The photons from the other channels “chirp” the signal frequencies on channel 1, which will interact with fiber chromatic dispersion and cause temporal distortion (see Fig. 21b). This effect is called cross-phase modulation (XPM). In a two-channel system, the frequency chirp in channel 1 due to power fluctuation within both channels is given by

$$\Delta B = \frac{d\Phi_{\text{NL}}}{dt} = \gamma L_{\text{eff}} \frac{dP_1}{dt} + 2\gamma L_{\text{eff}} \frac{dP_2}{dt} \quad (7)$$

where, dP_1/dt and dP_2/dt are the time derivatives of the pulse powers of channels 1 and 2, respectively. The first term on right hand side of the above equation is due to SPM, and the second term is due to XPM. Note that the XPM-induced chirp term is double that of the SPM-induced chirp term. As such, XPM can impose a much greater limitation on WDM systems than can SPM, especially in systems with many WDM channels.

Four-Wave-Mixing The optical intensity propagating through the fiber is related to the electric field intensity squared. In a WDM system, the total electric field is the sum of the electric fields of each individual channel. When squaring the sum of different fields, products emerge that are beat terms

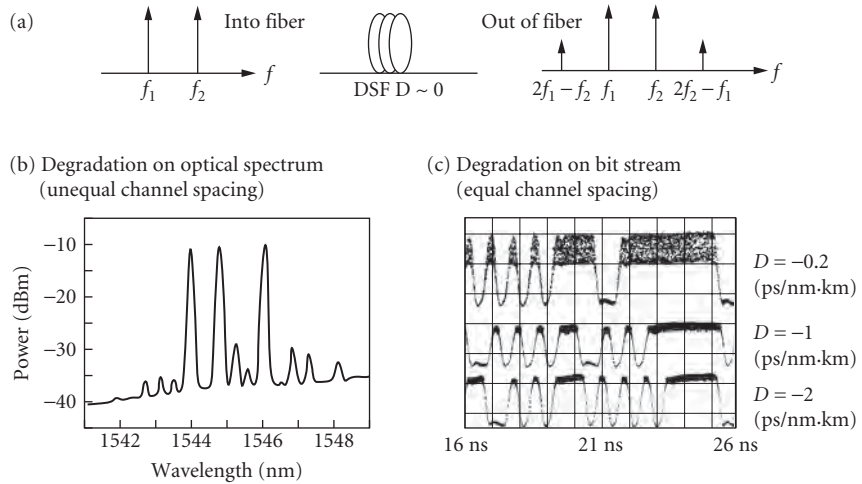


FIGURE 22 (a) and (b) FWM induces new spectral components via nonlinear mixing of two wavelength signals. (c) The signal degradation due to FWM products falling on a third data channel can be reduced by even small amounts of dispersion.⁷⁵

at various sum and difference frequencies to the original signals. Figure 22 depicts that if a WDM channel exists at one of the four-wave-mixing (FWM) beat-term frequencies, the beat term will interfere coherently with this WDM channel and potentially destroy the data.

Other Nonlinear Effects The nonlinear effects described above are governed by the power dependence of refractive index, and are elastic in the sense that no energy is exchanged between the electromagnetic field and the dielectric medium. A second class of nonlinear effects results from stimulated inelastic scattering in which the optical field transfers part of its energy to the nonlinear medium. Two important nonlinear effects fall in this category:⁷⁴ (1) stimulated Raman scattering (SRS) and (2) stimulated Brillouin scattering (SBS). The main difference between the two is that optical phonons participate in SRS, while acoustic phonons participate in SBS. In a simple quantum-mechanical picture applicable to both SRS and SBS, a photon of the incident field is annihilated to create a photon at a downshifted frequency. The downshifted frequency range where new photons can be generated is approximately 30 THz in SRS and only approximately 30 MHz in SBS.

The fiber nonlinearities, including SPM, XPM, FWM as well as stimulated scattering, will start to degrade the optical signals when the optical power in fiber becomes high. An important parameter when setting up spans in optical systems is the launch power to the fiber. The power must be large enough to provide an acceptable optical signal-to-noise ratio (OSNR) at the output of the span but below the limit where excited fiber nonlinearities distort the signal. The specific limit depends on several different factors such as the type of fiber used, the bit rate, amplifier spacing, and the applied dispersion map. In dense WDM systems, the trade-off relationship between OSNR degradation by accumulation of amplified spontaneous emission (ASE) noise from optical amplifiers and nonlinear waveform-distortion in transmission fibers determines the optimum transmission power and together they limit the regenerative repeater spacing.⁷⁶

Dispersion and Nonlinearities Management

In this section, we will address the concepts of chromatic dispersion and fiber nonlinearities management followed by some examples highlighting the need for tunability to enable robust optical WDM systems in dynamic environments.

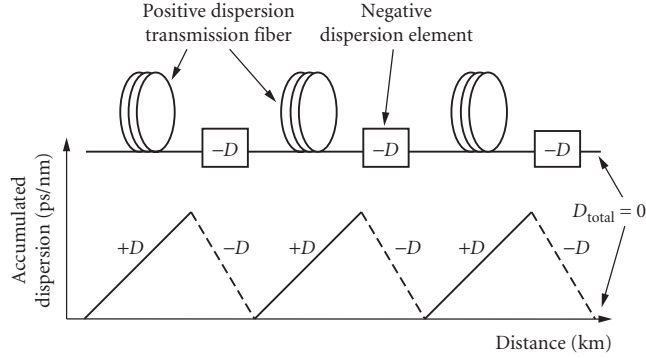


FIGURE 23 Dispersion map of a basic dispersion-managed system. Positive dispersion transmission fiber alternates with negative dispersion compensation elements such that the total dispersion is zero end-to-end.

In the preceding section, we saw that although chromatic dispersion is generally considered a negative characteristic, it is not always bad for fiber transmission. It is, in fact, a necessary evil for the deployment of WDM systems. When the fiber dispersion is near zero in a WDM system, different channels travel at almost the same speed. Any nonlinear effects that require phase matching between the different wavelength channels will accumulate at a higher rate than if wavelengths travel at widely different speeds (the case of higher dispersion fiber). Therefore, it may not be a good idea to reduce the fiber dispersion to zero by using dispersion-shifted fiber, which has both the dispersion zero and the loss minimum located at $1.55 \mu\text{m}$. As an alternative, we keep the local dispersion along the transmission link high enough to suppress nonlinear effects, while managing the total dispersion of the link to be close to zero, as shown in Fig. 23. This is a very powerful concept: at each point along the fiber the dispersion has some nonzero value, eliminating FWM and XPM, but the total dispersion at the end of the fiber link is zero, so that no pulse broadening is induced. The most advanced systems require periodic dispersion compensation, as well as pre- and postcompensation (before and after the transmission fiber).

The addition of negative dispersion to a standard fiber link has been traditionally known as “dispersion compensation,” however, the term “dispersion management” is more appropriate. Standard single-mode fiber (SMF) has positive dispersion, but some new varieties of nonzero dispersion-shifted fiber (NZDSF) come in both positive and negative dispersion varieties, as shown in Fig. 24. Reverse dispersion

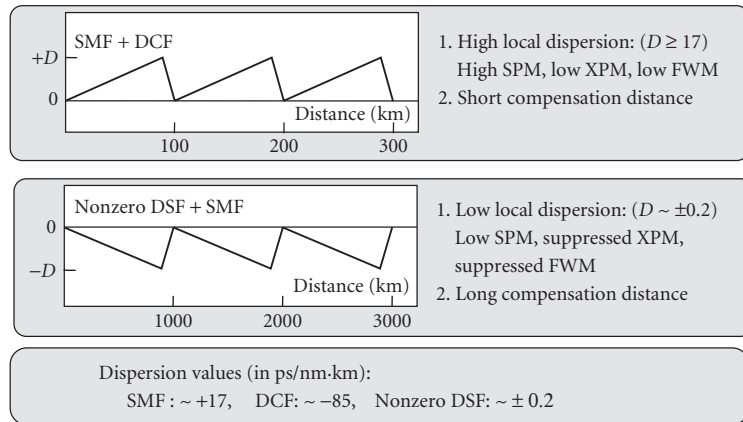


FIGURE 24 Various dispersion maps for SMF-DCF and NZDSF-SMF.

fiber is also now available, with a large dispersion comparable to that of SMF, but with the opposite sign. When such flexibility is available in choosing both the magnitude and sign of the dispersion of the fiber in a link, dispersion-managed systems can fully be optimized to the desired dispersion map using a combination of fiber and dispersion compensation devices (see Fig. 24). Dispersion is a linear process, so the first-order-dispersion maps can be understood as linear systems. However, the effects of nonlinearities cannot be ignored, especially in WDM systems with many tens of channels where the launch power may be very high. In particular, in systems deploying dispersion compensating fiber (DCF), the large nonlinear coefficient of the DCF can dramatically affect the dispersion map. We will review and highlight a few different dispersion management solutions in the following sections.

Fixed Dispersion Compensation From a systems point of view, there are several requirements for a dispersion compensating module: low loss, low optical nonlinearity, broadband (or multichannel) operation, small footprint, low weight, low power consumption, and clearly low cost. It is unfortunate that the first dispersion compensation modules, based on DCF only met two of these requirements: broadband operation and low power consumption. On the other hand, several solutions have emerged that can complement or even replace these first-generation compensators.

Dispersion compensating fiber One of the first dispersion compensation techniques was to deploy specially designed sections of fiber with negative chromatic dispersion. The technology for DCF emerged in the 1980s and has developed dramatically since the advent of optical amplifiers in 1990. DCF is the most widely deployed dispersion compensator, providing broadband operation and stable dispersion characteristics, and the lack of a dynamic, tunable DCF solution has not reduced its popularity.⁷⁷ In general, the core of the average DCF is much smaller than that of standard SMF, and beams with longer wavelengths experience relatively large changes in mode size (due to the waveguide structure) leading to greater propagation through the cladding of the fiber, where the speed of light is greater than that of the core. This leads to a large negative dispersion value. Additional cladding layers can lead to improved DCF designs that can include negative dispersion slope to counteract the positive dispersion slope of standard SMF.

In spite of its many advantages, DCF has a number of drawbacks. First of all, it is limited to a fixed compensation value. In addition, DCF has a weakly guiding structure and has a much smaller core cross-section, approximately $19 \mu\text{m}^2$, compared to the $85 \mu\text{m}^2$ (approximately) of SMF. This leads to higher nonlinearity, higher splice losses, as well as higher bending losses. Secondly, the length of DCF required to compensate for SMF dispersion is rather long, about one-fifth of the length of the transmission fiber for which it is compensating. Thus DCF modules induce loss, and are relatively bulky and heavy. The bulk is partly due to the mass of fiber, but also due to the resin used to hold the fiber securely in place. Another contribution to the size of the module is the higher bending loss associated with the refractive index profile of DCF; this limits the radius of the DCF loop to 6 to 8 inches, compared to the minimum bend radius of 2 inches for SMF.

Traditionally, DCF-based dispersion compensation modules are usually located at amplifier sites. This serves several purposes. First, amplifier sites offer relatively easy access to the fiber, without requiring any digging or unbraidings of the cable. Second, DCF has high loss (usually at least double that of standard SMF), so a gain stage is required before the DCF module to avoid excessively low signal levels. DCF has a cross section 4 times smaller than SMF, hence a higher nonlinearity, which limits the maximum launch power into a DCF module. The compromise is to place the DCF in the midsection of a two-section EDFA. This way, the first stage provides pre-DCF gain, but not to a power level that would generate excessive nonlinear effects in the DCF. The second stage amplifies the dispersion compensated signal to a power level suitable for transmission through the fiber link. This launch power level is typically much higher than the one that could be transmitted through DCF without generating large nonlinear effects. Many newer dispersion compensation devices have better performance than DCF, in particular lower loss and lower nonlinearities. For this reason, they may not have to be deployed at the midsection of an amplifier.

Chirped fiber bragg gratings Fiber Bragg gratings (FBGs) have emerged as major components for dispersion compensation because of their low loss, small footprint, and low optical nonlinearities.⁷⁸

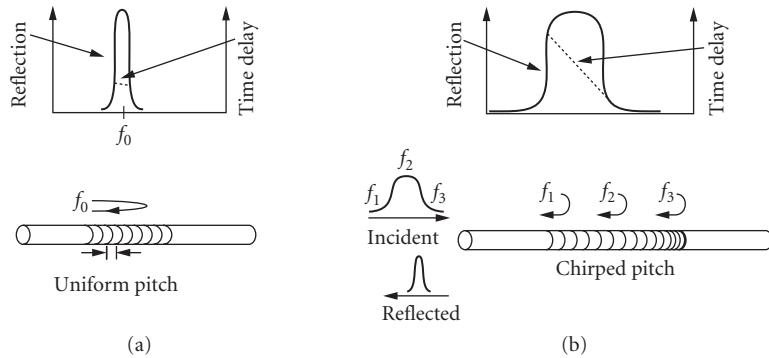


FIGURE 25 Uniform and chirped FBGs: (a) a grating with uniform pitch has a narrow reflection spectrum and a flat time delay as a function of wavelength and (b) a chirped FBG has a wider bandwidth, a varying time delay, and a longer grating length. Chirped gratings reflect different frequency components at different locations within the grating.

When the periodicity of the grating is varied along its length, the result is a chirped grating which can be used to compensate for chromatic dispersion. The chirp is understood as the rate of change of the spatial frequency as a function of position along the grating. In chirped gratings the Bragg matching condition for different wavelengths occurs at different positions along the grating length. Thus, the roundtrip delay of each wavelength can be tailored by designing the chirp profile appropriately. Figure 25 compares the chirped FBG with uniform FBG. In a data pulse that has been distorted by dispersion, different frequency components arrive with different amounts of relative delay. By tailoring the chirp profile such that the frequency components see a relative delay which is the inverse of the delay of the transmission fiber, the pulse can be compressed back. The dispersion of the grating is the slope of the time delay as a function of wavelength, which is related to the chirp. An optical circulator is traditionally used to separate the reflected output beam from the input beam.

The main drawback of Bragg gratings is that the amplitude profile and the phase profile as a function of wavelength have some amount of ripple. Ideally, the amplitude profile of the grating should have a flat (or rounded) top in the passband, and the phase profile should be linear (for linearly chirped gratings) or polynomial (for nonlinearly chirped gratings). The grating ripple is the deviation from the ideal profile shape. Considerable effort has been expended on reducing the ripple. While early gratings were plagued by more than 100 ps of ripple, published results have shown vast improvement to values close to ± 3 ps.

Ultimately, dispersion compensators should accommodate multichannel operation. Several WDM channels can be accommodated by a single chirped FBG in one of two ways: fabricating a much longer (i.e., meters-length) grating, or using a sampling function when writing the grating, thereby creating many replicas of transfer function of the FBG in the wavelength domain.⁷⁹

Tunable Dispersion Compensation

The need for tunability In a perfect world, all fiber links would have a known, discrete, and unchanging value of chromatic dispersion. Network operators would then deploy fixed dispersion compensators periodically along every fiber link to exactly match the fiber dispersion. Unfortunately, several vexing issues may necessitate that dispersion compensators have tunability, that is, they have the ability to adjust the amount of dispersion to match system requirements.

First, there is the most basic business issue of inventory management. Network operators typically do not know the exact length of a deployed fiber link nor its chromatic dispersion value. Moreover, fiber plants periodically undergo upgrades and maintenance, leaving new and nonexact lengths of fiber behind. Therefore, operators would need to keep in stock a large number

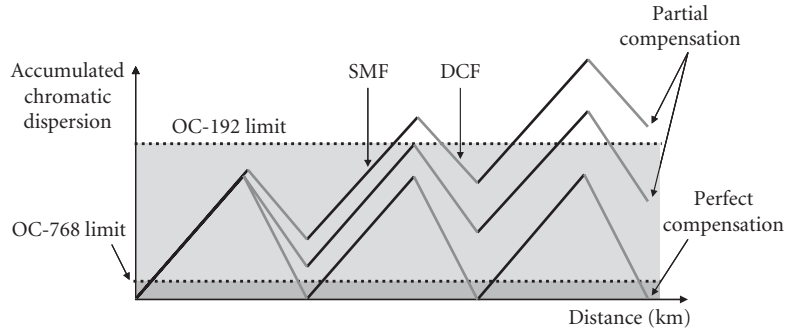


FIGURE 26 The need for tunability. The tolerance of OC-768 systems to chromatic dispersion is 16 times lower than that of OC-192 systems. Approximate compensation by fixed in-line dispersion compensators for a single channel may lead to rapid accumulation of unacceptable levels of residual chromatic dispersion.

of different compensator models, and even then the compensation would only be approximate. Second, we must consider the sheer difficulty of 40 Gb/s signals. The tolerable threshold for accumulated dispersion for a 40 Gb/s data channel is 16 times smaller than that at 10 Gb/s. If the compensation value does not exactly match the fiber to within a few percent of the required dispersion value, then the communication link will not work. Tunability is considered a key enabler for this bit rate (see Fig. 26). Third, the accumulated dispersion changes slightly with temperature, which begins to be an issue for 40 Gb/s systems and 10 Gb/s ultralong haul systems. In fiber, the zero-dispersion wavelength changes with temperature at a typical rate of 0.03 nm/°C. It can be shown that a not-uncommon 50°C variation along a 1000-km 40-Gb/s link can produce significant degradation (see Fig. 27). Fourth, we are experiencing the dawn of reconfigurable optical networking.

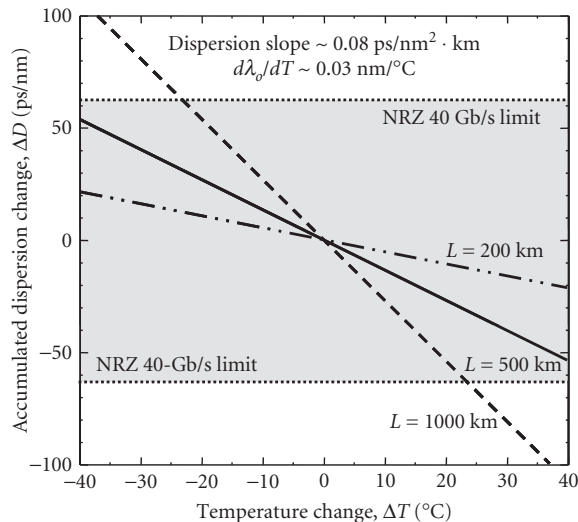


FIGURE 27 Accumulated dispersion changes as a function of the link length and temperature fluctuation along the fiber link.

In such systems, the network path, and therefore the accumulated fiber dispersion, can change. It is important to note that even if the fiber spans are compensated for span-by-span, the pervasive use of compensation at the transmitter and receiver suggests that optimization and tunability based on path will still be needed.

Other issues that increase the need for tunability include (1) laser and (de)mux wavelength drifts for which a data channel no longer resides on the flat-top portion of a filter, thereby producing a chirp on the signal that interacts with the fiber's chromatic dispersion, (2) changes in signal power that change both the link's nonlinearity and the optimal system dispersion map, and (3) small differences that exist in transmitter-induced signal chirp.

Approaches to tunable dispersion compensation A host of techniques for tunable dispersion compensation have been proposed in recent years. Some of these ideas are just interesting research ideas, but several have strong potential to become viable technologies. We will discuss FBG-based technology as an example.

If a FBG has a refractive-index periodicity that varies nonlinearly along the length of the fiber, it will produce a time delay that also varies nonlinearly with wavelength (see Fig. 28). Herein lays the key to tunability. When a linearly chirped grating is stretched uniformly by a single mechanical element, the time delay curve is shifted toward longer wavelengths, but the slope of the ps versus nm curve remains constant at all wavelengths within the passband. When a nonlinearly chirped grating is stretched, the time delay curve is shifted toward longer wavelengths, but the slope of the ps versus nm curve at a specific channel wavelength changes continuously.⁸⁰

Another solution was also reported, which is based on differential heating of the substrate. The thermal gradient induced a chirp gradient, which could be altered electrically⁸¹ and has a major advantage: no moving parts. However, this is countered by the disadvantage of slow tuning, limited to seconds or minutes. Additionally, the technology requires accurate deposition of a thin film of tapered thickness. The process of deposition of the tapered film seems to have some yield issues,

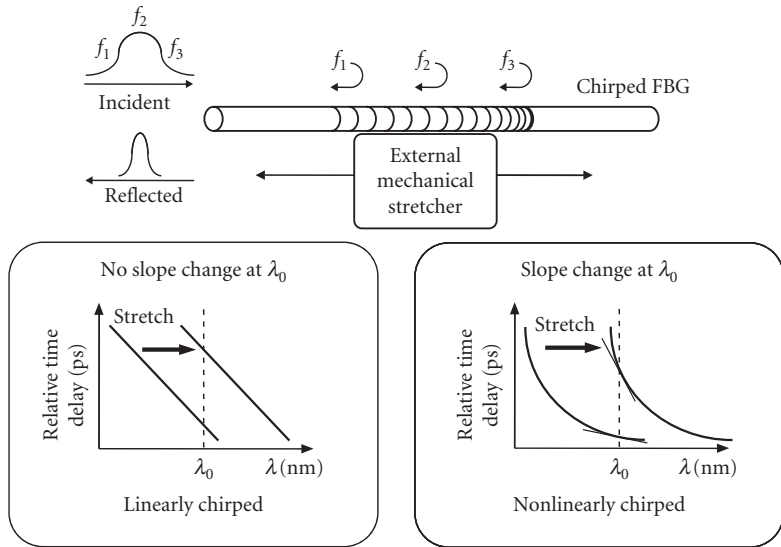


FIGURE 28 Tuning results for both linearly and nonlinearly chirped FBGs using uniform stretching elements. The slope of the dispersion curve at a given wavelength λ_0 is constant when the linearly chirped grating is stretched, but changes as the nonlinearly chirped grating is stretched.

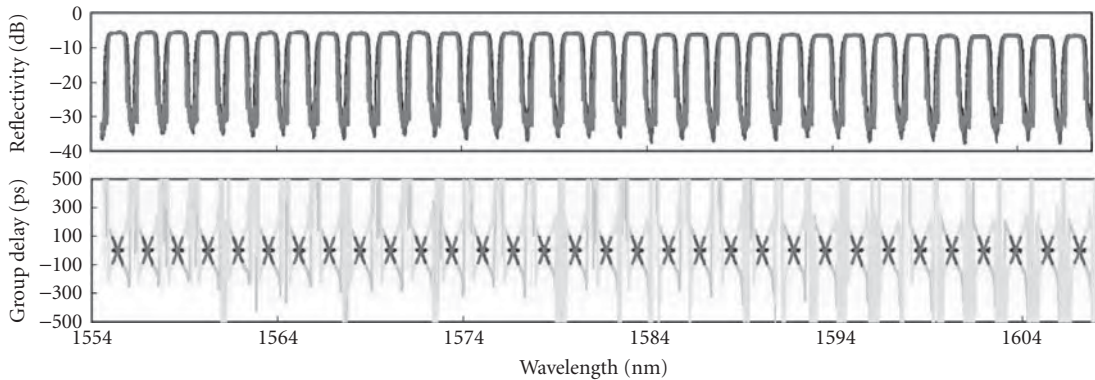


FIGURE 29 Thirty-two-channel, 100-GHz channel spacing, FBG-based tunable dispersion compensator made by Teraxion Inc.⁸²

making it rather difficult to manufacture. A 32-channel, 100-GHz channel spacing, FBG-based tunable dispersion compensator is demonstrated recently, with ± 400 ps/nm range.⁸² The parameters of the compensator are shown in Fig. 29. We can see that the tunable dispersion compensator exhibits uniform channel profiles, with flat top, steep edges, and low crosstalk.

Although currently no technology is a clear winner, the trend of dispersion compensation is toward tunable devices, or even actively self-tunable compensators, and such devices will allow system designers to cope with the shrinking system margins and with the emerging rapidly reconfigurable optical networks.

Electronic Solutions It is worth to mention that some of the most promising solutions of dispersion are electronic signal processing such as electronic equalizers and forward error correction (FEC) coding.^{83,84} The electronic equalizers rely on post-detection signal processing including filtering and adaptive signal processing to sharpen up distorted data pulses. Because the detection itself is nonlinear, the job of compensating for the linear distortions of chromatic dispersion is quite a bit more complicated. Many of the high-performance 40 Gb/s systems also incorporate FEC coding. Such coding adds some redundancy into the bits of a data stream to more easily find and correct errors. FEC is implemented using electronic chips, and it adds a system power margin that can ease the deleterious problems associated with fiber nonlinearities, chromatic dispersion, signal-to-noise ratio, and PMD. Note that electronic processing is potentially very cheap, and much easily scalable to large volume production, at least at data less than 10 Gb/s.

Analog equalizers that can combat the distortion produced by chromatic dispersion and PMD have been demonstrated at bit rates up to 43 Gb/s.⁸⁵ Hardware-implemented maximum likelihood sequence estimation (MLSE) has been demonstrated to reduce penalties from intersymbol interference and to extend the chromatic-dispersion-limited transmission length.^{86,87} Using electrical pre-equalization, 10-Gb/s transmission is demonstrated over 5120-km SMF without optical dispersion compensation.⁸⁸ The electrically precompensating transmitter is shown in Fig. 30a and Fig. 30b shows the BER measurement after transmission.

Another greatest trend in electronic signal processing for optical transmission is digital sampling and signal processing techniques for optical receivers.⁸⁹ Digital sampling techniques move several of the largest problems in optical transmission into the electrical domain. High-speed sampling followed by offline processing has been used in most experimental demonstrations of coherent detection with digital signal processing.^{90–93} It is still challenging for the benefits of coherent detection and digital signal processing to outweigh the cost of implementing real-time processing at high data rates.

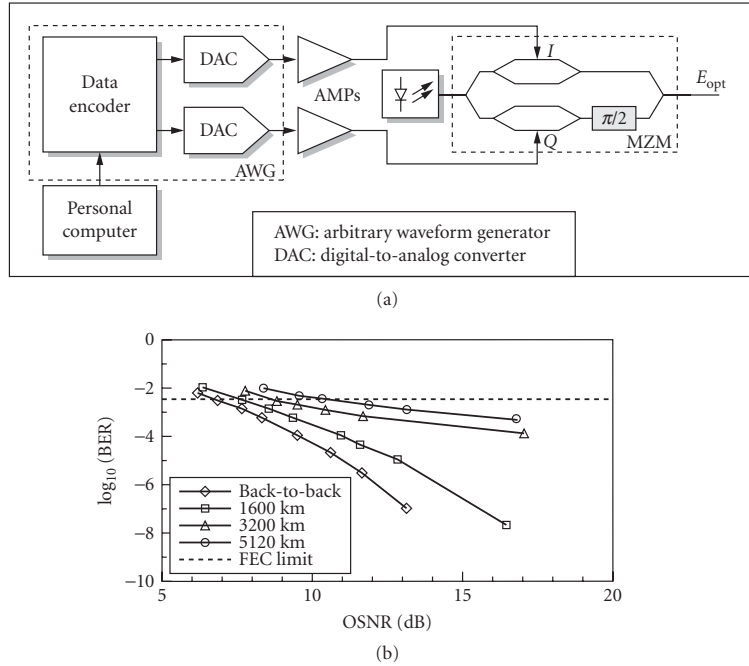


FIGURE 30 Ten Gb/s transmission over 5120-km SMF without optical dispersion compensation using electrical preequalization: (a) electrically precompensating transmitter and (b) BER versus OSNR after transmission.⁸⁸

21.4 OPTICAL MODULATION FORMATS FOR WDM SYSTEMS

Most of current fiber systems use binary modulation with error-control coding schemes. The spectral efficiency cannot exceed $1 \text{ b/s} \cdot \text{Hz}$ per polarization regardless of detection technique. With the increase of bit rate and decrease of channel spacing in WDM systems, higher spectral efficiency is required. To achieve spectral efficiencies above $1 \text{ b/s} \cdot \text{Hz}$ and increase overall capacity of a WDM transmission system, more advanced modulation formats will be needed. The types of data modulation formats also have substantial impacts on fiber impairments. Due to the fact that optical signals propagating in fibers offer several degrees of freedom, including amplitude, frequency, phase, polarization, and time, intense research efforts have been made toward the combination coding over these degrees of freedom as a means to increase fiber transmission capacity, especially as a way to combat or benefit from fiber impairments. In this section, we will discuss the optical modulation formats of digital signals in WDM systems. We will highlight a few examples and present their advantages and disadvantages based on fiber system performance characterization.

Basic Concepts

The digital signal, which may be modulated at approximately gigabits per second rates, is being transmitted on an optical carrier wave whose frequency is in the multiterahertz regime. This optical carrier wave, $A(t)$, has an intensity amplitude, A_0 , an angular frequency, ω_c , and a phase ϕ .⁹⁴

$$A(t) = A_0 \cos(\omega_c t + \phi) \quad (8)$$

A binary digital signal implies transmitting two different quantities of anything which can subsequently be detected as representing a “1” and “0,” that is, we can transmit blue and red, and this can represent “1” and “0” in the receiver electronics if blue and red can be distinguished. We can therefore modulate either the amplitude, frequency, or phase of the optical carrier between two different values to represent either a “1” or “0,” known respectively as amplitude-, frequency-, and phase-shift keying (ASK, FSK, and PSK), with the other two variables remaining constant:

$$\begin{aligned} A_{\text{ASK}}(t) &= [A_0 + m(\Delta A)]\cos(\omega_c t + \phi) & m &= \begin{cases} +1, & \text{“1”} \\ -1, & \text{“0”} \end{cases} \\ A_{\text{FSK}}(t) &= A_0 \cos\{[\omega_c + m(\Delta\omega)]t + \phi\} \\ A_{\text{PSK}} &= A_0 \cos[\omega_c t + (\phi + m\pi)] \end{aligned} \quad (9)$$

where ΔA is the amplitude modulation and is less than A_0 , and $\Delta\omega$ is the FSK frequency deviation. ASK has two different light amplitude levels; FSK has two different optical carrier wavelengths; and PSK has two different phases which can be detected as an amplitude change in the center of the bit time for which a “1” or “0” bit can be determined. It is important to emphasize that the differential-phase-shift-keying (DPSK) format, in which the phase of the preceding bit is used as a relative phase reference, has been reemerged in the last few years due to its less OSNR requirement and robustness to fiber nonlinearities.^{95–97} The DPSK modulation signal is not the binary code itself, but a code that records changes in the binary stream. The PSK signal can be converted to a DPSK signal by the following rules: a “1” in the PSK signal is denoted by no change in the DPSK, a “0” in the PSK signal is denoted by a change in the DPSK signal. For a DPSK signal, optical power appears in each bit slot, and can occupy the entire bit slot (NRZ-DPSK) or can appear as an optical pulse (RZ-DPSK).

Figure 31 shows the impression of a simple digital signal on the optical carrier. These three formats can be implemented by appropriately changing the optical source, whether by modulating the

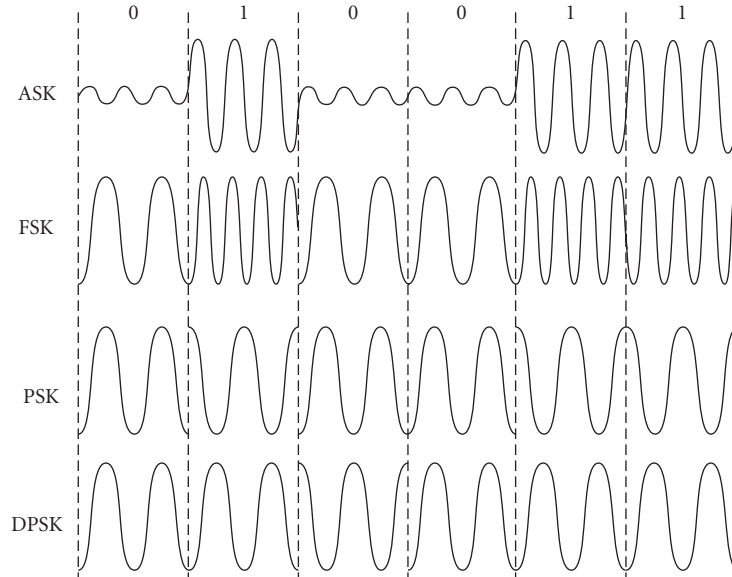


FIGURE 31 ASK, FSK, and PSK (DPSK) time modulation while employing an optical carrier wave.

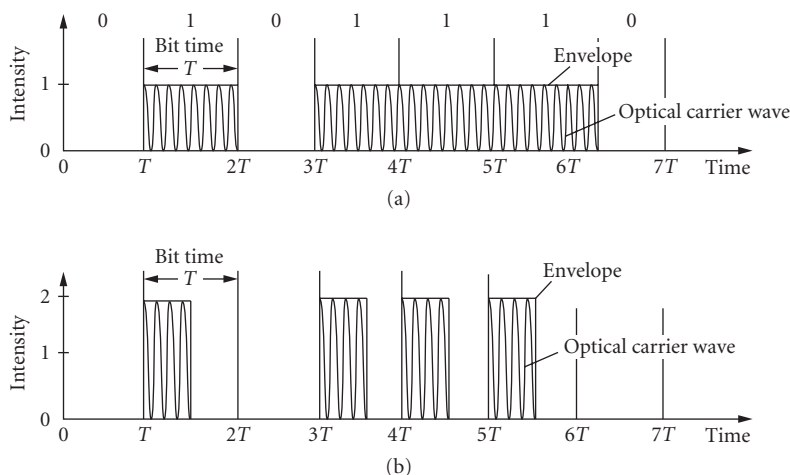


FIGURE 32 Modulation formats: (a) NRZ (non-return-to-zero) modulation format and (b) RZ (return-to-zero) modulation format.

light amplitude, laser output wavelength, or using an external phase shifter. ASK is important since it is the simplest to implement, FSK is important because a smaller chirp is incurred when direct modulation of a laser is used, and PSK is important because it, in theory, requires the least amount of optical power to enjoy error-free data recovery.⁹⁸

It should be mentioned that ASK, which is by far the most common form, is called on-off-keying (OOK) if the “0” level is really at zero amplitude. Figure 32 shows NRZ and RZ OOK. NRZ is the simplest format in which the amplitude level is high during the bit-time if a “1” is transmitted and is low if a “0” is transmitted. RZ format requires that a “1” always return to the low state during the bit-time even when two “1”s are transmitted in sequence, whereas the “0” bit remains at a low level. This eliminates the possibility of a long string of “1”s producing a constant high level but does not eliminate a long string of “0”s from producing a constant low level. The main attraction of the RZ format is its demonstrated improved immunity to fiber nonlinearities relative to NRZ. Note that the frequency of the optical carrier is so high that the optical detector, whose electronic bandwidth is usually $\ll 100$ GHz, will not detect it and only the envelope of the gigabits per second data will be electrically recovered.

Once a signal has been modulated and transmitted, it must be accurately detected. Two detection schemes include direct detection and coherent detection.^{99,100} Direct detection can be used for ASK and FSK modulation. PSK generally requires complicated coherent detection, which is based on mixing of two light waves to detect the phase information.

Direct detection, which is by far the simpler of the two schemes, involves detecting the amount of optical power incident on the optical detector. Direct detection of an ASK (or OOK) signal, that is, light “on” or light “off,” is extremely simple to accomplish using a detector and a high-bandwidth power meter (see Fig. 33a). This ASK signal is recovered by a detector of a certain electrical low-pass-filtering bandwidth. The electrical spectrum of the recovered ASK signal, which is sent in random NRZ format and can be measured on an RF spectrum analyzer, is shown in Fig. 33b. Only the first lobe of RF spectrum is necessary for recovering the data since only the transition edges would be affected by cutting off the higher lobes. The first lobe is considered the baseband signal representing the data stream.¹⁰¹ It is important to mention that the sensitivity of a direct detection receiver can be improved significantly by using a low noise optical preamplifier that is just before the photodetector. The output of the optical amplifier needs to be high enough so that at the photodetector the noise is dominated by the signal-spontaneous beat noise of the optical preamplifier.

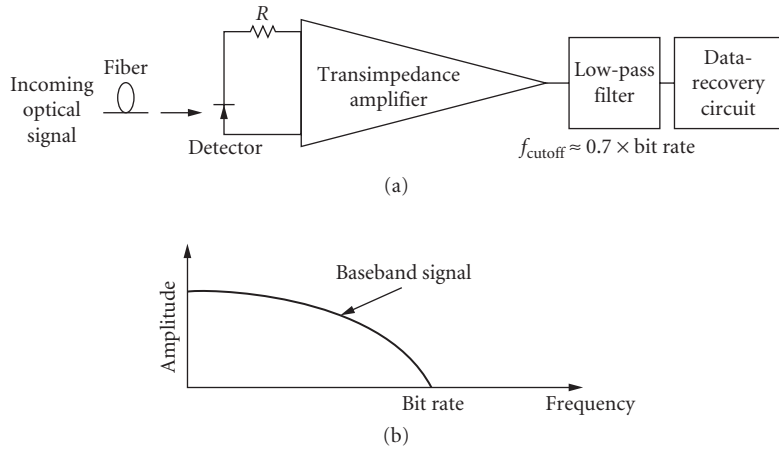


FIGURE 33 (a) Direct detection optical system and (b) baseband signal of a directly-detected NRZ signal.

Carrier-Suppressed Return-to-Zero and Duobinary

It is important to note that the type of data modulation formats have substantial impacts on the fiber dispersive and nonlinear effects. Some of these formats carry information through OOK, but also modulate the optical phase in a noninformation-bearing way in order to enhance the signals' robustness to chromatic dispersion, optical filtering, and/or nonlinearities. This group includes formats such as optical duobinary, chirped return-to-zero (CRZ), and alternating-phase OOK formats such as carrier-suppressed return-to-zero (CS-RZ). We will highlight a few examples here and present their advantages and disadvantages based on fiber system performance characterization.

CS-RZ optical signals have the feature of presenting bits that are π phase-shifted relative to neighboring bits such that on average all the phases cancel each other out for a net phase of zero. A CS-RZ optical data stream may consist of a plurality of pulses where half of the pulses have an alternating phase relationship with the other half of the pulses, which leads to carrier suppression, as shown in Fig. 34a. Because of this phase inversion between adjacent bit periods that reduces interbit interference, CS-RZ signals show increased tolerance dispersion and nonlinear penalties (see Fig. 34b).¹⁰²

The CS-RZ format can be generated by using a sinusoidal signal to drive the Mach-Zehnder modulator; the drive frequency is one half of the bit rate ($f_r/2$) and the amplitude is $2V_\pi$. As shown in Fig. 35, the generated pulse has the repetition rate of f_r , and the phase of the pulses alternates between 0 and π .

Optical duobinary has been proven more resilient to dispersion for more than 10 Gb/s data and is reasonably simple to implement. As shown in Fig. 36, optical duobinary signal is commonly generated by applying a baseband duobinary (three-level) electrical signal to the RF input of a lithium niobate (LiNbO_3) Mach-Zehnder modulator biased at maximum distinction (at V_π). The result is a binary, intensity-modulated optical signal, with a duobinary-modulated optical electric field caused by π -radian shift. With a zero ($c_k = 1$) input, no light is transmitted, but the $+1$ ($c_k = 2$) and -1 ($c_k = 0$) inputs are transmitted as $+E$ and $-E$ electrical fields. While this is a three-level signal in terms of the electric field, it is a two-level signal in terms of optical power. The same receiver that is used for a NRZ modulation scheme can be used for duobinary modulation. The power detector squares the electric field to detect power and hence the $+E$ and $-E$ outputs of the fiber get mapped to the same power level and are detected as logical 1s. Note that the received data is the invert of the original binary input. This choice significantly reduces the complexity of the receiver (the first optical duobinary system used a mapping that requires three levels of optical power). One of the key components is a driver that can produce a voltage swing of $2V_\pi$ V at high data rates such as more than 10 Gb/s in order to drive the Mach-Zehnder modulator.

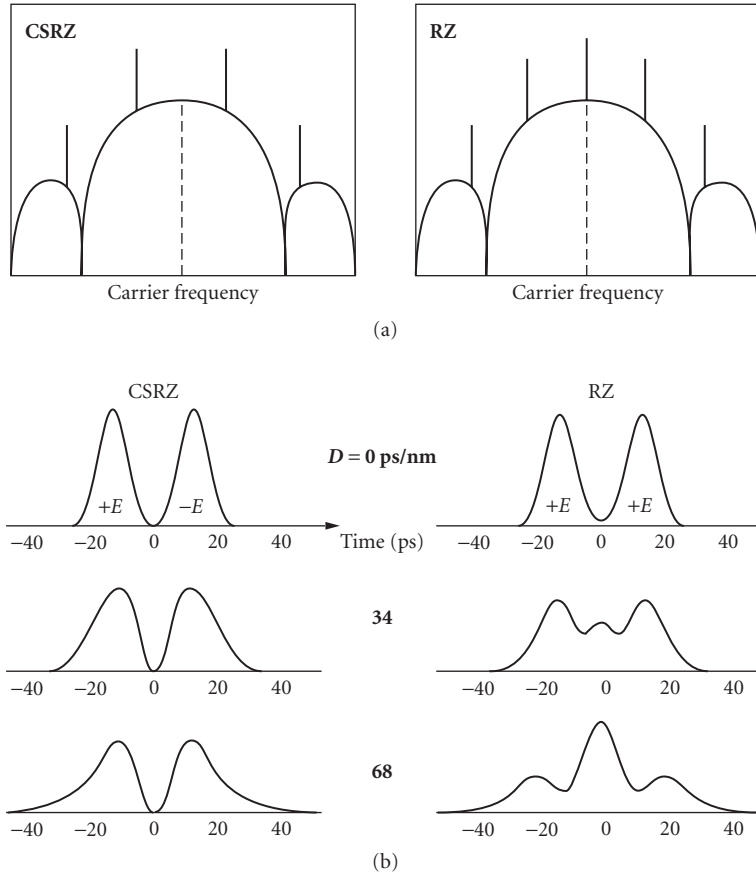


FIGURE 34 (a) Optical spectra of CS-RZ and RZ formats and (b) 40 Gb/s CS-RZ and RZ pulses under different dispersion values.¹⁰²

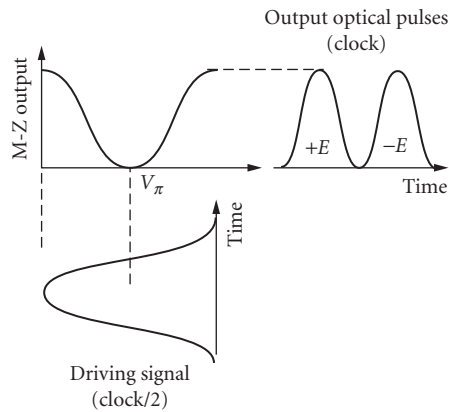


FIGURE 35 Generation of CS-RZ format using Mach-Zehnder modulator.

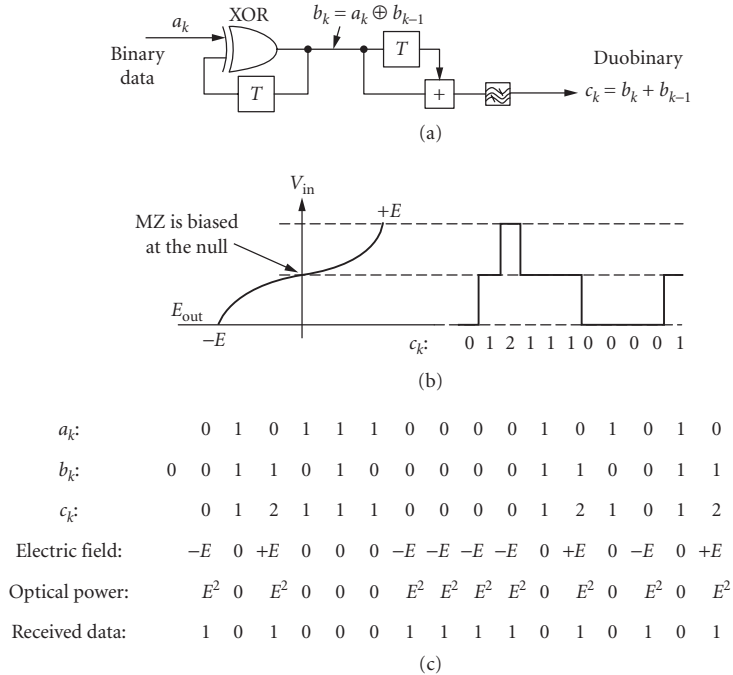


FIGURE 36 Optical duobinary modulation format: (a) duobinary encoder; (b) biasing of Mach-Zehnder modulator; and (c) an example of data transformation.

The combination of the duobinary encoder and the above mapping of electric fields help reduce the effects of dispersion in the fiber. The pulses spread out as they travel down the fiber. In an NRZ scheme, a data sequence of 1 0 1 is mapped onto the optical domain as $+E$ 0 $+E$. In the encoded duobinary sequence, a 1 0 1 sequence cannot occur, but a 1 0 -1 does occur, which is mapped to $+E$ 0 $-E$ in the optical domain. The effect of dispersion in the two cases is shown in Fig. 37, which depicts why the

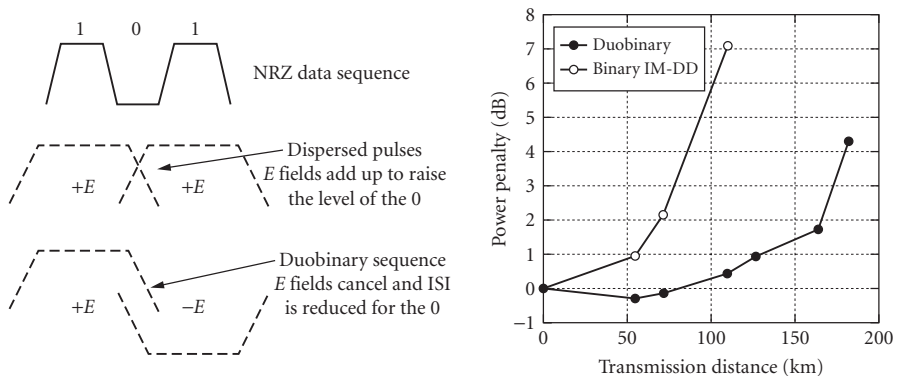


FIGURE 37 Dispersion effects on NRZ format and duobinary format. The power penalty due to fiber chromatic dispersion was measured at the BER of 10^{-9} .¹⁰³

resulting dispersion is less in the case of duobinary modulation. Figure 37 also shows an experimental result that optical duobinary technique expands the transmission distance to more than 150 km SMF for 10 Gb/s data.¹⁰³

The optical duobinary technique has been proven in experiment to expand the usable bandwidth and transmission distance in a four-channel multiplexed 40-Gb/s-based WDM system without individual channel dispersion slope compensation. That 160-Gb/s WDM transmission using four 40-Gb/s optical duobinary channels over a 100-km DSF is successful by virtue of the high dispersion tolerance of the optical duobinary signal.¹⁰⁴ Furthermore, the narrow optical spectrum of optical duobinary signals provides high spectral efficiency and reduces the coherent crosstalks in the ultradense WDM system.¹⁰⁵

DPSK and DQPSK

With the transmission capacity increases in WDM systems, PSK systems regain much interest in recent years. As we mentioned in the early section, PSK formats carry the information in the optical phase itself and DPSK formats carry the information in optical phase changes between bits. Since photodetector is inherently insensitive to the optical phase; a detector only converts the optical signal power into an electrical signal, directly detecting PSK signal is impractical due to the lack of an absolute phase reference. Therefore, PSK systems generally require complicated coherent detection which needs a local oscillator (laser) to mix with the received signal light.¹⁰⁶ However, we can detect the DPSK signal using a 1-bit delayed Mach-Zehnder interferometer followed by a balanced direct-detection system. A DPSK signal records the phase changes in the binary stream. Thus the demodulator only needs to determine these changes in the coming signal phase. A typical balanced DPSK receiver is shown in Fig. 38. The optical DPSK signal is first sent to a Mach-Zehnder interferometer (MZI) with the 1-bit period differential delay between the two arms. The MZI lets two adjacent bits interfere with each other at its output ports. This interference leads to the presence (absence) of power at an MZI output port if two adjacent bits interfere constructively (or destructively) with each other. Thus, the preceding bit in a DPSK-encoded bit stream acts as the phase reference for the

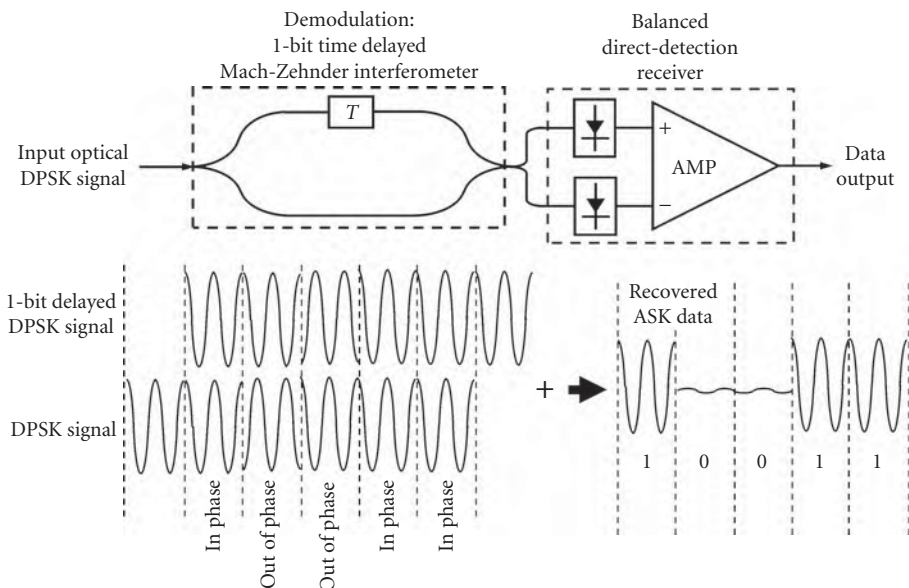


FIGURE 38 A typical DPSK receiver.

current bit. No phase change between these two bits will generate a “1” (or “0”) at the constructive port (or destructive port). Recall the generation of DPSK from PSK signals, this process is exactly the reverse process. Ideally, one of the MZI output ports is adjusted for destructive interference in the absence of phase modulation (*destructive port*), while the other output port then automatically exhibits constructive interference due to energy conservation (*constructive port*). For the same reason, the two MZI output ports will carry identical, but logically inverted data streams under DPSK modulation.⁹⁵

Using the balanced direct-detection scheme, the DPSK system has the advantage of requiring a lower OSNR than OOK to reach a given BER. Intuitively, this can be understood by comparing the signal constellations for DPSK and OOK.⁹⁹ To achieve the same symbol distance, the average optical power in DPSK is only half as compared to OOK. At 40 Gb/s, a sensitivity of about 38 photons/bit has been reported using RZ-DPSK.¹⁰⁷ This is approximately 3 dB better than the best OOK results of 78 photons/bit.¹⁰⁸ The lower OSNR requirement of DPSK can be used to extend transmission distance, reduce optical power requirements, or relax component specifications. Both numerical simulations and experiments have also shown DPSK to be more robust to some nonlinear effects than OOK⁹⁷ due to the following facts: (1) the optical power is evenly distributed (power is present in every bit slot for DPSK, which reduces bit-pattern-dependent nonlinear effects), and (2) the optical peak power is 3 dB lower for DPSK than for OOK for the same average optical power. Figure 39 shows the simulation results of comparison of the transmission of 43 Gb/s signals, NRZ, CSRZ and RZ-DPSK.¹⁰⁹

To further increase the bit rates and spectral efficiency in WDM systems, an extension to differential quadrature phase-shift keying (DQPSK) is introduced.⁹⁵ It transmits the four phase shifts $\{0, \pi/2, -\pi/2, \pi\}$ at a symbol rate of half the aggregate bit. DQPSK requires relatively complicated transmitter and receiver. As shown in Fig. 40, the transmitter consists of two parallel DPSK modulators that are integrated together in order to achieve phase stability. The receiver essentially consists of two DPSK receivers, although the phase difference in the arms of the delay interferometers is now set to $\pi/4$ and $-\pi/4$.

As compared to DPSK, the required OSNR to reach a given BER is increased by about 1 to 2 dB, depending on the BER.¹¹⁰ Also, the frequency offset tolerance between the laser and the delay interferometer is about 6 times less than for DPSK,¹¹¹ making the delay interferometer design and stabilization somewhat challenging. The benefit of DQPSK is that, for the same data rate, the symbol rate is reduced by a factor of two. Consequently, the spectral occupancy is reduced, the transmitter and receiver bandwidth requirements are reduced, and the chromatic dispersion and PMD limitations are

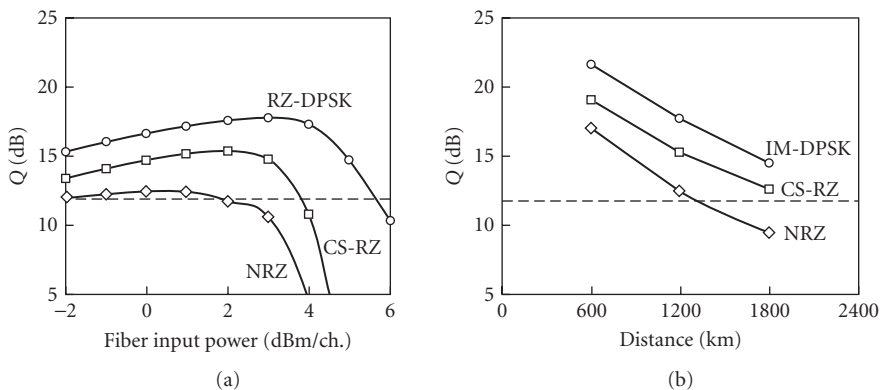


FIGURE 39 Comparison of 43 Gb/s signals, NRZ, CSRZ, and RZ-DPSK, with channel spacing of 75 GHz, (a) after 1200 km NZ-DSF (dispersion = 8 ps/nm · km) transmission and (b) at the optimum fiber input power. The broken line indicates the 11.8-dB Q-factor that corresponds a BER of 10^{-15} after Reed Solomon (255, 239) forward error correction.¹⁰⁹

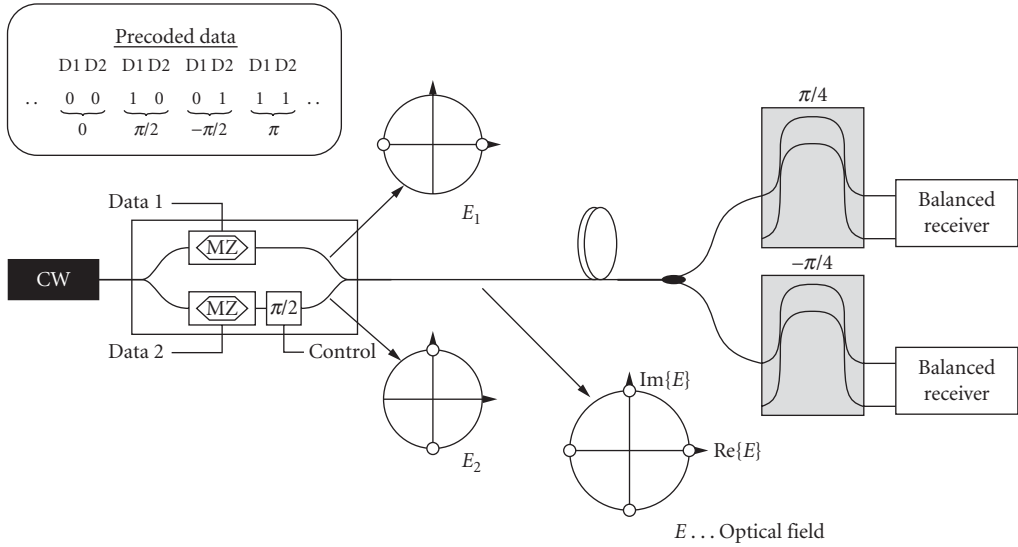


FIGURE 40 A typical DQPSK transmitter and receiver.⁹⁵

extended. The tolerance of DQPSK signal to PMD enables polarization-multiplexed data transmission to increase the spectral efficiency further. As shown in Fig. 41, a recent experiment demonstrated of $1.6\text{-b/s} \cdot \text{Hz}$ spectrally efficient transmission over 1700-km single-mode fiber, using 40-channel 85.6-Gb/s (totally 3.2-Tb/s) polarization-multiplexed RZ-DQPSK.¹¹²

Even higher spectral efficiency can be achieved using various combinations of phase- and amplitude-shift keying.^{113–115} Such multilevel modulation can also improve system tolerance to chromatic dispersion and PMD. However, these schemes quickly become quite complicated to implement, require higher OSNR, and are sensitive to nonlinear phase noise.

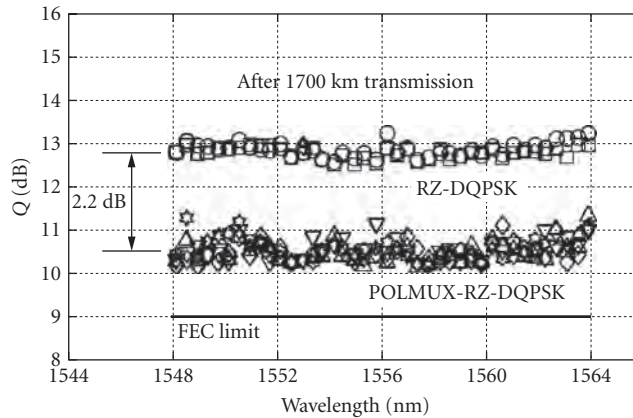


FIGURE 41 Experimental demonstration of $1.6\text{-b/s} \cdot \text{Hz}$ spectrally efficient transmission over 1700-km single-mode fiber, using 40-channel 85.6-Gb/s (totally 3.2-Tb/s) polarization-multiplexed RZ-DQPSK.¹¹²

TABLE 1 Overview of Modulation Formats and Some Performance Values at 42.7 Gb/s (Required OSNR at BER = 10^{-3})¹¹⁶

Modulation Format	TX Complexity	RX Complexity	Required OSNR (dB)	CD (ps/nm) (2-dB Penalty)	DGD (ps) (1-dB Penalty)
NRZ-OOK	1 MZM	1 PD	15.9	54	8
50% RZ-OOK	1–2 MZMs	1 PD	14.4	48	10
67% CSRZ-OOK	2 MZMs	1 PD	14.9	42	11
Duobinary	1 MZM	1 PD	16.6	211	6
NRZ-DPSK	1 MZM	1 DI + 2 PDs	11.7	74	10
50% RZ-DPSK	1–2 MZMs	1 DI + 2 PDs	11.1	50	10
NRZ-DQPSK	2 nested MZMs	2 DI + 4 PDs	13.2	168	20
50% RZ-DQPSK	2 nested MZMs	2 DI + 4 PDs	12.2	161	21

PD: photodiode; OF: optical filter; DI: delay interferometer; MZM: Mach–Zehnder modulator; PC: pulse carver; CD: chromatic dispersion; DGD: differential group delay.

Table 1 gives an overview of some key characteristics of the optical modulation formats.¹¹⁶ The second and third columns summarize the transmitter and receiver hardware complexities, respectively, in terms of the optoelectronic component requirements. The fourth column specifies required OSNR for BER equal to 10^{-3} based on BER simulations that properly take into account the non-gaussian noise statistics of beat-noise-limited detection. The assumed 42.7 Gb/s are representative of a 40-Gb/s per-channel bit rate, including a 7 percent overhead for FEC, as standardized for terrestrial fiber transmission systems. The fifth column quantifies the accumulated chromatic dispersion that yields a 2-dB penalty in OSNR, and the sixth column quantifies the tolerance (1-dB OSNR penalty) of different modulation formats to the first-order PMD.

While actually measured values for required OSNR may differ somewhat from the numbers given in Table 1 due to various optical and electronic hardware implementation aspects, some general facts are

1. RZ formats in general require 1 to 3 dB less OSNR for identical BER than their NRZ equivalents, which is mostly due to the reduced impact of intersymbol interference (ISI) on RZ formats.
2. Using DPSK instead of intensity modulation, OSNR requirements are significantly reduced. The gain of balanced-detection DPSK over OOK is generally independent of the target BER and typically amounts to around 3 dB. Depending on the modulation waveforms, extinction ratios, and optical as well as electrical filters, the gain of DPSK can also even exceed 3 dB.
3. DQPSK requires only 1 to 1.5 dB higher OSNR than DPSK at poor BER (e.g., 10^{-3}), though the OSNR gap between DPSK and DQPSK increases at good BER (e.g., 10^{-12}).¹¹⁷ The good OSNR performance at FEC error ratios makes DQPSK an attractive candidate for optically routed networks that require narrow optical signal spectra.¹¹⁸
4. Duobinary and DQPSK have significantly better dispersion tolerance than other formats due to their narrower spectra.
5. For most modulation formats, a 1-dB penalty occurs at a DGD between 30 and 40 percent of the symbol duration, with RZ formats being in general more resilient to PMD than NRZ formats.¹¹⁹ Since the tolerance to the first-order PMD scales linearly with symbol duration, DQPSK has about twice the PMD tolerance of binary modulation formats at the same bit rate.

Table 2 shows a summary comparison of the various modulation formats relative to NRZ-OOK.¹²⁰ Note that the performance of a modulation format may depend significantly on its implementation details. The choice of the most suitable format depends on the application (metropolitan, regional, or long-haul), bit rate, wavelength spacing, optical power level, fiber type and dispersion map, the number of pass-through nodes and associated multiplexing filtering technique, the amplification scheme, as well as other system requirements.¹²¹ The “optimum” implementation characteristics of a modulation format often depend on a subtle interplay of several parameters and require extensive studies to maximize system performance.¹²²

TABLE 2 Comparison of the Various Modulation Formats Relative to NRZ-OOK at 40 Gb/s¹²⁰

Mod. Format Performance vs. NRZ	CS-RZ	Duobinary	DPSK	DQPSK	DPSK-RZ	DQPSK-RZ
OSNR sensitivity	Slightly better	Slightly worse	Much better	Slightly better	Much better	Better
CD tolerance and spectral efficiency	Slightly worse	Much better	Slightly better	Much better	Slightly worse	Much better
PMD tolerance	Better	Equivalent	Slightly better	Much better	Better	Much better
Nonlinearity tolerance	Better	Equivalent	Better	Equivalent	Much better	Equivalent
Cost and complexity	Slightly worse	Equivalent	Slightly worse	Much worse	Worse	Much worse

21.5 OPTICAL AMPLIFIERS IN WDM NETWORKS

The optical amplifier is ideally a transparent box which provides gain and is also insensitive to the bit-rate, modulation-format, power, and wavelengths of the signal(s) passing through it. The signals remain in optical form during amplification. Optical amplifiers have played a key role in the optical telecommunications world due to rapid device progress and revolutionary systems results.¹²³ In fact, much of advances in optical communications can be traced to the incorporation of optical amplifiers.

EDFA is the most widely used optical amplifier. In this section, we will consider some important issues about EDFAs with regards to their implementation in WDM systems. EDFAs have been used in multichannel WDM systems to compensate for: (1) fiber attenuation losses in transmission, (2) component excess losses, and (3) optical network splitting losses. These optical splitting losses can occur in a passive star, in which the optical power is divided by the number of users (N), or in a ring/bus in which there may possibly be optical tapping losses at each node.

Figure 42a shows WDM transmission in a conventional electrically regenerated system. Regenerators can correct for fiber attenuation and chromatic dispersion by detecting an optical signal and then

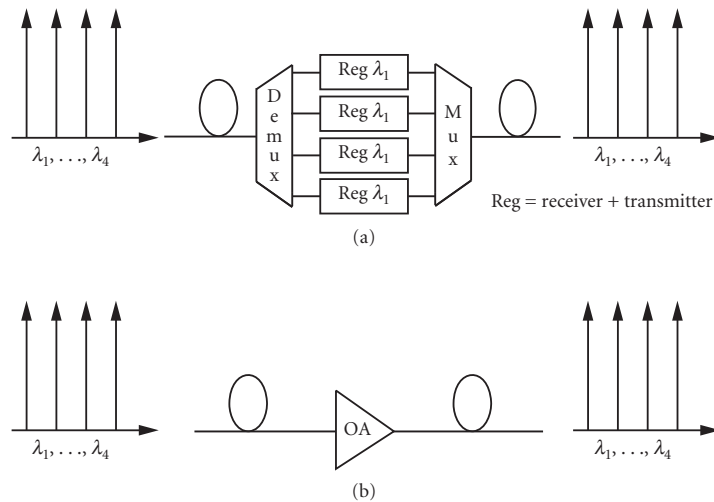


FIGURE 42 Wideband amplifiers enable WDM: (a) regeneration and (b) optical amplification.

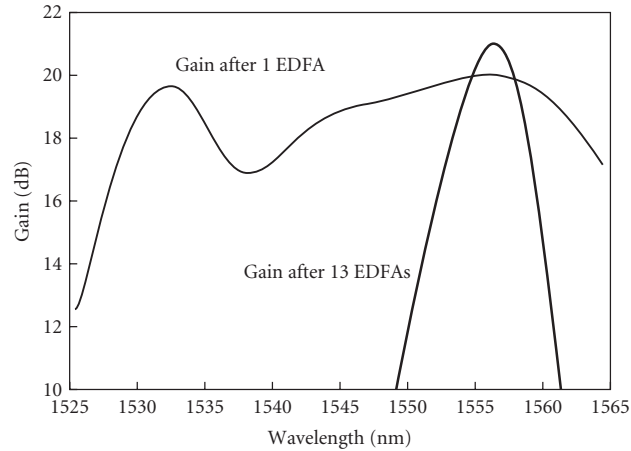


FIGURE 43 EDFA gain nonuniformity accumulation.

retransmitting it as a new signal using an internal laser. However, regenerators (being a hybrid of optics and electronics) are expensive, bit-rate and modulation-format specific, and waste much power and time in converting from photons to electrons and back again to photons. In contrast, as shown in Fig. 42b, the EDFA is ideally a transparent box which is insensitive to the bit-rate, modulation-format, power, and wavelengths of the signal(s) passing through it, and most importantly, provides gain for all the WDM channels simultaneously. Since all the channels remain in optical form during amplification, optically amplified WDM systems are potentially cheaper and more reliable than electrically regenerated systems.

Gain Peaking in EDFA Cascades

The EDFA is an almost ideal optical amplifier for WDM systems except for one major flaw: the gain is not uniform with wavelength, whereas the inter-amplifier losses are nearly wavelength independent.^{124–127} For a single amplifier, as shown in Fig. 43, the gain exhibits a peak at 1530 nm and a relatively flat region near 1555 nm. Moreover, the gain shape of an EDFA is dependent on the inversion of Er^{3+} in the Erbium-doped fiber.¹²⁸ When the inversion is low, which can be achieved by operating the amplifier in deep saturation, the gain peak at 1530 nm can be suppressed and the gain flatness around 1555 nm would become quite flat.

If several channels are located on the relatively flat shoulder region of the gain spectrum, then the gain differential after a single amplifier will be within a few decibels. However, when a cascade of EDFAs is used to periodically compensate for losses, the differential in gain and resultant OSNR can become quite severe. A large differential in SNR among many channels can be deleterious for proper system performance. Figure 43 shows the gain spectrum after a single amplifier and after 13 cascaded amplifiers. The gain does not accumulate linearly from stage to stage, and the resultant wavelength-dependent gain shape dramatically changes in a cascade. Along the cascade, gain is gradually “pulled” away from the shorter wavelengths and made available at the longer wavelengths, resulting in a usable bandwidth of only several nanometers.

EDFA Gain Flattening

We have shown the bandwidth reduction due to nonuniform gain in a cascade of EDFAs. It is clear that gain flattening is an important issue in optically amplified networks. Several methods have been reported for equalizing nonuniform EDFA gain. These methods include

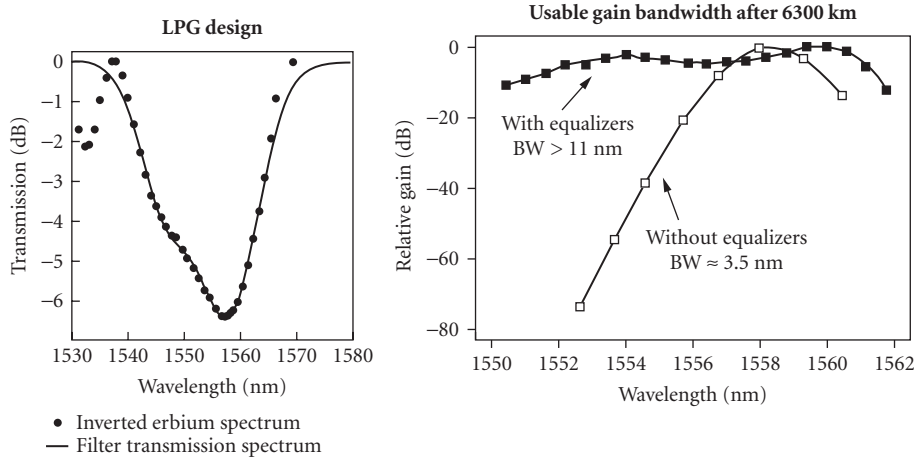


FIGURE 44 LPG design and gain equalization results.¹²⁹

1. Long period grating filters: A long period grating (LPG) with an index-varying period of approximately $100\ \mu\text{m}$ provides coupling between the core modes and the cladding modes, creating a wavelength-dependent loss to equalize the EDFA gain shape,^{129–131} as shown in Fig. 44.
2. Mach-Zehnder filters: The wavelength dependent transmission characteristics of cascaded Mach-Zehnder filters can be tailored to compensate for the gain nonuniformity of EDFAs.¹³²
3. Special designed EDFAs: A new coaxial dual-core gain-flattened EDF refractive index profile (RIP) is demonstrated recently, which is based on resonant coupling analogous to that in an asymmetric directional coupler. It has median gains more than 28 dB and gain excursion within $\pm 2\ \text{dB}$ across the C-band.¹³³

Fast Power Transients

The lifetime of a stimulated erbium ion is generally approximately 10 ms, which seems to be long enough to be transparent to signals modulated by data at the rates of several gigabits per second or higher. However, the EDFAs could be critically affected by the adding or dropping of WDM channels, network reconfiguration, or link failures, as illustrated in Fig. 45. To achieve optimal channel SNRs, the EDFAs are typically operated in the gain-saturation regime where all channels must share

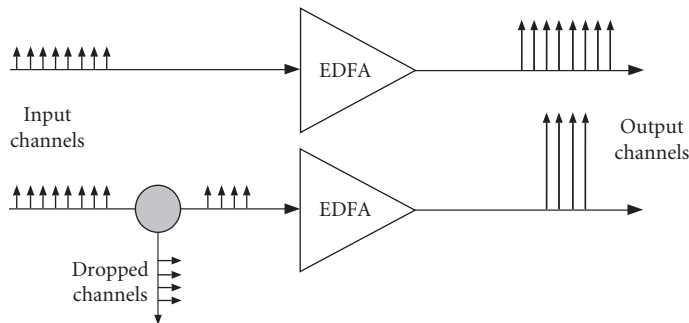


FIGURE 45 EDFA gain transients.

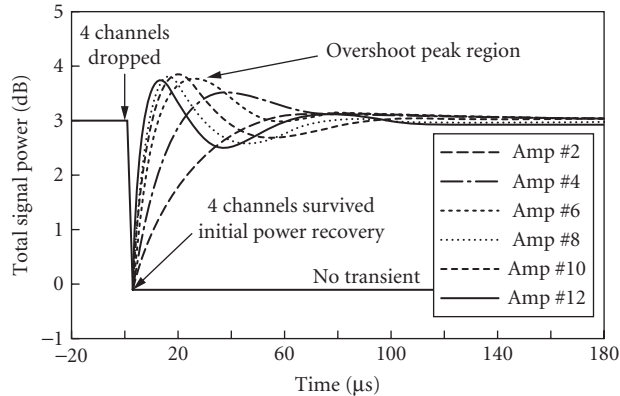


FIGURE 46 Fast power transients in EDFA cascades.¹³⁶

the available gain.^{134,135} Therefore, when channels are added or dropped, the power of the remaining channels will increase resulting transient effects.

The transients can be very fast in EDFA cascades.¹³⁶ As shown in Fig. 46, with an increase in the number of cascaded EDFAs, the transients can occur in approximately 2 μ s. These fast power transients in chain-amplifier systems should be controlled dynamically, and the response time required scales as the size of the network. For large-scale networks, response times shorter than 100 ns may be necessary.

From a system point of view, fiber nonlinearity may become a problem when too much channel power exists, and a small SNR at the receiver may arise when too little power remains.¹³⁷ The corresponding fiber transmission penalty of the surviving channel is shown in Fig. 47 in terms of the Q factor, for varying numbers of cascaded EDFAs. When 15 channels are dropped or added, the penalties are quite severe. Note that this degradation increases with the number of channels N simply because of enhanced SPM due to a large power excursion as a result of dropping $N - 1$ channels.

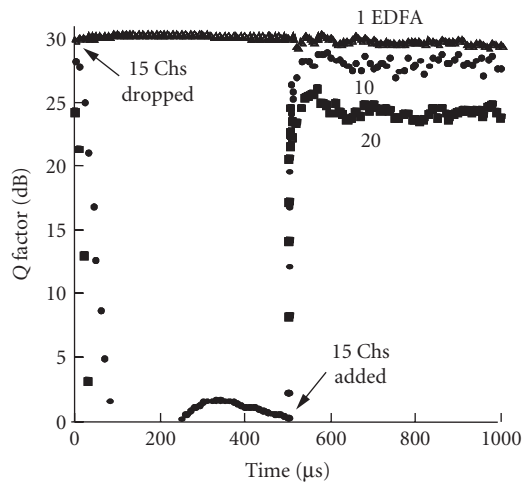


FIGURE 47 Q factor versus time for adding and dropping 15 channels of a 16-channel system at a bit rate of 10 Gb/s.¹³⁷

In order to maintain the quality of service, the surviving channels must be protected when channel add or drop or network reconfiguration occurs. The techniques include (1) optical attenuation, by adjusting optical attenuators between the gain stages in the amplifier to control the amplifier gain,¹³⁸ (2) pump power control, by adjusting the drive current of the pump lasers to control the amplifier gain,¹³⁹ (3) link control, using a power-variable control channel propagating with the signal channels to balance the amplifier gain,¹⁴⁰ and (4) EDFA gain clamping, by an automatic optical feedback control scheme to achieve all-optical gain clamping.¹⁴¹

Static Gain Dynamic and Channel Power Equalization

We just discussed EDFA gain flattening, which is a passive channel power equalization scheme effective only for a static link. However, in the nonstatic optical networks, the power in each channel suffers from dynamic network changes, including wavelength drift of components, changes in span loss, and channel add or drop. As an example, Fig. 48 shows how the gain shape of a cascaded EDFA chain varies significantly with link loss changes due to environmental problems. This is because the EDFA gain spectra are dependent on the saturation level of the amplifiers. The results in Fig. 48 are for a cascade of 10 gain-flattened EDFAs, each with 20-dB gain, saturated by 16 input channels with -18 dBm per channel.

System performance can be degraded due to unequalized WDM channel power. These degrading effects include SNR differential (reduced system dynamic range), widely varying channel crosstalk, non-linear effects, and low signal power at the receiver. Therefore, channel power needs to be equalized dynamically in WDM networks to ensure stable system performance. To obtain feedback for control purposes, a channel power monitoring scheme is very important. A simple way to accomplish this is to demultiplex all the channels and detect the power in each channel using different photodetectors or detector arrays. To avoid the high cost of many discrete components in WDM systems with large numbers of channels, other monitoring techniques that take advantage of wavelength-to-time mapping have also been proposed including the use of concatenated FBGs or swept acousto-optic tunable filters.

Various techniques have been proposed for dynamic channel power equalization, including parallel loss elements,¹⁴² individual bulk devices (e.g., AOTFs),¹⁴³ serial filters,¹⁴⁴ micro-optomechanics (MEMS),¹⁴⁵ and integrated devices.^{146,147} As an example, Fig. 49 shows the parallel loss element scheme, where the channels are demultiplexed and attenuated by separate loss elements. An additional advantage of this scheme is that ASE noise is reduced by the WDM multiplexer and demultiplexer. Possible candidates for the loss elements in this scheme include optomechanical attenuators, acousto-optic modulators, and FBGs.

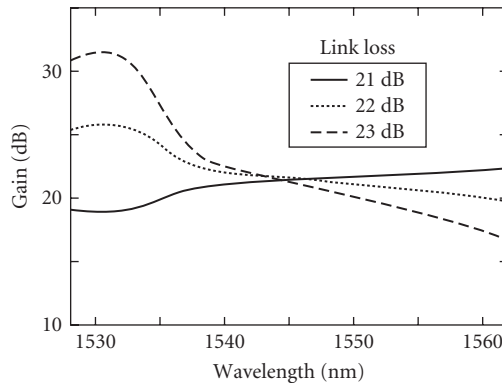


FIGURE 48 Gain spectra variation due to link loss changes for a cascade of 10 EDFAs.

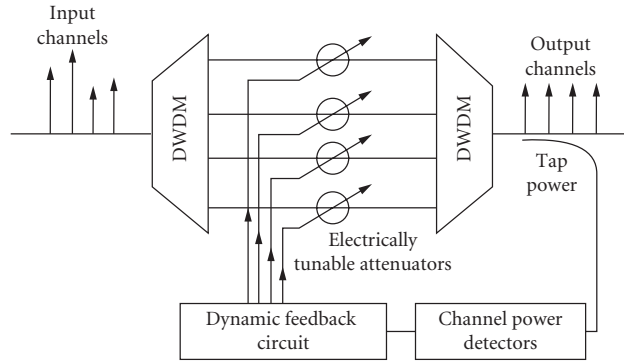


FIGURE 49 Parallel loss element scheme for dynamic channel power equalization.

Raman Amplifier

Raman amplifier is another important type of optical amplifier for WDM systems. The fundamental principles are based on Raman scattering as follows. The pump light photon is absorbed and sets the fiber molecules into mechanical vibrations. A photon is again radiated at the Stokes frequency, but since mechanical vibrations are not uniform in a fiber, the Stokes frequency is not a set number. Furthermore, the pump and signal may co- or counter-propagate in the fiber. It is worth to mention that practical, efficient, and high-power pump sources have diminished the disadvantage of the relatively poor efficiency of the Raman process over the last few years. Interest in Raman amplification has steadily increased.^{148,149}

The most important feature of Raman amplifiers is their capability to provide gain at any signal wavelength, as opposed to EDFAs based on the doped ions in the fibers. The position of the gain bandwidth within the wavelength domain can be adjusted simply by tuning the pump wavelength. Thus, Raman amplification potentially can be achieved in every region of the transmission window of the optical transmission fiber. It only depends on the availability of powerful pump sources at the required wavelengths. Figure 50 illustrates the Raman gain coefficient in a few different fibers.

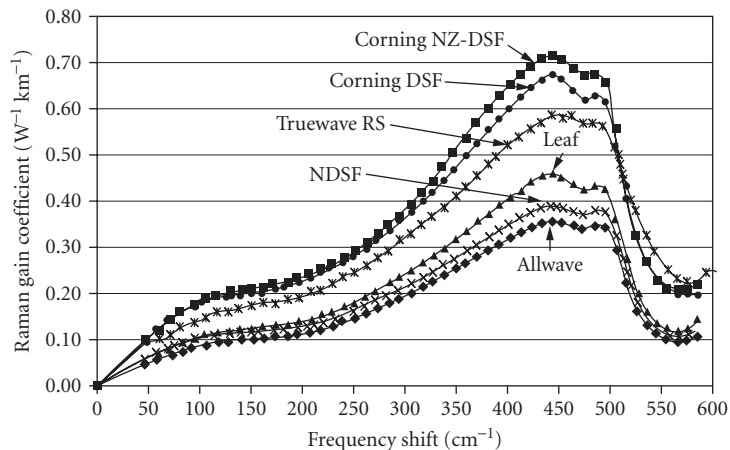


FIGURE 50 Raman gain spectra for different commercial fibers (gain peak is shifted 13 THz from the pump wavelength toward longer wavelength).¹⁴⁸

The disadvantage of Raman amplification is the need for high pump powers to provide a reasonable gain. However, the Raman effect can be used for signal amplification in transmission windows that cannot be covered properly by EDFAs. The upgrade of already existing systems by opening another transmission window where Raman amplification is applied could be an attractive application. Another application of the Raman effect is given with hybrid EDFA/Raman amplifiers characterized by a flat gain over especially large bandwidths. Repeaters can be built that compensate for the nonflatness of the EDFA gain with a more flexible Raman gain. Multiwavelength pumping could be used to shape the Raman gain such that it equalizes for the EDFA gain shaping.

Figure 51 shows a typical Raman amplifier that is backward pumped and the gain is distributed over the long transmission fibers.^{148,149} The spectral flexibility of Raman amplification allows the gain spectrum to be shaped by combining multiple pump wavelengths to make a polychromatic pump spectrum. There have been many studies searching for optimization approaches that give the flattest gain with the fewest number of pumps. Using this broadband pumping approach, amplifiers with gain bandwidths greater than 100 nm have been demonstrated.¹⁵⁰ When designing such broadband Raman amplifiers, one must consider the strong Raman interaction between the pumps. The short wavelength pumps amplify the longer wavelengths, and so more power is typically needed at the

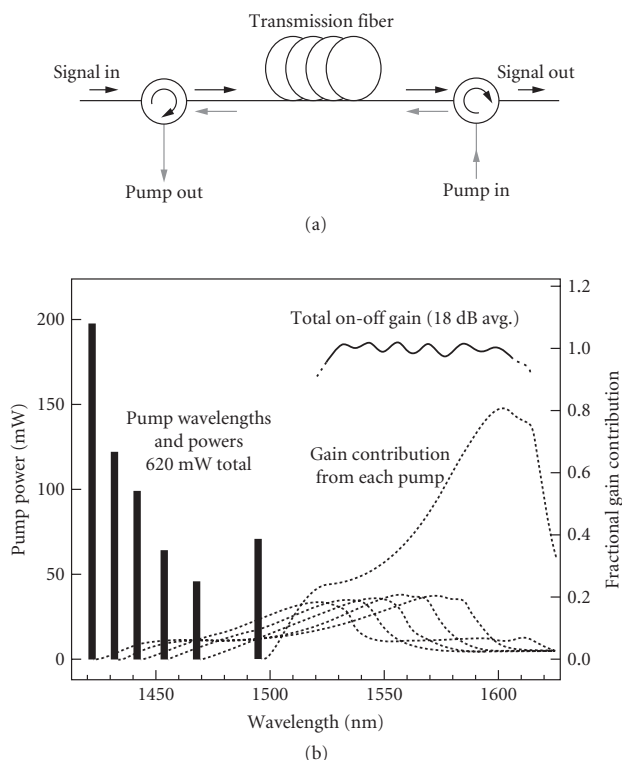


FIGURE 51 (a) Basic setup of backward pumped Raman amplifiers. (b) A numerical example of broadband Raman gain obtained using a broadband spectrum to pump a NZDSF. Bars show the counter-pump wavelengths and its power. Solid line shows the total small-signal on-off gain. Dashed lines show the fractional gain contribution from each pump wavelength.¹⁴⁹

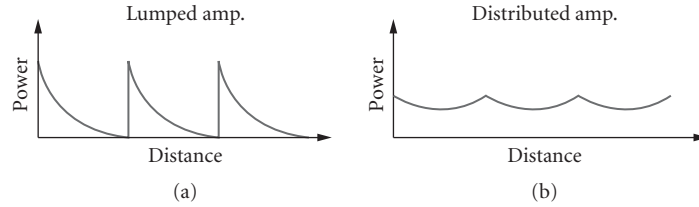


FIGURE 52 Power evolution along distance with (a) lump amplification (EDFA) only and (b) distributed amplification.

shortest wavelengths (see Fig. 51b). This interaction between the pumps also affects the noise properties of broadband amplifiers.

Another advantage of Raman amplifier is the feature of distributed amplification, since the transmission fiber itself can be used as a gain medium. As shown in Fig. 52a, in the conventional EDFA repeater systems, the signal monotonically attenuates in the fiber span, which is amplified at a point of the EDFA (lumped amplifier) location to recover the original level before entering the next fiber span. On the other hand, Raman amplifiers are mostly used in a distributed configuration, as shown in Fig. 52b. The transmission impairments are caused mostly by signal quality degradation due to optical nonlinearity in the transmission fiber and ASE noise entailed by optical amplifiers. In the presence of distributed Raman amplifier, the magnitude of the signal level excursion is smaller than the case with EDFA only, which can reduce both nonlinearity and degradation of OSNR due to ASE noise.^{151,152} A transmission of 6.4 Tb/s (160×42.7 Gb/s) over 3200 km of fiber has been demonstrated in a distributed Raman amplified system.¹⁵³

21.6 SUMMARY

In this chapter, we have covered many different aspects of high-speed WDM fiber-optic communication networks. We have endeavored to treat the most important topics—those that will likely impact these networks for years to come. The enormous growth of these systems is due to the revolutionary introduction of the EDFA. With the increasing knowledge, more development, higher data rates, and increasing channel count, WDM network limitations are being continually redefined. Network reconfigurability can offer great benefits for future WDM networks. However, a number of new degrading effects must be solved before reconfigurable networks become a reality. Yet the push for more bandwidth in WDM systems continues due to the enormous inherent potential of the optical fiber.

21.7 ACKNOWLEDGMENTS

We would like to extend our gratitude to Phillip Regan, Jing Yang, and Leroy Chee for their generous help to this chapter.

21.8 REFERENCES

1. F. P. Kapron, "Fiber-Optic System Tradeoffs," *IEEE Spectrum Magazine* **22**:68–75, 1985.
2. J. MacMillan, "Advanced Fiber Optics," *U.S. News and World Report*, p. 58, May 1994.
3. A. H. Gnauck, G. Charlet, P. Tran, P. J. Winzer, C. R. Doerr, J. C. Centanni, E. C. Burrows, T. Kawanishi, T. Sakamoto, and K. Higuma, "25.6-Tb/s WDM Transmission of Polarization-Multiplexed RZ-DQPSK Signals," *IEEE Journal of Lightwave Technology* **26**:79–84, 2008.

4. J. Seoane, A. T. Clausen, L. K. Oxenlowe, M. Galili, T. Tokle, and P. Jeppesen, "Enabling Technologies for OTDM Networks at 160 Gbit/s and beyond," pp. 22–28, Paper MG1, Orlando, Fla., October 2005.
5. M. Nakazawa, T. Yamamoto, and K. R. Tamura, "1.28 Tbit/s-70 km OTDM Transmission Using Third- and Fourth-Order Simultaneous Dispersion Compensation with a Phase Modulator," *IEEE Electronics Letters* **36**:2027–2029, 2000.
6. P. J. Winzer and R. -J. Essiambre, "Advanced Modulation Formats for High-Capacity Optical Transport Networks," *IEEE Journal of Lightwave Technology* **24**:4711–4728, 2006.
7. G. Vareille, F. Pitel, and J. F. Marcero, "3 Tbit/s (300×11.6 Gbit/s) Transmission over 7380 km Using C+L Band with 25 GHz Spacing and NRZ Format," *Optical Fiber Communication Conference and Exhibit*, Paper PD22, Anaheim, Calif, March 2001.
8. J. -X. Cai, M. Nissov, C. R. Davidson, Y. Cai, A. N. Pilipetskii, H. Li, M. A. Mills, et al., "Transmission of Thirty-Eight 40 Gb/s Channels (> 1.5 Tb/s) over Transoceanic Distance," *Conference on Optical Fiber Communication (OFC) '02*, Paper PD FC-4, Anaheim, Calif., March 2002.
9. Y. Frignac, G. Charlet, W. Idler, R. Dischler, P. Tran, S. Lanne, S. Borne, et. al., "Transmission of 256 Wavelength-Division and Polarization-Division-Multiplexed Channels at 42.7 Gb/s (10.2 Tbit/s capacity) over 3×100 km TeraLight™ Fiber," *Conference on Optical Fiber Communication (OFC) '02*, Paper: PD FC-5, Anaheim, Calif, March 2002.
10. J. Berthold, A. A. M. Saleh, L. Blair, and J. M. Simmons, "Optical Networking: Past, Present, and Future," *IEEE Journal of Lightwave Technology* **26**:1104–1118, 2008.
11. L. G. Kazovsky, W. -T. Shaw, D. Gutierrez, N. Cheng, and S. -W. Wong, "Next-Generation Optical Access Networks," *IEEE Journal of Lightwave Technology* **25**:3428–3442, 2007.
12. P. Kaiser and D. B. Keck, "Fiber Types and Their Status," *Optical Fiber Telecommunications II*, Chap. 2, p. 40, S. E. Miller and I. P. Kaminow, eds., Academic Press, New York, 1988.
13. C. A. Brackett, "Dense Wavelength Division Multiplexing: Principles and Applications," *IEEE Journal on Selected Areas in Communications* **8**(6):948–964, 1990.
14. I. P. Kaminow, "FSK with Direct Detection in Optical Multiple-Access FDM Networks," *IEEE Journal on Selected Areas in Communications* **8**:1005–1014, 1990.
15. P. E. Green, Jr., *Fiber Optic Networks*, Prentice Hall, Englewood Cliffs, N.J., 1993.
16. N. K. Cheung, K. Nosu, and G. Winzer, Special Issue on Wavelength Division Multiplexing, *IEEE Journal on Selected Areas in Communications* **8**, 1990.
17. A. E. Willner, I. P. Kaminow, M. Kuznetsov, J. Stone, and L. W. Stulz, "1.2 Gb/s Closely-Spaced FDMA-FSK Direct-Detection Star Network," *IEEE Photonics Technology Letters* **2**:223–226, 1990.
18. N. R. Dono, P. E. Green, K. Liu, R. Ramaswami, and F. F. Tong, "A Wavelength Division Multiple Access Network for Computer Communication," *IEEE Journal on Selected Areas in Communications* **8**:983–994, 1990.
19. W. I. Way, D. A. Smith, J. J. Johnson, and H. Izadpanah, "A Self-Routing WDM High-Capacity SONET Ring Network," *IEEE Photonics Technology Letters* **4**:402–405, 1992.
20. T. -H. Wu, *Fiber Network Service Survivability*, Artech House, Boston, Mass., 1992.
21. A. S. Acampora, M. J. Karol, and M. G. Hluchyj, "Terabit Lightwave Networks: The Multihop Approach," *AT&T Technical Journal* **66**:21–34, November/December 1987.
22. M. Schwartz, *Telecommunication Networks, Protocols, Modeling, and Analysis*, Addison Wesley, New York, 1987.
23. M. Jeong, H. C. Cankaya, and C. Qiao, "On a New Multicasting Approach in Optical Burst Switched Networks," *IEEE Communications Magazine* **40**:96–103, 2002.
24. I. Baldine, H. G. Perros, G. N. Rouskas, and D. Stevenson, "JumpStart: A Just-in-Time Signaling Architecture for WDM Burst-Switched Networks," *IEEE Communications Magazine* **40**:82–89, 2002.
25. J. E. Berthold, "Networking Fundamentals," *Conference on Optical Fiber Communications (OFC) '94*, Tutorial TuK, San Jose, Calif., February 1994.
26. J. Y. Hui, *Switching and Traffic Theory for Integrated Broadband Networks*, Kluwer Academic Publishers, Boston, 1990.
27. J. B. Yoo and G. K. Chang, "High-Throughput, Low-Latency Next Generation Internet Using Optical-Tag Switching," *U.S. Patent 6,111,673*, 1997.
28. M. W. Maeda, A. E. Willner, J. R. Wullert II, J. Patel, and M. Allersma, "Wavelength-Division Multiple-Access Network Based on Centralized Common-Wavelength Control," *IEEE Photonics Technology Letters* **5**:83–86, 1993.

29. K. K. Goel, "Nonrecirculating and Recirculating Delay Line Loop Topologies of Fiber-Optic Delay Line Filters," *IEEE Photonics Technology Letters* **5**:1086–1088, 1993.
30. I. Chlamtac, A. Fumagalli, and S. Chang-Jin, "Multibuffer Delay Line Architectures for Efficient Contention Resolution in Optical Switching Nodes," *IEEE Transactions on Communications* **48**:2089–2098, 2000.
31. A. Agrawal, L. Wang, Y. Su, and P. Kumar, "All-Optical Erasable Storage Buffer Based on Parametric Nonlinearity in Fiber," *Conference on Optical Fiber Communication (OFC) '01*, Paper ThH5, Anaheim, Calif., March 2001.
32. A. Rader and B. L. Anderson, "Demonstration of a Linear Optical True-Time Delay Device by Use of a Microelectromechanical Mirror Array," *Applied Optics* **42**:1409–1416, 2003.
33. D. R. Pape and A. P. Goutzoulis, "New Wavelength Division Multiplexing True-Time-Delay Network for Wideband Phased Array Antennas," *Journal of Optics A: Pure Applied Optics* **1**:320–323, 1999.
34. C. J. Chang-Hasnain, P. Ku, and J. Kim, S. Chuang, "Variable Optical Buffer Using Slow Light in Semiconductor Nanostructures," *Proceedings of the IEEE*, **91**:1884–1897, 2003.
35. S. Rangarajan, H. Zhaoyang, L. Rau, and D. J. Blumenthal, "All-Optical Contention Resolution with Wavelength Conversion for Asynchronous Variable-Length 40 Gb/s Optical Packets," *IEEE Photonics Technology Letters* **16**:689–691, 2004.
36. J. Elmırghani and H. Mouftah, "All-Optical Wavelength Conversion: Techniques and Applications in DWDM Networks," *IEEE Communications Magazine* **38**:86–92, 2000.
37. D. Nasset, T. Kelly, and D. Marcenac, "All-Optical Wavelength Conversion Using SOA Nonlinearities," *IEEE Communications Magazine* **36**:56–61, 1998.
38. I. Brener, M. H. Chou, and M. M. Fejer, "Efficient Wideband Wavelength Conversion Using Cascaded Second-Order Nonlinearities in LiNbO_3 Waveguides," *Conference on Optical Fiber Communication (OFC) '99*, Paper FB6, San Diego, Calif., February 1999.
39. A. Hsu and S. L. Chuang, "Wavelength Conversion by Cross-Absorption Modulation Using an Integrated Electroabsorption Modulator/Laser," *Summaries of Papers Presented at the Conference on Lasers and Electro-Optics (CLEO) '99*, Paper CThV3, Baltimore, Md., May 1999.
40. M. Baresi, S. Bregni, A. Pattavina, and G. Vegetti, "Deflection Routing Effectiveness in Full-Optical IP Packet Switching Networks," *IEEE International Conference on Communications* **2**:1360–1364, Anchorage, Alaska, May 2003.
41. F. -S. Choa, X. Zhao, Y. Xiuqin, J. Lin, J. P. Zhang, Y. Gu, G. Ru, et al., "An Optical Packet Switch Based on WDM Technologies," *IEEE Journal of Lightwave Technology* **23**:994–1014, 2005.
42. A. E. Willner, D. Gurkan, A. B. Sahin, J. E. McGeehan, and M. C. Hauer, "All-Optical Address Recognition for Optically-Assisted Routing in Next-Generation Optical Networks," *IEEE Communications Magazine* **41**: S38–S44, May 2003.
43. N. Calabretta, H. de Waardt, G. D. Khoe, and H. J. S. Dorren, "Ultrafast Asynchronous Multioutput All-Optical Header Processor," *IEEE Photonics Technology Letters* **16**:1182–1184, 2004.
44. D. Gurkan, M. C. Hauer, A. B. Sahin, Z. Pan, S. Lee, A. E. Willner, K. R. Parameswaran, and M. M. Fejer, "Demonstration of Multi-Wavelength All-Optical Header Recognition Using a PPLN and Optical Correlates," *European Conference on Optical Communication (ECOC)'01*, Paper We.B.2.5, Amsterdam, NL, September/October 2001.
45. J. Bannister, J. Touch, P. Kamath, and A. Patel, "An Optical Booster for Internet Routers," *8th International Conference. High Performance Computing*, pp. 399–413, Hyderabad, India, December 2001.
46. D. J. Blumenthal, J. E. Bowers, L. Rau, Hsu-Feng Chou, S. Rangarajan, Wei Wang, and K. N. Poulsen, "Optical Signal Processing for Optical Packet Switching Networks," *IEEE Communications Magazine* **41**:S23–S29, 2003.
47. L. Rau, S. Rangarajan, D. J. Blumenthal, H. -F. Chou, Y. -J. Chiu, and J. E. Bowers, "Two-Hop All-Optical Label Swapping with Variable Length 80 Gb/s Packets and 10 Gb/s Labels Using Nonlinear Fiber Wavelength Converters, Unicast/Multicast Output and a Single EAM for 80- to 10 Gb/s Packet Demultiplexing," *Optical Fiber Communication Conference (OFC)'02*, Pages FD2-1, Anaheim, Calif., March 2002.
48. S. Yao, B. Mukherjee, and S. Dixit, "Advances in Photonic Packet Switching: An Overview," *IEEE Communications Magazine* **38**:84–94, 2000.
49. V. W. S. Chan, K. L. Hall, E. Modiano, and K. A. Rauschenbach, "Architectures and Technologies for High-Speed Optical Data Networks," *IEEE Journal of Lightwave Technology* **16**:2146–2168, 1998.
50. Winston I. Way, "Rules of the ROADM, A Deployment Guide," *Electronic Engineering Times*, pp. 60–64, November 2005. <http://www.eetimes.com/news/latest/showArticle.jhtml?articleID=173601985>.

51. R. S. Bernhey and M. Kanaan, "ROADM Deployment, Challenges, and Applications," *Optical Fiber Communication and the National Fiber Optic Engineers Conference (OFC/NFOEC)'07*, Paper NWD1, Anaheim, CA, USA, March 2007.
52. B. P. Keyworth, "ROADM Subsystems and Technologies," *Optical Fiber Communication and the National Fiber Optic Engineers Conference (OFC/NFOEC)'05*, Paper OWB5, Anaheim, CA, USA, March 2005.
53. K. Grobe, "Applications of ROADMS and Metro Planes in Control and Regional Networks," *Optical Fiber Communication and the National Fiber Optic Engineers Conference (OFC/NFOEC)'07*, Paper NTuC1, Anaheim, CA, USA, March 2007.
54. P. Roorda and B. Collings, "Evolution to Colorless and Directionless ROADM Architectures," *Optical Fiber Communication and the National Fiber Optic Engineers Conference (OFC/NFOEC)'08*, Paper NWE2, San Diego, CA, USA, February 2008.
55. <http://www.nortel.com/solutions/optical/collateral/nn115940.pdf>, April 2006.
56. K. Guild, "Impairment-Aware Routing for OBS and OPS Networks," *International Conference on Transparent Optical Networks, 2006* 3:61, Nottingham, UK, June 2006.
57. Y. Huang, J. P. Heritage, and B. Mukherjee, "Connection Provisioning with Transmission Impairment Consideration in Optical WDM Networks with High-Speed Channels," *IEEE Journal of Lightwave Technology* 23:982–993, 2005.
58. T. Carpenter, D. Shallcross, J. Gannett, J. Jackel, and A. Von Lehmen, "Maximizing the Transparency Advantage in Optical Networks," *Optical Fiber Communications Conference (OFC)'03* 2:616–617, Atlanta, Ga., March 2003.
59. N. Andrioli, P. Castoldi, J. Cornelias, F. Cugini, G. Junyent, R. Martinez, C. Pinart, L. Vakarenghi, and L. Wosinska, "Challenges and Requirements for Introducing Impairment-Awareness into the Management and Control Planes of ASON/GMPLS WDM Networks," *IEEE Communications Magazine* 44:76–85, 2006.
60. G. P. Agrawal, *Fiber-Optics Communication Systems*, 2nd ed., Wiley, New York, 2002.
61. A. E. Willner and B. Hoanca, "Fixed and Tunable Management of Fiber Chromatic Dispersion," *Optical Fiber Telecommunications IV*, Ivan P. Kaminow and Tingye Li, eds., New York: Academic Press, 2002.
62. L. D. Garrett, "All about Chromatic Dispersion in Dense WDM Optical Fiber Transmission," Invited Short Course, *Optical Fiber Communication Conference (OFC)'01*, Anaheim, Calif., March 2001.
63. C. D. Poole, "Statistical Treatment of Polarization Dispersion in Single-Mode Fiber," *Optics Letters* 13: 687–689, 1988.
64. C. D. Poole, "Measurement of Polarization-Mode Dispersion in Single-Mode Fibers with Random Mode Coupling," *Optics Letters* 14:523–525, 1989.
65. Y. NamiHIRA and H. Wakabayashi, "Fiber Length Dependence of Polarization Mode Dispersion Measurement in Long-Length Optical Fibers and Installed Optical Submarine Cables," *Journal of Optical Communications* 12:2–9, 1991.
66. C. D. Poole, R. W. Tkach, A. R. Chraplyvy, and D. A. Fishman, "Fading in Lightwave Systems due to Polarization-Mode Dispersion," *IEEE Photonics Technology Letters* 3:68–70, 1991.
67. A. E. Willner, "Polarization Mode Dispersion: Playing Russian Roulette with Your Network," *Lightwave Magazine* 19:79–82, 2002.
68. V. J. Mazurczyk and J. L. Zyskind, "Polarization Hole Burning in Erbium Doped Fiber Amplifiers," *Lasers and Electro-Optics/Quantum Electronics and Laser Science (CLEO/QELS)'93*, Paper CPD26, Baltimore, Md., May 1993.
69. H. F. Haunstein and H. M. Kallert, "Influence of PMD on the Performance of Optical Transmission Systems in the Presence of PDL," *Conference on Optical Fiber Communication (OFC) '01*, Paper WT4, Anaheim, Calif., March 2001.
70. N. Gisin and B. Huttner, "Combined Effects of Polarization Mode Dispersion and Polarization Dependent Loss," *Optics Communications* 142:119–125, 1997.
71. B. Huttner, C. Geiser, and N. Gisin, "Polarization-Induced Distortions in Optical Fiber Networks with Polarization-Mode Dispersion and Polarization-Dependent Losses," *IEEE Journal of Selected Topics in Quantum Electronics* 6:317–329, 2000.
72. L. S. Yan, Q. Yu, Y. Xie, and A. E. Willner, "Experimental Demonstration of the System Performance Degradation due to the Combined Effect of Polarization Dependent Loss with Polarization Mode Dispersion," *IEEE Photonics Technology Letters* 14:224–226, 2002.

73. L. -S. Yan, Q. Yu, T. Luo, J. E. McGeehan, and A. E. Willner, "Deleterious System Effects due to Low-Frequency Polarization Scrambling in the Presence of Nonnegligible Polarization-Dependent Loss," *IEEE Photonic Technology Letters* **15**:464–466, 2003.
74. G. P. Agrawal, *Nonlinear Fiber Optics*, Academic Press, New York, 1990.
75. R. W. Tkach, A. R. Chraplyvy, F. Forghieri, A. H. Gnauck, and R. M. Derosier, "Four-Photo Mixing and High-Speed WDM System," *IEEE Journal of Lightwave Technology* **13**:841–849, 1995.
76. P. P. Mitra and K. B. Stark, "Nonlinear Limits to the Information Capacity of Optical Fiber Communications," *Nature* **411**:1027–1030, 2001.
77. A. J. Antos and D. K. Smith, "Design and Characterization of Dispersion Compensating Fiber Based on the LP_{01} Mode," *IEEE Journal of Lightwave Technology* **12**:1739–1745, 1994.
78. A. H. Gnauck, L. D. Garrett, F. Forghieri, V. Gusmeroli, and D. Scarano, " 8×20 Gb/s 315-km, 8×10 Gb/s 480-km WDM Transmission Over Conventional Fiber Using Multiple Broad-Band Fiber Gratings," *IEEE Photonics Technology Letters* **10**:1495–1497, 1998.
79. M. Ibsen, M. K. Durkin, M. J. Cole, and R. I. Laming, "Sinc-Sampled Fiber Bragg Gratings for Identical Multiple Wavelength Operation," *IEEE Photonics Technology Letters* **10**:842–845, 1998.
80. K. -M. Feng, J. -X. Cai, V. Grubsky, D. S. Starodubov, M. I. Hayee, S. Lee, X. Jiang, A. E. Willner, and J. Feinberg, "Dynamic Dispersion Compensation in a 10-Gb/s Optical System Using a Novel Voltage Tuned Nonlinearly Chirped Fiber Bragg Grating," *IEEE Photonics Technology Letters* **11**:373–375, 1999.
81. Benjamin J. Eggleton, John A. Rogers, Paul S. Westbrook, and Thomas A. Strasser, "Electrically Tunable Power Efficient Dispersion Compensating Fiber Bragg Grating," *IEEE Photonics Technology Letters* **11**:854–856, 1999.
82. Y. Painchaud, "Dispersion Compensation Module," <http://www.teraxion.com>, 2008.
83. J. H. Winters and R. D. Gitlin, "Electrical Signal Processing Techniques in Long-Haul Fiber-Optic Systems," *IEEE Transactions on Communications* **38**:1439–1453, 1990.
84. H. Bülow, "Electronic Equalization of Transmission Impairments," *Conference on Optical Fiber Communication (OFC) '02*, Invited Paper, TuE4, Anaheim, Calif, March 2002.
85. B. Franz, D. Rosener, F. Buchali, H. Bulow, and G. Veith, "Adaptive Electronic Feed-Forward Equaliser and Decision Feedback Equaliser for the Mitigation of Chromatic Dispersion and PMD in 43 Gbit/s Optical Transmission Systems," *European Conference on Optical Communication (ECOC) '06*, Paper We1.5.1, Cannes, France, September 2006.
86. N. Alic, G. C. Papan, R. E. Saperstein, R. Jiang, C. Marki, Y. Fainman, and S. Radic, "Experimental Demonstration of 10 Gb/s NRZ Extended Dispersion-Limited Reach Over 600 km-SMF Link Without Optical Dispersion Compensation," *Conference on Optical Fiber Communication (OFC) '06*, Paper OWB7, Anaheim, Calif. March 2006.
87. P. Poggiolini, G. Bosco, S. Savory, Y. Benlachtar, R. I. Killey, and J. Prat, "1040 km Uncompensated IMDD Transmission Over G.652 Fiber at 10 Gbit/s Using a Reduced-State SQRT-Metric MLSE Receiver," *European Conference on Optical Communication (ECOC) '06*, Postdeadline Paper Th4.4.6., Cannes, France, September 2006.
88. D. McGhan, C. Laperle, A. Savchenko, C. Li, G. Mak, and M. O'Sullivan, "5120-km RZ-DPSK Transmission Over G.652 Fiber at 10 Gb/s without Optical Dispersion Compensation," *IEEE Photonics Technology Letters* **18**:400–402, 2006.
89. A. H. Gnauck, R. W. Tkach, A. R. Chraplyvy, and T. Li, "High-Capacity Optical Transmission Systems," *IEEE Journal of Lightwave Technology* **26**:1032–1045, 2008.
90. C. Laperle, B. Villeneuve, Z. Zhang, D. McGhan, H. Sun, and M. O'Sullivan, "Wavelength Division Multiplexed (WDM) and Polarization Mode Dispersion (PMD) Performance of a Coherent 40 Gbit/s Dual Polarization Quadrature Phase Shift Keying (DP-QPSK) Transceiver," *Conference on Optical Fiber Communication (OFC) '07*, Postdeadline Paper PDP16, Anaheim, Calif., March 2007.
91. G. Charlet, J. Renaudier, H. Mardoyan, O. B. Pardo, F. Cerou, P. Tran, and S. Bigo, "12.8 Tbit/s Transmission of 160 PDM-QPSK ($160 \times 2 \times 40$ Gbit/s) Channels with Coherent Detection over 2550 km," *European Conference on Optical Communication (ECOC) '07*, Postdeadline Paper PD1.6., Berlin, Germany, 2007.
92. C. R. S. Fludger, T. Duthel, D. v. d. Borne, C. Schülten, E.-D. Schmidt, T. Wuth, E. D. Man, G. D. Khoe, and H. D. Waardt, " 10×111 Gbit/s, 50 GHz Spaced, POLMUX-RZ-DQPSK Transmission over 2375 km Employing Coherent Equalisation," *Conference on Optical Fiber Communication (OFC) '07*, Postdeadline Paper PDP22, Anaheim, Calif., USA, March 2007.

93. X. Liu, S. Chandrasekhar, A. H. Gnauck, C. R. Doerr, I. Kang, D. Kilper, L. L. Buhl, and J. Centanni, "DSP-Enabled Compensation of Demodulator Phase Error and Sensitivity Improvement in Direct-Detection 40-Gb/s DQPSK," *European Conference on Optical Communication (ECOC) '06*, Postdeadline Paper Th4.4.5, Cannes, France, 2006.
94. H. B. Killen, *Fiber Optic Communications*, Prentice Hall, Englewood Cliffs, N.J., 1991.
95. A. H. Gnauck and P. J. Winzer, "Optical Phase-Shift-Keyed Transmission," *IEEE Journal of Lightwave Technology* **23**:115–130, 2005.
96. S. R. Chinn, D. M. Boroson, and J. C. Livas, "Sensitivity of Optically Preamplified DPSK Receivers with Fabry-Perot Filters," *IEEE Journal of Lightwave Technology* **14**:370–376, 1996.
97. C. Xu, X. Liu, and L. Mollenauer, "Comparison of Return-to-Zero Phase Shift Keying and On-Off Keying in Long Haul Dispersion Managed Transmissions," *Conference on Optical Fiber Communication (OFC) '03*, Paper ThE3, Atlanta, Ga., March 2003.
98. P. S. Henry, R. A. Linke, and A. H. Gnauck, "Introduction to Lightwave Systems," *Optical Fiber Telecommunications II*, Chap. 21, S. E. Miller and I. P. Kaminow, eds., Academic Press, New York, 1988.
99. J. M. Senior, *Optical Fiber Communications*, Prentice Hall, New York, 1985.
100. S. Betti, G. de Marchis, and E. Iannone, *Coherent Optical Communication Systems*, chaps. 5 and 6, Wiley, New York, 1995.
101. M. S. Roden, *Analog and Digital Communication Systems*, 4th ed., Prentice Hall, Upper Saddle River, N. J., 1995.
102. K. Sato, S. Kuwahara, Y. Miyamoto, K. Murata, and H. Miyazawa, "Carrier-Suppressed Return-to-Zero Pulse Generation Using Mode-Locked Lasers for 40-Gbit/s Transmission," *IEICE Transactions on Communications* **E85-B**:410–415, 2002.
103. K. Yonenaga and S. Kuwano, "Dispersion-Tolerant Optical Transmission System Using Duobinary Transmitter and Binary Receiver," *IEEE Journal of Lightwave Technology* **15**:1530–1537, 1997.
104. K. Yonenaga, M. Yoneyama, Y. Miyamoto, K. Hagimoto, and K. Noguchi, "160 Gbit/s WDM Transmission Experiment Using Four 40 Gbit/s Optical Duobinary Channels," *IEEE Electronics Letters* **34**:1506–1507, 1998.
105. T. Ono, Y. Yano, K. Fukuchi, T. Ito, H. Yamazaki, M. Yamaguchi, and K. Emura, "Characteristics of Optical Duobinary Signals in Terabit/s Capacity, High-Spectral Efficiency WDM Systems," *IEEE Journal of Lightwave Technology* **16**:788–797, 1998.
106. A. E. Willner, "Simplified Model of a FSK-to-ASK Direct-Detection System Using a Fabry-Perot Demodulator," *IEEE Photonics Technology Letters* **2**:363–366, 1990.
107. J. H. Sinsky, A. Adamiecki, A. Gnauck, C. Burrus, J. Leuthold, O. Wohlgenuth, and A. Umbach, "A 42.7-Gb/s Integrated Balanced Optical Front End with Record Sensitivity," *Conference on Optical Fiber Communication (OFC) '03*, Postdeadline Paper PD39, Atlanta, Ga., March 2003.
108. P. J. Winzer, A. H. Gnauck, G. Raybon, S. Chandrasekhar, Y. Su, and J. Leuthold, "40-Gb/s Return-to-Zero Alternate-Mark-Inversion (RZ-AMI) Transmission over 2000 km," *IEEE Photonics Technology Letters* **15**:766–768, 2003.
109. T. Hoshida, O. Vassilieva, K. Yamada, S. Choudhary, R. Pecqueur, and H. Kuwahara, "Optimal 40 Gb/s Modulation Formats for Spectrally Efficient Long-Haul DWDM Systems," *IEEE Journal of Lightwave Technology* **20**:1989–1996, 2002.
110. G. Kramer, A. Ashikhmin, A. J. van Wijngaarden, and X. Wei, "Spectral Efficiency of Coded Phase-Shift Keying for Fiber-Optic Communication," *IEEE Journal of Lightwave Technology* **21**:2438–2445, 2003.
111. H. Kim and P. J. Winzer, "Robustness to Laser Frequency Offset in Direct Detection DPSK and DQPSK Systems," *IEEE Journal of Lightwave Technology* **21**:1887–1891, 2003.
112. D. van den Borne, S. L. Jansen, E. Gottwald, P. M. Krummrich, G. D. Khoe, and H. de Waardt, "Spectral Efficiency of Coded Phase-Shift Keying for Fiber-Optic Communication," *IEEE Journal of Lightwave Technology* **25**:222–232, 2007.
113. M. Ohm and J. Speidel, "Quaternary Optical ASK-DPSK and Receivers with Direct Detection," *IEEE Photonics Technology Letters* **15**:159–161, 2003.
114. S. Hayase, N. Kikuchi, K. Sekine, and S. Sasaki, "Proposal of 8-State per Symbol (Binary ASK and QPSK) 30-Gbit/s Optical Modulation/Demodulation Scheme," *European Conference on Optical Communication (ECOC) '03*, Paper Th2.6.4, Rimini, Italy, September 2003.

115. X. Liu, X. Wei, Y. -H. Kao, J. Leuthold, C. R. Doerr, and L. F. Mollenauer, "Quaternary Differential-Phase Amplitude-Shift-Keying for DWDM Transmission," *European Conference on Optical Communication (ECOC)* '03, Paper Th2.6.5, Rimini, Italy, September 2003.
116. P. J. Winzer, and R. -J. Essiambre, "Advanced Modulation Formats for High-Capacity Optical Transport Networks," *IEEE Journal of Lightwave Technology* **24**:4711–4728, 2006.
117. J. G. Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.
118. A. H. Gnauck, P. J. Winzer, S. Chandrasekhar, and C. Dorrer, "Spectrally Efficient (0.8 b/s/Hz) 1-Tb/s (25×42.7 Gb/s) RZ-DQPSK Transmission over 28 100-km SSMF Spans with 7 Optical Add/Drops," *Conference on Optical Fiber Communication (OFC)* '04, Paper Th4.4.1, Los Angeles, Calif., February 2004.
119. P. J. Winzer, H. Kogelnik, C. H. Kim, H. Kim, R. M. Jopson, L. E. Nelson, and K. Ramanan, "Receiver Impact on First-Order PMD Outage," *IEEE Photonics Technology Letters* **15**:1482–1484, 2003.
120. E. B. Basch, R. Egorov, S. Gringeri, and S. Elby, "Architectural Tradeoffs for Reconfigurable Dense Wavelength-Division Multiplexing Systems," *IEEE Journal of Selected Topics in Quantum Electronics* **12**:615–626, 2006.
121. A. H. Gnauck, P. J. Winzer, and S. Chandrasekhar, "Hybrid 10/40-G Transmission on a 50-GHz Grid through 2800 km of SSMF and Seven Optical Add-Drops," *IEEE Photonics Technology Letters* **17**:2203–2205, 2005.
122. S. Appathurai, V. Mikhailov, R. I. Killey, and P. Bayvel, "Investigation of the Optimum Alternate-Phase RZ Modulation Format and Its Effectiveness in the Suppression of Intrachannel Nonlinear Distortion in 40-Gbit/s Transmission Over Standard Single-Mode Fiber," *IEEE Journal of Selected Topics in Quantum Electronics* **10**:239–249, 2004.
123. T. Li, "The Impact of Optical Amplifiers on Long-Distance Lightwave Telecommunications," *Proceedings of the IEEE* **81**:1568–1579, 1993.
124. E. L. Goldstein, A. F. Elrefaie, N. Jackman, and S. Zaidi, "Multiwavelength Fiber-Amplifier Cascades in Unidirectional Interoffice Ring Networks," *Conference on Optical Fiber Communication (OFC)* '93, Paper TuJ3, San Jose, Calif., February 1993.
125. J. P. Blondel, A. Pitel, and J. F. Marcereou, "Gain-Filtering Stability in Ultralong-Distance Links," *Conference on Optical Fiber Communication (OFC)* '93, paper TuI3, San Jose, Calif., February 1993.
126. H. Taga, N. Edagawa, Y. Yoshida, S. Yamamoto, and H. Wakabayashi, "IM-DD Four-Channel Transmission Experiment over 1500 km Employing 22 Cascaded Optical Amplifiers," *IEEE Electronics Letters* **29**:485–486, 1993.
127. A. E. Willner and S. -M. Hwang, "Transmission of Many WDM Channels through a Cascade of EDFA's in Long-Distance Link and Ring Networks," *IEEE Journal of Lightwave Technology* **13**:802–816, 1995.
128. I. Kaminow and T. Koch, *Optical fiber Telecommunications IIIB*, Academic Press, San Diego, 1997.
129. A. M. Vengsarkar, P. J. Lemaire, J. B. Judkins, V. Bhatia, T. Erdogan, and J. E. Sipe, "Long-Period Fiber Gratings as Band-Rejection Filters," *IEEE Journal of Lightwave Technology* **14**:58–65, 1996.
130. Y. Sun, J. B. Judkins, A. K. Srivastava, L. Garrett, J. L. Zyskind, J. W. Suloff, C. Wolf, et. al., "Transmission of 32-WDM 10-Gb/s Channels over 640 km Using Broad-Band, Gain-Flattened Erbium-Doped Silica Fiber Amplifiers," *IEEE Photonics Technology Letters* **9**:1652–1654, 1997.
131. P. F. Wysocki, J. B. Judkins, R. P. Espindola, M. Andrejco, and A. M. Vengsarkar, "Broad-Band Erbium-Doped Fiber Amplifier Flattened beyond 40 nm Using Long-Period Grating Filter," *IEEE Photonics Technology Letters* **9**:1343–1345, 1997.
132. J. Y. Pan; M. A. Ali, A.F. Elrefaie, and R. E. Wagner, "Multiwavelength Fiber-Amplifier Cascades with Equalization Employing Mach-Zehnder Optical Filter," *IEEE Photonics Technology Letters* **7**:1501–1503, 1995.
133. B. Nagaraju, M. C. Paul, M. Pal, A. Pal, R. K. Varshney, B. P. Pal, S. K. Bhadra, G. Monnom, and B. Dussardier, "Design and Realization of an Inherently Gain Flattened Erbium Doped Fiber Amplifier," *Conference on Lasers and Electro-Optics (CLEO)* '08, Paper JTuA86, San Jose, Calif., May 2008.
134. E. Desurvire, R. Giles, and J. Simpson, "Gain Saturation Effects in High-Speed, Multi Channel Erbium-Doped Fiber Amplifiers at 1.53 μm ," *IEEE Journal of Lightwave Technology* **7**:2095–2104, 1989.
135. D. C. Kilper, S. Chandrasekhar, and C. A. White, "Transient Gain Dynamics of Cascaded Erbium Doped Fiber Amplifiers with Re-Configured Channel Loading," *Conference on Optical Fiber Communication (OFC)* '06, Paper OTuK6, Anaheim, Calif., March 2006.

136. J. Zyskind, Y. Sun, A. Srivastava, J. Sulhoff, A. Lucero, C. Wolf, and R. Tkach, "Fast Power Transients in Optically Amplified Multiwavelength Optical Networks," *Conference on Optical Fiber Communication (OFC)* '96, Paper PD-31, San Jose, Calif., February 1996.
137. M. Hayee and A. Willner, "Fiber Transmission Penalties due to EDFA Power Transients Resulting from Fiber Nonlinearity and ASE Noise in Add/Drop Multiplexed WDM Networks," *IEEE Photonics Technology Letters* **11**:889–891, 1999.
138. J.-X. Cai, K.-M. Feng, and A. E. Willner, "Simultaneous Compensation of Fast Add/Drop Power-Transients and Equalization of Inter-Channel Power Differentials for Robust WDM Systems with EDFAs," *Conference on Optical Amplifiers and Their Applications*, Paper MC6, Victoria, Canada, July 1997.
139. K. Motoshima, L. Leba, D. Chen, M. Downs, T. Li, and E. Desurvire, "Dynamic Compensation of Transient Gain Saturation in Erbium-Doped Fiber Amplifiers by Pump Feedback Control," *IEEE Photonics Technology Letters* **5**:1423–1426, 1993.
140. A. Srivastava, J. Zyskind, Y. Sun, J. Ellson, G. Newsome, R. Tkach, A. Chraplyvy, J. Sulhoff, T. Strasser, C. Wolf, and J. Pedrazzani, "Fast-Link Control Protection of Surviving Channels in Multiwavelength Optical Networks," *IEEE Photonics Technology Letters* **9**:1667–1669, 1997.
141. G. Luo, J. Zyskind, Y. Sun, A. Srivastava, J. Sulhoff, and M. Ali, "Performance Degradation of All-Optical Gain-Clamped EDFA's due to Relaxation-Oscillations and Spectral-Hole Burning in Amplified WDM Networks," *IEEE Photonics Technology Letters* **9**:1346–1348, 1997.
142. J. Cai, K. Feng, X. Chen, and A. Willner, "Experimental Demonstration of Dynamic High-Speed Equalization of Three WDM Channels Using Acousto-Optic Modulators and a Wavelength Demultiplexer," *IEEE Photonics Technology Letters* **9**:678–680, 1997.
143. S. Huang, X. Zou, A. Willner, Z. Bao, and D. Smith, "Experimental Demonstration of Active Equalization and ASE Suppression of Three 2.5 Gb/s WDM-Network Channels over 2500 km Using AOTF as Transmission Filters," *IEEE Photonics Technology Letters* **9**:389–391, 1997.
144. D. Starodubov, V. Grubsky, J. Feinberg, J. Cai, K. Feng, and A. Willner, "Novel Fiber Amplitude Modulators for Dynamic Channel Power Equalization in WDM Systems," *Conference on Optical Fiber Communication (OFC)* '98, Paper PD-8, San Jose, Calif., February 1998.
145. J. Ford and J. Walker, "Dynamic Spectral Power Equalization Using Micro-Opto-Mechanics," *IEEE Photonics Technology Letters* **10**:1440–1442, 1998.
146. C. Doerr, C. Joyner, and L. Stulz, "Integrated WDM Dynamic Power Equalizer with Low Insertion Loss," *IEEE Photonics Technology Letters* **10**:1443–1445, 1998.
147. C. Doerr, C. Joyner, and L. Stulz, "16-Band Integrated Dynamic Gain Equalization Filter with Less than 2.8-dB Insertion Loss," *IEEE Photonics Technology Letters* **14**:334–336, 2002.
148. C. Fludger, A. Maroney, and N. Jolley, "An Analysis of the Improvements in OSNR from Distributed Raman Amplifiers Using Modern Transmission Fibres," *Conference on Optical Fiber Communication (OFC)* '00, Paper FF2, Baltimore, Md., March 2000.
149. J. Bromage, "Raman Amplification for Fiber Communications Systems," *IEEE Journal of Lightwave Technology* **22**:79–93, 2004.
150. Y. Emori, K. Tanaka, and S. Namiki, "100 nm Bandwidth Flat-Gain Raman Amplifiers Pumped and Gain-Equalised by 12 Wavelength-Channel WDM Laser Diode Unit," *IEEE Electronics Letters* **35**:1355–1356, 1999.
151. S. Namiki and Y. Emori, "Ultrabroad-Band Raman Amplifiers Pumped and Gain-Equalized by Wavelength-Division-Multiplexed Highpower Laser Diodes," *IEEE Journal of Selected Topics in Quantum Electronics* **7**:3–16, 2001.
152. S. Namiki, K. Seo, N. Tsukiji, and S. Shikii, "Challenges of Raman Amplification," *Proceedings of IEEE* **94**:1024–1035, 2006.
153. B. Zhu, L. E. Nelson, S. Stulz, A. H. Gnauck, C. Doerr, J. Leuthold, L. Grüner-Nielsen, M. O. Pedersen, J. Kim, R. Lingle Jr., et al., "6.4-Tb/s (160×42.7 Gb/s) Transmission with 0.8 bit/s/Hz Spectral Efficiency over 32×100 km of Fiber Using CSRZ-DPSK Format," *Conference on Optical Fiber Communication (OFC)* '03, Paper PD-19, Atlanta, Ga., March 2003.

This page intentionally left blank

SOLITONS IN OPTICAL FIBER COMMUNICATION SYSTEMS

Pavel V. Mamyshev

*Bell Laboratories—Lucent Technologies
Holmdel, New Jersey*

22.1 INTRODUCTION

To understand why optical solitons are needed in optical fiber communication systems, we should consider the problems that limit the distance and/or capacity of optical data transmission. A fiber-optic transmission line consists of a transmitter and a receiver connected with each other by a transmission optical fiber. Optical fibers inevitably have chromatic dispersion, losses (attenuation of the signal), and nonlinearity. Dispersion and nonlinearity can lead to the distortion of the signal. Because the optical receiver has a finite sensitivity, the signal should have a high-enough level to achieve error-free performance of the system. On the other hand, by increasing the signal level, one also increases the nonlinear effects in the fiber. To compensate for the fiber losses in a long distance transmission, one has to periodically install optical amplifiers along the transmission line. By doing this, a new source of errors is introduced into the system—an amplifier spontaneous emission noise. (Note that even ideal optical amplifiers inevitably introduce spontaneous emission noise.) The amount of noise increases with the transmission distance (with the number of amplifiers). To keep the signal-to-noise ratio (SNR) high enough for the error-free system performance, one has to increase the signal level and hence the potential problems caused by the nonlinear effects. Note that the nonlinear effects are proportional to the product of the signal power P and the transmission distance L , and both of these multipliers increase with the distance. Summarizing, we can say that all the problems—dispersion, noise, and nonlinearity—grow with the transmission distance. The problems also increase when the transmission bit rate (speed) increases. It is important to emphasize that it is very difficult to deal with the signal distortions when the nonlinearity is involved, because the nonlinearity can couple all the detrimental effects together [nonlinearity, dispersion, noise, polarization mode dispersion (i.e., random birefringence of the fiber), polarization-dependent loss/gain, etc.]. That happens when the nonlinear effects are out of control. The idea of soliton transmission is to guide the nonlinearity to the desired direction and use it for your benefit. When soliton pulses are used as an information carrier, the effects of dispersion and nonlinearity balance (or compensate) each other and thus don't degrade the signal quality with the propagation distance. In such a regime, the pulses propagate through the fiber without changing their spectral and temporal shapes. This mutual compensation of dispersion and nonlinear effects takes place continuously with the distance

in the case of “classical” solitons and periodically with the so-called dispersion map length in the case of dispersion-managed solitons. In addition, because of the unique features of optical solitons, soliton transmission can help to solve other problems of data transmission, like polarization mode dispersion. Also, when used with frequency guiding filters (sliding guiding filters in particular), the soliton systems provide continuous all-optical regeneration of the signal suppressing the detrimental effects of the noise and reducing the penalties associated with wavelength-division multiplexed (WDM) transmission. Because the soliton data looks essentially the same at different distances along the transmission, the soliton type of transmission is especially attractive for all-optical data networking. Moreover, because of the high quality of the pulses and return-to-zero (RZ) nature of the data, the soliton data is suitable for all-optical processing.

22.2 NATURE OF THE CLASSICAL SOLITON

Signal propagation in optical fibers is governed by the nonlinear Schroedinger equation (NSE) for the complex envelope of the electric field of the signal.¹⁻³ This equation describes the combined action of the self-phase modulation and dispersion effects, which play the major role in the signal evolution in most practical cases. Additional linear and nonlinear effects can be added to the modified NSE.⁴ Mathematically, one can say that solitons are stable solutions of NSE.^{1,2} In this paper, however, we will give a qualitative physical description of the soliton regimes of pulse propagation, trying to avoid mathematics as much as possible.

Consider first the effect of dispersion. An optical pulse of width τ has a finite spectral bandwidth $\text{BW} \approx 1/\tau$. When the pulse is transform limited, or unchirped, all the spectral components have the same phase. In time domain, one can say that all the spectral components overlap in time, or sit on top of each other (see Fig. 1). Because of the dispersion, different spectral components propagate in the fiber with different group velocities, V_{gr} . As a result of the dispersion action alone, the initial

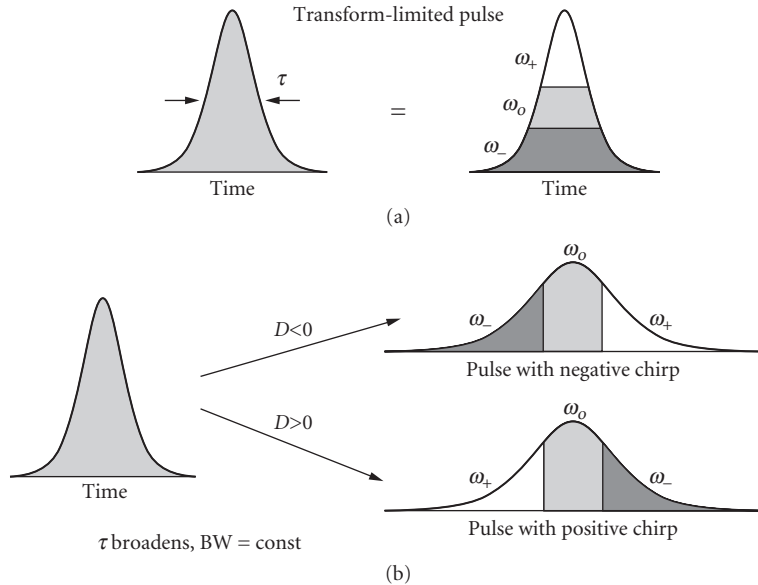


FIGURE 1 (a) Transform-limited pulse: all spectral components of the pulse “sit” on top of each other. (b) Effect of group velocity dispersion on a transform-limited pulse.

unchirped pulse broadens and gets chirped (frequency modulated). The sign of the chirp depends on the sign of the fiber group velocity dispersion (see Fig. 1).

$$D = d \left(\frac{1}{V_{gr}} \right) / d\lambda \quad (1)$$

(λ is the light wavelength). A characteristic fiber length called the *dispersion length*, at which the pulse broadens by a factor sqrt (2), is determined both by the fiber dispersion and the pulse width:

$$z_d = \frac{2\pi c 0.322 \tau^2}{\lambda^2 D} \quad (2)$$

(c is the speed of light). Note that the pulse spectral bandwidth remains unchanged because the dispersion is a linear effect.

Consider now the nonlinear effect of self-phase modulation (SPM).⁵ Due to the Kerr effect, the fiber refractive index depends on the signal intensity, $n(I) = n_0 + n_2 I$, where n_2 is the nonlinear refractive index and intensity is $I = P/A$, P is the signal power and A is the fiber effective cross-section mode area. During a pulse propagation through the fiber, different parts of the pulse acquire different values of the nonlinear phase shift: $\phi(t) = 2\pi/\lambda n_2 I(t)L$. Here $I(t)$ is the intensity pulse shape in time domain and L is the transmission distance. This time-dependent nonlinear phase shift means that different parts of the pulse experience different frequency shifts:

$$\delta\omega(t) = \frac{d\phi}{dt} = -\frac{2\pi}{\lambda} n_2 L \frac{dI(t)}{dt} \quad (3)$$

As one can see, the frequency shift is determined by the time derivative of the pulse shape. Because the nonlinear refractive index in silica-based fibers is positive, the self-phase modulation effect always shifts the front edge of the pulse to the “red” spectral region (downshift in frequency), and the trailing edge of the pulse to the “blue” spectral region (upshift in frequency). This means that an initially unchirped pulse spectrally broadens and gets negatively chirped (Fig. 2). A characteristic fiber length called the *nonlinear length*, at which the pulse spectrally broadens by a factor of two, is

$$z_{NL} = \left(\frac{2\pi}{\lambda} n_2 I_0 \right)^{-1} \quad (4)$$

Note that, when acting alone, SPM does not change the temporal intensity profile of the pulse.

As it was mentioned earlier, when under no control, both SPM and dispersion may be very harmful for the data transmission distorting considerably the spectral and temporal characteristics of the signal. Consider now how to control these effects by achieving the soliton regime of data transmission when the combined action of these effects results in a stable propagation of data pulses without changing their spectral and temporal envelopes.

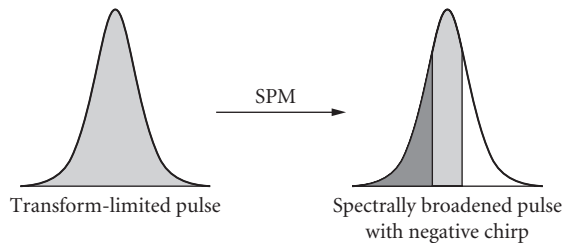


FIGURE 2 Effect of self-phase modulation on a transform-limited pulse.

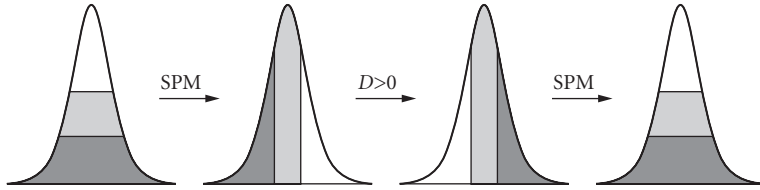


FIGURE 3 Qualitative explanation of classical soliton. Combined action of dispersion and nonlinearity (self-phase modulation) results in a stable pulse propagation with constant spectral and temporal widths. See text.

In our qualitative consideration, consider the combined action of dispersion and nonlinearity (SPM) as an alternative sequence of actions of dispersion and nonlinearity. Assume that we start with a chirp-free pulse (see Fig. 3). The self-phase modulation broadens the pulse spectrum and produces a negative frequency chirp: The front edge of the pulse becomes red-shifted, and the trailing edge becomes blue-shifted. When positive GVD is then applied to this chirped pulse, the red spectral components are delayed in time with respect to the blue ones. If the right amount of dispersion is applied, the sign of the pulse chirp can be reversed to negative: The blue spectral components shift in time to the front pulse edge, while the red spectral components move to the trailing edge. When the nonlinearity is applied again, it shifts the frequency of the front edge to the red spectral region and upshifts the frequency of the trailing edge. That means that the blue front edge becomes green again, the red trailing edge also becomes green, and the pulse spectrum bandwidth narrows to its original width. The described regime of soliton propagation is achieved when the nonlinear and dispersion effect compensate each other exactly. In reality, the effects of dispersion and SPM act simultaneously, so that the pulse spectral and temporal widths stay constant with the distance, and the only net effect is a (constant within the entire pulse) phase shift of 0.5 rad per dispersion length of propagation.⁶ The condition of the soliton regime is equality of the nonlinear and dispersion lengths: $z_d = z_{NL}$. One can rewrite this expression to find a relationship between the soliton peak power, pulse width, and fiber dispersion:

$$P_0 = \frac{\lambda^3 DA}{0.3224 \pi^2 n_2 \tau^2} \quad (5)$$

Here, P_0 is the soliton peak power and τ is the soliton FWHM. Soliton pulses have a sech^2 form. Note that as it follows from our previous consideration, classical soliton propagation in fibers requires a positive sign of the fiber's dispersion, D (assuming that n_2 is positive). Consider a numerical example. For a pulse of width $\tau = 20$ ps propagating in a fiber with $D = 0.5$ ps nm⁻¹ km⁻¹, fiber cross-section mode area $A = 50$ μm², $\lambda = 1.55$ μm, and typical value of $n_2 = 2.6$ cm²/W, one can find the soliton peak power is 2.4 mW. The dispersion length is $z_d = 200$ km in this case.

22.3 PROPERTIES OF SOLITONS

The most important property of optical solitons is their robustness.^{6–20} Consider what robustness means from a practical point of view. When a pulse is injected into the fiber, the pulse does not have to have the exact soliton shape and parameters [Eq. (5)] to propagate as a soliton. As long as the input parameters are not too far from the optimum, during the nonlinear propagation the pulse “readjusts” itself, shaping into a soliton and shedding off nonsoliton components. For example, an unchirped pulse of width τ will be reshaped into a single soliton as long as its input power P is greater than $P_0/4$ and less than $2.25P_0$. Here, P_0 is the soliton power determined by Eq. (5).³

Solitons are also robust with respect to the variations of the pulse energy and of the fiber parameters along the transmission line. As long as these variations are fast enough (period of perturbations is much smaller than the soliton dispersion length z_d), the soliton “feels” only the average values of these parameters. This feature is extremely important for practical systems. In particular, it makes it possible to use solitons in long distance transmission systems where fiber losses are periodically compensated by lumped amplifiers. As long as the amplifier spacing is much less than the soliton dispersion length $L_{\text{amp}} \ll z_d$, classical solitons work very well in these systems. Note that all soliton perturbations result in a loss of some part of the soliton energy, which is radiated into dispersive waves.

Consider now a case of slow variations of parameters along the transmission when a characteristic length at which a fiber parameter (or pulse energy) changes considerably is much longer than the soliton dispersion length. Soliton parameters follow adiabatically these changes. That means that all the parameters in Eq. (5) can be considered as distance dependent, and Eq. (5) remains valid. It can be rewritten in the following form:

$$\tau(z) = \text{const} \frac{D(z)A(z)}{P(z)\tau(z)} = \text{const} \frac{D(z)A(z)}{\text{energy}(z)} \quad (6)$$

One can derive many important consequences from this equation.^{13–22} One example would be the pulse broadening (and spectral narrowing) in a fiber with loss [assuming $D(z)$ and $A(z)$ are constant].^{13–15} Note that the soliton broadening can be used in repeaterless data transmission systems when high-input signal power is required.¹⁵ On the other hand, one can get a pulse compression in a fiber with adiabatic gain. Similar effects can be obtained by changing the fiber dispersion and/or mode area along the length. For example, adiabatic soliton compression can be obtained in a fiber with slowly decreasing dispersion (dispersion-tapered fiber).^{16–22}

It is important to emphasize that the adiabatic soliton propagation does not necessarily require that each of these parameters—pulse energy, fiber dispersion, and mode area—changes adiabatically with the distance, as long as the whole expression, $[D(z)A(z)]/[\text{energy}(z)]$ changes adiabatically with the distance. For example, soliton propagation in a dispersion-tapered fiber with losses is equivalent to transmission in a lossless, constant-dispersion fiber if the dispersion decreases with the same rate with the distance as the pulse energy [i.e., if $D(z)/\text{energy}(z) = \text{const}$]. Note that this is true no matter what the fiber loss and the pulse width are.

So far, we’ve been discussing a single pulse propagation. In communication systems, one has to deal with streams of pulses. When two or more soliton pulses propagate in the fiber at the same wavelength, they can interact with each other: Tails from one soliton pulse may overlap with the other pulse. Due to the cross-phase modulation effect, this overlap leads to the frequency shifts of the interacting solitons. The signs of the frequency shifts are opposite for the two solitons. Through the fiber dispersion, the frequency changes result in the changes of the soliton group velocities. The strength of the interaction decreases very fast with the soliton separation and for most practical applications can be considered to be negligible when the separation is 4 to 5 times greater than the soliton pulse width τ .^{23,24} The character of interaction depends on the mutual optical phases of the solitons: When they are the same, the solitons attract to each other; when they are out of phase, the solitons repel from each other; when the phase difference is $\pi/2$, the solitons do not interact.

22.4 CLASSICAL SOLITON TRANSMISSION SYSTEMS

The soliton properties described earlier determine the engineering rules for designing the soliton-based transmission systems. First, to make sure that every individual pulse is stable in the transmission line with constant fiber dispersion and loss periodically compensated by lump amplifiers, the amplifier spacing L_{amp} should be much smaller than the soliton dispersion length z_d . To avoid considerable pulse-to-pulse interaction, the minimum distance between adjacent pulses should be $T \geq 4\tau$, where $1/T$ is the transmission bit rate and τ is the soliton pulse width. The pulse power determined from Eq. (5) should be considered as a path-average power P_{av} . If the signal energy

decreases with the distance in the fiber spans between the amplifiers as $\exp(\gamma z)$ (here, γ is the loss rate), the path-average power is related to the pulse power at the output of each amplifier (input to the fiber span) P_{in} , as

$$P_0 = P_{\text{in}} \frac{1 - \exp(\gamma L_{\text{amp}})}{|\gamma| L_{\text{amp}}} \quad (7)$$

Here, L_{amp} is the amplifier spacing. As it was stated earlier, the dispersion and nonlinear effects “compensate” each other in the soliton regime of transmission, so that the pulses propagate practically without changing their temporal and spectral shapes. As long as the length scale of perturbations of the transmission parameters is much shorter than the soliton dispersion length, the pulses “feel” only the average parameters. Note, however, that perturbations may lead to shedding of dispersive waves by solitons.¹²

There are two main sources of errors in the soliton transmission systems: fluctuations of the pulse energies and fluctuations of the pulse arrival times.²⁵ The origin of the energy fluctuations is the same as in the other types of systems—spontaneous emission noise generated by the amplifiers. Each amplifier contributes a noise with a spectral density (power per unit bandwidth):

$$P_v = (G - 1) n_{\text{sp}} h \nu \quad (8)$$

Here, G is the power gain of the amplifier, $h \nu$ is the photon energy, and $n_{\text{sp}} \geq 1$ is the spontaneous emission factor that characterizes the quality of the amplifier. In the best case, when the amplifier is highly inverted, n_{sp} is close to unity. In a broadband transmission system (i.e., without in-line spectral filters), when the lumped amplifiers compensate exactly for the fiber loss, the noise grows linearly with the distance (with the number of amplifiers). At the output of a transmission line of length L , the path-averaged spectral density is

$$P_{\text{vav}} = |\gamma| L n_{\text{sp}} h \nu F(G) \quad (9)$$

Here, function $F(G)$ describes the penalty one has to pay for having high-gain amplifiers (or long amplifier spacing):

$$F(G) = \frac{(G - 1)^2}{G \ln^2 G} \quad (10)$$

The penalty function has its minimum [$F(G) = 1$] in the case of distributed amplification (when $G \rightarrow 1$) and grows with G . The SNR at the output of transmission should be high enough to have error-free transmission. Note that the noise spectral density P_v has units of energy. It is also the noise energy received in any time T in a spectral bandwidth $1/T$. That is why P_v is also called the *equipartition energy*. To have the error probability less than 10^{-9} and 10^{-15} , the ratio of the pulse energy to the equipartition energy should be, correspondently, 100 and 160. For example, consider a transmission system with the average loss of 0.21 dB/km, $n_{\text{sp}} = 1.5$, amplifier spacing of 50 km. The minimum pulse energy at the input of each fiber span to have the error probability less than 10^{-9} in such a system of length $L = 5000$ km is 20 fJ, and for $L = 10,000$ km, it is 40 fJ.

Another type of error in the soliton systems is the fluctuation in the pulse arrival times, or timing jitter. The timing jitter can be caused by several factors. The adjacent pulse-to-pulse interactions can cause the pulses to shift in time. As we have stated earlier, interaction problems can be practically eliminated by spacing the solitons in time by more than 4 or 5 of their width. A very important source of the timing jitter is the spontaneous emission noise. Every time the noise is added to the signal, it modulates the carrier frequencies of the solitons at random. The chromatic dispersion of the fiber then converts these frequency variations in a variation of the pulses' arrival times. This effect

is known as the *Gordon-Haus effect*.^{6,26} The variance of the timing jitter produced by the Gordon-Haus effect is

$$\sigma_{\text{GH}}^2 \approx 0.2n_2hn_{\text{sp}}F(G)\frac{|\gamma|}{A}\frac{D}{\tau}L^3 \quad (11)$$

An error occurs when a pulse arrives outside of the acceptance time window W of the detection system (this window is usually slightly less than the bit slot, T). To have the error probability less than 10^{-9} , the acceptance window should be greater than 12 standard deviations of the timing jitter:

$$W \geq 12\sigma_{\text{GH}} \quad (12)$$

The Gordon-Haus jitter limits the maximum bit rate and transmission distance. As one can see from Eq. (11), the jitter increases very fast with the distance; it also increases when τ decreases. Another factor that limits the maximum transmission distance is that σ_{GH}^2 is proportional to the pulse energy [because the pulse energy is proportional to (D/τ)], and long-distance transmission systems should have high-enough pulse energies to keep the SNR high. Consider a numerical example, $L = 9000$ km, $\tau = 20$ ps, $n_{\text{sp}} = 1.4$, $\gamma = -0.048$ km⁻¹, amplifier spacing = 30 km, $D = 0.5$ ps/(nm⁻¹ km⁻¹), $A = 50$ μm². Equation (11) then gives the standard deviation of the Gordon-Haus timing jitter $\sigma = 11.7$ ps. As one can see, according to Eq. (12), this jitter is too high for 10 Gbit/s transmission ($1/T = 100$ ps) to be error-free, because $12\sigma_{\text{GH}} > 1/T$ in this case.

Another source of the timing jitter is the acoustic interaction of pulses.²⁷⁻³⁰ Due to the electrostriction effect in the fiber, each propagating pulse generates an acoustic wave in the fiber. Other pulses experience the refractive index change caused by the acoustic wave. The resultant frequency changes of the pulses lead, through the effect of the fiber chromatic dispersion, to the fluctuation in the arrival times. The acoustic effect causes a “long-range” interaction: Pulses separated by a few nanoseconds can interact through this effect. One can estimate the acoustic timing jitter from the following simplified equation:

$$\sigma_a \approx 4.3 \frac{D^2}{\tau} (R - 0.99)^{1/2} L^2 \quad (13)$$

Here, standard deviation σ_a is in picoseconds; dispersion D is in picoseconds per nanometer per kilometer; the bit rate $R = 1/T$, is in gigabits per second; and the distance L , is in megameters. Equation (13) also assumes the fiber mode area of $A = 50$ μm². The acoustic jitter increases with the bit rate, and it has even stronger dependence on the distance than the Gordon-Haus jitter.

As it follows from the previous considerations, the timing jitter can impose severe limitations on the distance and capacity of the systems, and it has to be controlled.

22.5 FREQUENCY-GUIDING FILTERS

The Gordon-Haus and acoustic timing jitters originate from the frequency fluctuations of the pulses. That means that by controlling the frequency of the solitons, one can control the timing jitter as well. The frequency control can be done by periodically inserting narrowband filters (so-called frequency-guiding filters) along the transmission line, usually at the amplifier locations.^{31,32} If, for some reason, the center frequency of a soliton is shifted from the filter peak, the filter-induced differential loss across the pulse spectrum “pushes” the pulse frequency back to the filter peak. As a result, the pulse spectrum returns back to the filter peak in a characteristic damping length Δ . If the damping length is considerably less than the transmission distance L the guiding filters dramatically reduce the timing jitter. To calculate the timing jitter in a filtered system, one should replace L^3 by

$3L\Delta^2$ in Eq. (11), and L^2 in Eq. (13) should be replaced by $2L\Delta$. Then, we get the following expression for the Gordon-Haus jitter:

$$\sigma_{\text{GH},f}^2 \approx 0.6n_2hn_{\text{sp}}F(G)\frac{|\gamma|}{A}\frac{D}{\tau}L\Delta^2 \quad (14)$$

The damping properties of the guiding filters are determined mainly by the curvature of the filter response in the neighborhood of its peak. That means that shallow Fabry-Perot etalon filters can be used as the guiding filters. Fabry-Perot etalon filters have multiple peaks, and different peaks can be used for different WDM channels. The ability of the guiding filters to control the frequency jitter is determined both by the filter characteristics and by the soliton spectral bandwidth. In the case of Fabry-Perot filters with the intensity mirror reflectivity R , and the free spectral range (FSR), the damping length is

$$\Delta = 0.483(\tau \text{FSR})^2 \frac{(1-R)^2}{R} L_f \quad (15)$$

Here, L_f is the spacing between the guiding filters; usually, L_f equals the amplifier spacing L_{amp} .

Note that the Gordon-Haus and acoustic jitters are not specific for soliton transmission only. Any kind of transmission systems, including so-called linear transmission, are subject to these effects. However, the guiding filters can be used in the soliton systems only. Every time a pulse passes through a guiding filter, its spectrum narrows. Solitons can quickly recover their bandwidth through the fiber nonlinearity, whereas for a linear transmission the filter action continuously destroys the signal.

Note that even a more effective reduction of the timing jitter can be achieved if, in addition to the frequency-guiding filters, an amplitude and/or phase modulation at the bit rate is applied to the signal periodically with the distance. “Error-free” transmission over practically unlimited distances can be achieved in this case (1 million kilometers at 10 Gbit/s has been demonstrated).^{33,34} Nevertheless, this technique is not passive, high-speed electronics is involved, and the clock recovery is required each time the modulation is applied. Also, in the case of WDM transmission, all WDM channels have to be demultiplexed before the modulation and then multiplexed back afterward; each channel has to have its own clock recovery and modulator. As one can see, this technique shares many drawbacks of the electronic regeneration schemes.

The frequency-guiding filters can dramatically reduce the timing jitter in the systems. At the same time, though, in some cases they can introduce additional problems. Every time a soliton passes through the filter, it loses some energy. To compensate for this loss, the amplifiers should provide an additional (excess) gain. Under this condition, the spontaneous emission noise and other nonsoliton components with the spectrum in the neighborhood of the filter peak experience exponential growth with the distance, which reduces the SNR and can lead to the soliton instabilities. As a result, one has to use weak-enough filters to reduce the excess gain. In practice, the filter strength is chosen to minimize the total penalty from the timing jitter and the excess gain.

22.6 SLIDING FREQUENCY-GUIDING FILTERS

As one can see, the excess gain prevents one from taking a full advantage of guiding filters. By using the sliding frequency-guiding filters,³⁵ one can essentially eliminate the problems associated with the excess gain. The trick is very simple: The transmission peak of each guiding filter is shifted in frequency with respect to the peak of the previous filter, so that the center frequency slides with the distance with the rate of $f' = df/dz$. Solitons, thanks to the nonlinearity, can follow the filters and slide in frequency with the distance. But all unwanted linear radiation (e.g., spontaneous emission noise, nonsoliton components shedded from the solitons, etc.) cannot slide and eventually is killed by the filters. The sliding allows one to use strong guiding filters and even to reduce the amount of

noise at the output of transmission in comparison with the broadband (no guiding filters) case. The maximum filter strength³⁶ and maximum sliding rate³⁵ are determined by the soliton stability. The error-free transmission of 10 Gbit/s signal over 40,000 km and 20 Gbit/s over 14,000 km was demonstrated with the sliding frequency-guiding filters technique.^{37,38}

It is important to emphasize that by introducing the sliding frequency-guiding filters into the transmission line, one converts this transmission line into an effective, all-optical passive regenerator (compatible with WDM). Solitons with only one energy (and pulse width) can propagate stably in such a transmission line. The parameters of the transmission line (the filter strength, excess gain, fiber dispersion, and mode area) determine the unique parameters of these stable solitons. The system is opaque for a low-intensity radiation (noise, for example). However, if the pulse parameters at the input of the transmission line are not too far from the optimum soliton parameters, the transmission line reshapes the pulse into the soliton of that line. Note, again, that the parameters of the resultant soliton do not depend on the input pulse parameters, but only on the parameters of the transmission line. Note also that all nonsoliton components generated during the pulse reshaping are absorbed by the filters. That means, in particular, that the transmission line removes the energy fluctuations from the input data signal.⁶ Note that the damping length for the energy fluctuations is close to the frequency damping length of Eq. (15). A very impressive demonstration of regenerative properties of a transmission line with the frequency-guiding filters is the conversion of a nonreturn-to-zero (NRZ) data signal (frequency modulated at the bit rate) into a clean soliton data signal.³⁹ Another important consequence of the regenerative properties of a transmission line with the frequency-guiding filters is the ability to self-equalize the energies of different channels in WDM transmission.⁴⁰ Negative feedback provided by frequency-guiding filters locks the energies of individual soliton channels to values that do not change with distance, even in the face of considerable variation in amplifier gain among the different channels. The equilibrium values of the energies are independent of the input values. All these benefits of sliding frequency-guiding filters are extremely valuable for practical systems. Additional benefits of guiding filters for WDM systems will be discussed later.

22.7 WAVELENGTH DIVISION MULTIPLEXING

Due to the fiber chromatic dispersion, pulses from different WDM channels propagate with different group velocities and collide with each other.⁴¹ Consider a collision of two solitons propagating at different wavelengths (different channels). When the pulses are initially separated and the fast soliton (the soliton at shorter wavelength, with higher group velocity) is behind the slow one, the fast soliton eventually overtakes and passes through the slow soliton. An important parameter of the soliton collision is the collision length L_{coll} , the fiber length at which the solitons overlap with each other. If we let the collision begin and end with the overlap of the pulses at half power points, then the collision length is

$$L_{\text{coll}} = \frac{2\tau}{D\Delta\lambda} \quad (16)$$

Here, $\Delta\lambda$ is the solitons wavelengths difference. Due to the effect of cross-phase modulation, the solitons shift each other's carrier frequency during the collision. The frequency shifts for the two solitons are equal in amplitudes (if the pulse widths are equal) and have opposite signs. During the first half of collision, the fast accelerates even faster (carrier frequency increases), while the slow soliton slows down. The maximum frequency excursion δf_{max} , of the solitons is achieved in the middle of the collision, when the pulses completely overlap with each other:

$$\delta f_{\text{max}} = \pm \frac{1}{3\pi^2 0.322 \Delta f \tau^2} = \pm \frac{1.18 n_2 \epsilon}{A \tau D \Delta \lambda} \quad (17)$$

Here, $\Delta f = -c \Delta\lambda / \lambda^2$ is the frequency separation between the solitons, and $\epsilon = 1.13 P_0 \tau$ is the soliton energy. In the middle of collision, the accelerations of the solitons change their signs. As a result, the

frequency shifts in the second half of collision undo the frequency shifts of the first half, so that the soliton frequency shifts go back to zero when the collision is complete. This is a very important and beneficial feature for practical applications. The only residual effect of complete collision in a lossless fiber is the time displacements of the solitons:

$$\delta t_{cc} = \pm \frac{0.1786}{\Delta f^2 \tau} = \pm \frac{2\epsilon n_2 \lambda}{cDA \Delta \lambda^2} \quad (18)$$

The symmetry of the collision can be broken if the collision takes place in a transmission line with loss and lumped amplification. For example, if the collision length L_{coll} is shorter than the amplifier spacing L_{amp} , and the center of collision coincides with the amplifier location, the pulses intensities are low in the first half of collision and high in the second half. As a result, the first half of collision is practically linear. The soliton frequency shifts acquired in the first half of collision are very small and insufficient to compensate for the frequency shifts of opposite signs acquired by the pulses in the second half of collision. This results in nonzero residual frequency shifts. Note that similar effects take place when there is a discontinuity in the value of the fiber dispersion as a function of distance. In this case, if a discontinuity takes place in the middle of collision, one half of the collision is fast (where D is higher) and the other half is slow. The result is nonzero residual frequency shifts. Nonzero residual frequency shifts lead, through the dispersion of the rest of the transmission fiber, to variations in the pulses arrival time at the output of transmission. Nevertheless, if the collision length is much longer than the amplifier spacing and of the characteristic length of the dispersion variations in the fiber, the residual soliton frequency shifts are zero, just like in a lossless uniform fiber. In practice, the residual frequency shifts are essentially zero as long as the following condition is satisfied:⁴¹

$$L_{\text{coll}} \geq 2L_{\text{amp}} \quad (19)$$

Another important case is so-called half-collisions (or partial collisions) at the input of the transmission.⁴² These collisions take place if solitons from different channels overlap at the transmission input. These collisions result in residual frequency shifts of δf_{max} and the following pulse timing shifts δt_{pc} at the output of transmission of length L :

$$\delta t_{\text{pc}} \approx \delta f_{\text{max}} \frac{\lambda^2}{c} D(L - L_{\text{coll}}/4) = \pm \frac{1.18\epsilon n_2 \lambda}{c\tau A \Delta \lambda} (L - L_{\text{coll}}/4) \quad (20)$$

One can avoid half-collisions by staggering the pulse positions of the WDM channels at the transmission input.

Consider now the time shifts caused by all complete collisions. Consider a two-channel transmission, where each channel has a $1/T$ bit rate. The distance between subsequent collisions is

$$l_{\text{coll}} = \frac{T}{D \Delta \lambda} \quad (21)$$

The maximum number of collisions that each pulse can experience is L/l_{coll} . This means that the maximum time shift caused by all complete collisions is

$$\delta t_{\Sigma cc} \approx \delta t_{cc} L/l_{\text{coll}} = \pm \frac{2\epsilon n_2 \lambda}{cTA \Delta \lambda} L \quad (22)$$

It is interesting to note that $\delta t_{\Sigma cc}$ does not depend on the fiber dispersion. Note also that Eq. (22) describes the worst case when the pulse experiences the maximum number of possible collisions. Consider a numerical example. For a two-channel transmission, 10 Gbit/s each ($T = 100$ ps), pulse energy ($\epsilon = 50$ fJ), channel wavelength separation ($\Delta \lambda = 0.6$ nm), fiber mode area ($A = 50 \mu\text{m}^2$ and $L = 10$ mm), we find $\delta t_{\Sigma cc} = 45$ ps. Note that this timing shift can be reduced by increasing the

channel separation. Another way to reduce the channel-to-channel interaction by a factor of two is to have these channels orthogonally polarized to each other. In WDM transmission, with many channels, one has to add timing shifts caused by all other channels. Note, however, that as one can see from Eq. (22), the maximum penalty comes from the nearest neighboring channels.

As one can see, soliton collisions introduce additional jitter to the pulse arrival time, which can lead to considerable transmission penalties. As we saw earlier, the frequency-guiding filters are very effective in suppressing the Gordon-Haus and acoustic jitters. They can also be very effective in suppressing the timing jitter induced by WDM collisions. In the ideal case of parabolical filters and the collision length being much longer than the filter spacing $L_{\text{coll}} \gg L_f$, the filters make the residual time shift of a complete collision δt_{cc} exactly zero. They also considerably reduce the timing jitter associated with asymmetrical collisions and half-collisions. Note that for the guiding filters to work effectively in suppressing the collision penalties, the collision length should be at least a few times greater than the filter spacing. Note also that real filters, such as etalon filters, do not always perform as good as ideal parabolic filters. This is true especially when large-frequency excursions of solitons are involved, because the curvature of a shallow etalon filter response reduces with the deviation of the frequency from the filter peak. In any case, filters do a very good job in suppressing the timing jitter in WDM systems.

Consider now another potential problem in WDM transmission, which is the four-wave mixing. During the soliton collisions, the four-wave mixing spectral sidebands are generated. Nevertheless, in the case of a lossless, constant-dispersion fiber, these sidebands exist only during the collision, and when the collision is complete, the energy from the sidebands regenerates back into the solitons. That is why it was considered for a long time that the four-wave mixing should not be a problem in soliton systems. But this is true only in the case of a transmission in a lossless fiber. In the case of lossy fiber and periodical amplification, these perturbations can lead to the effect of the pseudo-phase-matched (or resonance) four-wave mixing.⁴³ The pseudo-phase-matched four-wave mixing lead to the soliton energy loss to the spectral sidebands and to a timing jitter (we called that effect an *extended Gordon-Haus effect*).⁴³ The effect can be so strong that even sliding frequency-guiding filters are not effective enough to suppress it. The solution to this problem is to use dispersion-tapered fiber spans. As we have discussed earlier, soliton propagation in the condition:

$$\frac{D(z)A(z)}{\text{Energy}(z)} = \text{const} \quad (23)$$

is identical to the case of lossless, constant-dispersion fiber. That means that the fiber dispersion in the spans between the amplifiers should decrease with the same rate as the signal energy. In the case of lumped amplifiers, this is the exponential decay with the distance. Note that the dispersion-tapered spans solve not just the four-wave mixing problem. By making the soliton transmission perturbation-free, they lift the requirements to have the amplifier spacing much shorter than the soliton dispersion length. The collisions remain symmetrical even when the collision length is shorter than the amplifier spacing. (Note, however, that the dispersion-tapered fiber spans do not lift the requirement to have guiding filter spacing as short as possible in comparison with the collision length and with the dispersion length.) The dispersion-tapered fiber spans can be made with the present technology.²² Stepwise approximation of the exact exponential taper made of fiber pieces of constant dispersion can also be used.⁴³ It was shown numerically and experimentally that by using fiber spans with only a few steps one can dramatically improve the quality of transmission.^{44,45} In the experiment, each fiber span was dispersion tapered typically in three or four steps, the path-average dispersion value was $0.5 \pm 0.05 \text{ ps nm}^{-1} \text{ km}^{-1}$ at 1557 nm. The use of dispersion-tapered fiber spans together with sliding frequency-guiding filters allowed transmission of eight 10-Gbit/s channels with the channel spacing, $\Delta\lambda = 0.6 \text{ nm}$, over more than 9000 km. The maximum number of channels in this experiment was limited by the dispersion slope, $dD/d\lambda$, which was about $0.07 \text{ ps nm}^{-2} \text{ km}^{-1}$. Because of the dispersion slope, different WDM channels experience different values of dispersion. As a result, not only the path average dispersion changes with the wavelength, but the dispersion tapering has exponential behavior only in a vicinity of one particular wavelength in the center of the transmission band. Wavelength-division multiplexed channels located far from that wavelength propagate in far from

the optimal conditions. One solution to the problem is to use dispersion-flattened fibers (i.e., fibers with $dD/d\lambda = 0$). Unfortunately, these types of fibers are not commercially available at this time. This and some other problems of classical soliton transmission can be solved by using dispersion-managed soliton transmission.^{46–63}

22.8 DISPERSION-MANAGED SOLITONS

In the dispersion-managed (DM) soliton transmission, the transmission line consists of the fiber spans with alternating signs of the dispersion. Let the positive and negative dispersion spans of the map have lengths and dispersions, L_+ , D_+ and L_- , D_- , respectively. Then, the path-average dispersion D_{av} is

$$D_{av} = (D_+ L_+ + L_- D_-) / L_{map} \quad (24)$$

Here, L_{map} is the length of the dispersion map:

$$L_{map} = L_+ + L_- \quad (25)$$

Like in the case of classical soliton, during the DM soliton propagation, the dispersion and non-linear effects cancel each other. The difference is that in the classical case, this cancellation takes place continuously, whereas in the DM case, it takes place periodically with the period of the dispersion map length L_{map} . The strength of the DM is characterized by a parameter S , which is determined as^{47,50,52}

$$S = \frac{\lambda^2}{2\pi c} \frac{(D_+ - D_{av})L_+ - (D_- - D_{av})L_-}{\tau^2} \quad (26)$$

The absolute values of the local dispersion are usually much greater than the path average dispersion: $|D_+|, |D_-| \gg |D_{av}|$. As one can see from Eq. (26), the strength of the map is proportional to the number of the local dispersion lengths of the pulse in the map length: $S \approx L_{map} / z_{d,local}$. The shape of the DM solitons are close to Gaussian. A very important feature of DM solitons is the so-called power enhancement. Depending on the strength of the map, the pulse energy of DM solitons, $\mathcal{E}_{DM}$, is greater than that of classical solitons, $\mathcal{E}_0$, propagating in a fiber with constant dispersion, $D = D_{av}$.^{47,50}

$$\mathcal{E}_{DM} \approx \mathcal{E}_0 (1 + 0.75S^2) \quad (27)$$

Note that this equation assumes lossless fiber. The power enhancement effect is very important for practical applications. It provides an extra degree of freedom in the system design by giving the possibility to change the pulse energy while keeping the path-average fiber dispersion constant. In particular, because DM solitons can have adequate pulse energy (to have a high-enough SNR) at or near zero path average dispersion, timing jitter from the Gordon-Haus and acoustic effects is greatly reduced (e.g., the variance of the Gordon-Haus jitter, σ^2 , scales almost as $1/\mathcal{E}_{DM}$).⁴⁹ Single-channel high-bit-rate DM soliton transmission over long distances with weak guiding filters and without guiding filters was experimentally demonstrated.^{46,51}

Dispersion-managed soliton transmission is possible not only in transmission lines with positive dispersion, $D_{av} > 0$, but also in the case of $D_{av} = 0$ and even $D_{av} < 0$.⁵² To understand this, consider qualitatively the DM soliton propagation (Fig. 4). Locally, the dispersive effects are always stronger than the nonlinear effect (i.e., the local dispersion length is much shorter than the nonlinear length). In the zero approximation, the pulse propagation in the map is almost linear. Let's call the middle of

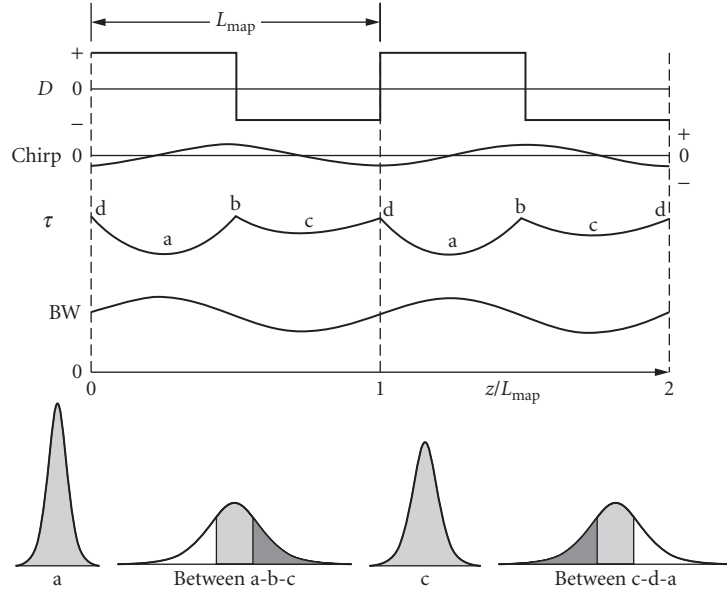


FIGURE 4 Qualitative description of dispersion-managed (DM) soliton transmission. Distance evolution of the fiber dispersion $[D(z)]$, pulse chirp, pulse width $[\tau(z)]$, and pulse bandwidth $[BW(z)]$. Evolution of the pulse shape in different fiber sections is shown in the bottom.

the positive D sections “point a,” the middle of the negative sections “point c,” transitions between positive and negative sections “point b,” and transitions between negative and positive sections “point d.” The chirp-free (minimum pulse width) positions of the pulse are in the middle of the positive- and negative- D sections (points a and c). The pulse chirp is positive between points a, b, and c (see Fig. 4). That means that the high-frequency (blue) spectral components of the pulse are at the front edge of the pulse, and the low-frequency (red) components are at the trailing edge. In the section c-d-a, the pulse chirp is negative. The action of the nonlinear SPM effect always downshifts in frequency the front edge of the pulse and up shifts in frequency the trailing edge of the pulse. That means that the nonlinearity decreases the spectral bandwidth of positively chirped pulses (section a-b-c) and increases the spectral bandwidth of negatively chirped pulses (section c-d-a). This results in the spectral bandwidth behavior also shown in Fig. 4: The maximum spectral bandwidth is achieved in the chirp-free point in the positive section, whereas the minimum spectral bandwidth is achieved in the chirp-free point in the negative section. The condition for the pulses to be DM solitons is that the nonlinear phase shift is compensated by the dispersion-induced phase shift over the dispersion map length. That requires that $\int DBW^2 dz > 0$ (here, BW is the pulse spectral bandwidth). Note that in the case of classical solitons, when spectral bandwidth is constant, this expression means that dispersion D must be positive. In the DM case, however, the pulse bandwidth is wider in the positive- D section than in the negative- D section. As a result, the integral can be positive, even when $D_{av} = \int D dz / L_{map}$ is zero or negative. Note that the spectral bandwidth oscillations explain also the effect of power enhancement of DM solitons.

Consider interaction of adjacent pulses in DM systems.⁵⁴ The parameter that determines the strength of the interaction is the ratio τ/T (here, τ is the pulse width and T is the spacing between adjacent pulses). As in the case of classical soliton transmission, the cross-phase modulation effect (XPM) shifts the frequencies of the interacting pulses, Δf_{XPM} , which, in turn, results in timing jitter at the output of the transmission. As it was discussed earlier, the classical soliton interaction increases

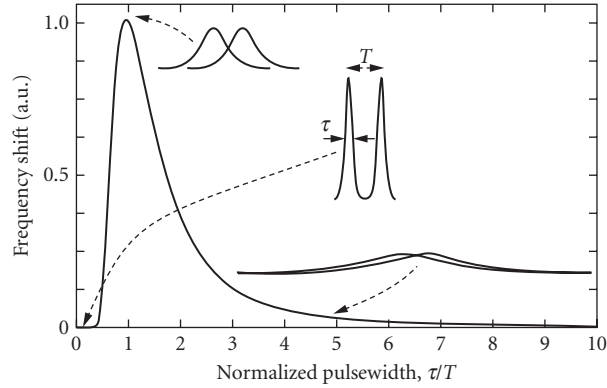


FIGURE 5 Dimensionless function $\Phi(\tau/T)$ describing the XPM-induced frequency shift of two interacting chirped Gaussian pulses as a function of the pulse width normalized to the pulse separation.

very quickly with τ/T . To avoid interaction-induced penalties in classical soliton transmission systems, the pulses should not overlap significantly with each other: τ/T should be less than 0.2 to 0.3. In the DM case, the situation is different. The pulse width in the DM case oscillates with the distance $\tau(z)$; that means that the interaction also changes with distance. Also, because the pulses are highly chirped when they are significantly overlapped with each other, the sign of the interaction is essentially independent of the mutual phases of the pulses. Cross-phase modulation always shifts the leading pulse to the red spectral region, and the trailing pulse shifts to the blue spectral region. The XPM-induced frequency shifts of interacting solitons per unit distance is

$$\frac{d\Delta f_{\text{XPM}}}{dz} \approx \pm 0.15 \frac{2\pi n_2 \epsilon}{\lambda T^2 A} \Phi(\tau/T) \quad (28)$$

The minus sign in Eq. (28) corresponds to the leading pulse, and the plus sign corresponds to the trailing pulse. Numerically calculated dimensionless function, $\Phi(\tau/T)$, is shown in Fig. 5. As it follows from Eq. (28), $\Phi(\tau/T)$ describes the strength of the XPM-induced interaction of the pulses as a function of the degree of the pulse overlap. One can see that the interaction is very small when τ/T is smaller than 0.4 (i.e., when the pulses barely overlap), which is similar to the classical soliton propagation. The strength of the interaction of DM solitons also increases with τ/T , but only in the region $0 < \tau/T < 1$. In fact, the interaction reaches its maximum at $\tau/T \approx 1$ and then decreases and becomes very small again when $\tau/T \gg 1$ (i.e., when the pulses overlap nearly completely). There are two reasons for such an interesting behavior at $\tau/T \gg 1$. The XPM-induced frequency shift is proportional to the time derivative of the interacting pulse's intensity, and the pulse derivative reduces with the pulse broadening. Also, when the pulses nearly completely overlap, the sign of the derivative changes across the region of overlap so that the net effect tends to be canceled out.

Based on Eq. (28) and Fig. 5, one can distinguish three main regimes of data transmission in DM systems. In all these regimes, the minimum pulse width is, of course, less than the bit slot, T . The regimes differ from each other by the maximum pulse breathing with the distance. In the first, “non-pulse-overlapped,” regime, adjacent pulses barely overlap during most of the transmission, so that the pulse interaction is not a problem in this case. This is the most stable regime of transmission. In the “partially-pulse-overlapped” regime, the adjacent pulses spend a considerable portion of the transmission being partially overlapped [$\tau(z)$ being around T]. Cross-phase modulation causes the frequency and timing jitter in this case. In the third, “pulse-overlapped,” regime, the adjacent pulses are almost completely overlapped with each other during most of the transmission [$\tau_{\min}(L_{\text{map}}/z_{d,\text{local}}) \gg T$].

The XPM-induced pulse-to-pulse interaction is greatly reduced in this case in comparison with the previous one. The main limiting factor for this regime of transmission is the intrachannel four-wave mixing taking place during strong overlap of adjacent pulses.⁵⁴ The intrachannel four-wave mixing leads to the amplitude fluctuations of the pulses and “ghost” pulse generation in the “zero” slots of the data stream.

22.9 WAVELENGTH-DIVISION MULTIPLEXED DISPERSION-MANAGED SOLITON TRANSMISSION

One of the advantages of DM transmission over classical soliton transmission is that the local dispersion can be very high ($|D_+|, |D_-| \gg |D_{av}|$), which efficiently suppresses the four-wave mixing from soliton-soliton collisions in WDM. Consider the timing jitter induced by collisions in the non-pulse-overlapped DM transmission. The character of the pulse collisions in DM systems is quite different from the case of a transmission line with uniform dispersion: In the former, the alternating sign of the high local dispersion causes the colliding solitons to move rapidly back and forth with respect to each other, with the net motion determined by D_{av} .^{56–59} Because of this rapid breathing of the distance between the pulses, each net collision actually consists of many fast or “mini” collisions. The net collision length can be estimated as⁵⁹

$$L_{\text{coll}} \approx \frac{2\tau}{D_{av}\Delta\lambda} + \frac{(D_+ - D_{av})L_+}{D_{av}} \approx \frac{2\tau}{D_{av}\Delta\lambda} + \frac{\tau_{\text{eff}}}{D_{av}\Delta\lambda} \quad (29)$$

Here, τ is the minimum (unchirped) pulse width. Here, we also defined the quantity $\tau_{\text{eff}} \equiv L_+ D_+ \Delta\lambda$, which plays the role of an effective pulse width. For strong dispersion management, τ_{eff} is usually much bigger than τ . Thus, L_{coll} becomes almost independent of $\Delta\lambda$ and much longer than it is for classical solitons subject to the same D_{av} . As a result, the residual frequency shift caused by complete pulse collisions tends to become negligibly small for transmission using strong maps.⁵⁸ The maximum frequency excursion during the DM soliton collision is⁵⁹

$$\delta f_{\text{max}} \approx \pm \frac{2n_2\epsilon}{L_+ D_+ A D_{av} \lambda \Delta\lambda^2} = \pm \frac{2n_2\epsilon}{A D_{av} \lambda \Delta\lambda \tau_{\text{eff}}} \quad (30)$$

Now, we can estimate the time shift of the solitons per complete collision:

$$\delta t_{\text{cc}} \approx D_{av} \lambda^2 / c \int \delta f dz \approx \alpha D_{av} L_{\text{coll}} \delta f_{\text{max}} \lambda^2 / c \approx \pm \alpha \frac{2n_2\epsilon\lambda}{c A D_{av} \Delta\lambda^2} \quad (31)$$

Here, $\alpha \leq 1$ is a numerical coefficient that takes into account the particular shape of the frequency shift as a function of distance. Consider now the time shifts caused by all collisions. In a two-channel transmission, the distance between subsequent collisions is $l_{\text{coll}} = T / (D_{av} \Delta\lambda)$. The maximum number of complete collisions at the transmission distance L is $(L - L_{\text{coll}}) / l_{\text{coll}}$ (we assume that $L > L_{\text{coll}}$), and the number of incomplete collisions at the end of transmission is $L_{\text{coll}} / l_{\text{coll}}$. The timing shift caused by all these collisions can be estimated as

$$\delta t_{\Sigma c} \approx \delta t_{\text{cc}} (L - L_{\text{coll}}/2) / l_{\text{coll}} = \pm \alpha \frac{2n_2\epsilon\lambda}{c A T \Delta\lambda} (L - L_{\text{coll}}/2) \quad (32)$$

Consider the problem of initial partial collisions. As it was discussed earlier for the case of classical solitons, initial partial collisions can be a serious problem by introducing large timing jitter at the output of transmission. On the other hand, for the classical case, one could avoid the half-collisions

by staggering the pulse positions of the WDM channels at the transmission input. The situation is very different for the DM case. In the DM case, the collision length is usually longer than the distance between subsequent collisions (i.e., $L_{\text{coll}} > l_{\text{coll}}$). Thus, a pulse can collide *simultaneously* with several pulses of another channel. The maximum number of such simultaneous collisions is $N_{\text{sc}} \approx L_{\text{coll}}/l_{\text{coll}} = 2\tau/T + [(D_+ - D_{\text{av}})L_c \Delta\lambda]/T$. Note that N_{sc} increases when the channel spacing $\Delta\lambda$ increases. The fact that the collision length is greater than the distance between collisions also means that initial partial collisions are inevitable in DM systems. Moreover, depending on the data pattern in the interacting channel, each pulse can experience up to N_{sc} initial partial collisions with that channel (not just one as in the classical case). As a consequence, the residual frequency shifts can be bigger than δf_{max} . The total time shift caused by the initial partial collisions at distance $L > L_{\text{coll}}$ can be estimated as

$$\delta\tau_{\text{pc}} = \beta \delta f_{\text{max}} N_{\text{sc}} (L - L_{\text{coll}}/2) D_{\text{av}} \lambda^2/c \approx \pm \beta \frac{2n_2 \varepsilon \lambda}{c A T \Delta \lambda} (L - L_{\text{coll}}/2) \quad (33)$$

Here, $\beta \leq 1$ is a numerical coefficient that takes into account the particular shape of the frequency shift as a function of distance for a single collision.

Equations (32) and (33) assume that the transmission distance is greater than the collision length. When $L > L_{\text{coll}}$, these equations should be replaced by

$$\delta t_{\Sigma, \text{pc}} \approx (\alpha, \beta) D_{\text{av}} \delta f_{\text{max}} \frac{\lambda^2}{c} \frac{L^2}{2l_{\text{coll}}} \approx \pm (\alpha, \beta) \frac{n_2 \varepsilon \lambda}{c A T \Delta \lambda} \frac{L^2}{L_{\text{coll}}} \quad (34)$$

Note that the signs of the timing shifts caused by initial partial collisions and by complete collisions are opposite. Thus, the maximum (worst-case) spread of the pulse arriving times caused by pulse collisions in the two-channel WDM transmission is described by:

$$\delta t_{\text{max}} = |\delta t_{\text{pc}}| + |\delta t_{\Sigma c}| \quad (35)$$

In a WDM transmission with more than two channels, one has to add contributions to the time shift from all the channels. Note that the biggest contribution makes the nearest neighboring channels, because the time shift is inversely proportional to the channel spacing, $\Delta\lambda$. Now, we can summarize the results of Eqs. (32) through (35) as follows. When $L > L_{\text{coll}}$ [Eqs. (32) to (33)], corresponding to very long distance transmission, δt_{max} increases linearly with the distance and almost independently of the path-average dispersion D_{av} . When $L < L_{\text{coll}}$ [Eq. (34)], which corresponds to short-distance transmission and/or very low path-average dispersion, δt_{max} increases quadratically with the distance and in proportion to D_{av} . Note also that the WDM data transmission at near zero path-averaged dispersion $D_{\text{av}} = 0$, may not be desirable, because $L_{\text{coll}} \rightarrow \infty$ and frequency excursions $\delta f_{\text{max}} \rightarrow \infty$ when $D \rightarrow 0$ [see Eq. (30)]. Thus, even though Eq. (34) predicts the time shift to be zero when D_{av} is exactly zero, the frequency shifts of the solitons can be unacceptably large and Eq. (34) may be no longer valid. There are also practical difficulties in making maps with $D_{\text{av}} < 0.1 \text{ ps nm}^{-1} \text{ km}^{-1}$ over the wide spectral range required for dense WDM transmission.

It is interesting to compare these results with the results for the case of classical solitons [Eqs. (17) to (22)]. The time shifts per complete collisions [Eqs. (18) and (31)] are about the same, the time shifts from all initial partial collisions [Eqs. (20) and (33)] are also close to each other. The total maximum time shifts from all collisions are also close to each other for the case of long distance transmission. That means that, similar to the classical case, one has to control the collision-induced timing jitter when it becomes too large. As it was discussed earlier, the sliding frequency-guiding filters are very effective in suppressing the timing jitter. Because the collision length in DM systems is much longer than in classical systems, and, at the same time, it is almost independent of the channel wavelength separation, the requirement that the collision length is much greater than the filter spacing, $L_{\text{coll}} \gg L_p$, is easy to meet. As a result, the guiding filters suppress the timing jitter in DM systems even more

effective than in classical soliton systems. The fact that the frequency excursions during collisions are much smaller in DM case, also makes the filters to work more effectively.

As we have discussed previously, many important features of DM solitons come from the fact that the soliton spectral bandwidth oscillates with the distance. That is why guiding filters alter the dispersion management itself and give an additional degree of freedom in the system design.⁶⁰ Note also that the position of the filters in the dispersion map can change the soliton stability in some cases.⁶¹ It should also be noted that because of the weak dependence of the DM soliton spectral bandwidth on the soliton pulse energy, the energy fluctuations damping length provided by the guided filters is considerably longer than the frequency damping length.⁶² This is the price one has to pay for many advantages of DM solitons. From the practical point of view, the most important advantage is the flexibility in system design and freedom in choosing the transmission fibers. For example, one can upgrade existing systems by providing an appropriate dispersion compensation with dispersion compensation fibers or with lumped dispersion compensators (fiber Bragg gratings, for example). The biggest advantage of DM systems is the possibility to design dispersion maps with essentially zero dispersion slope of the path-average dispersion, $dD_{av}/d\lambda$, by combining commercially available fibers with different signs of dispersion and dispersion slopes. (Note that it was a nonzero dispersion slope that limited the maximum number of channels in classical soliton long distance WDM transmission.) This was demonstrated in the experiment where almost flat average dispersion $D_{av} = 0.3 \text{ ps nm}^{-1} \text{ km}^{-1}$ was achieved by combining standard, dispersion-compensating, and True-Wave (Lucent nonzero dispersion-shifted) fibers.⁶³ By using sliding frequency-guiding filters and this dispersion map, “error-free” DM soliton transmission of twenty-seven 10-Gbit/s WDM channels was achieved over more than 9000 km without using forward error correction. It was shown that once the error-free transmission with about 10 channels is achieved, adding additional channels practically does not change performance of the system. (This is because, for each channel, only the nearest neighboring channels degrade its performance.) The maximum number of WDM channels in this experiment was limited only by the power and bandwidth of optical amplifiers used in the experiment. One can expect that the number of channels can be increased by a few times if more powerful and broader-bandwidth amplifiers are used.

22.10 CONCLUSION

We considered the basic principles of soliton transmission systems. The main idea of the “soliton philosophy” is to put under control, balance, and even to extract the maximum benefits from otherwise detrimental effects of the fiber dispersion and nonlinearity. The “soliton approach” is to make transmission systems intrinsically stable. Soliton technology is a very rapidly developing area of science and engineering, which promises a big change in the functionality and capacity of optical data transmission and networking.

22.11 REFERENCES

1. V. E. Zaharov and A. B. Shabat, “Exact Theory of Two Dimensional Self Focusing and One-Dimensional Self-Modulation of Waves in Nonlinear Media,” *Zh. Eksp. Teor. Fiz.* **61**:118–134 (1971) [*Sov. Phys. JETP* **34**:62–69 (1972)].
2. A. Hasegawa and F. D. Tappert, “Transmission of Stationary Nonlinear Optical Pulses in Dispersive Dielectric Fibers. I. Anomalous Dispersion,” *Appl. Phys. Lett.* **23**:142–144 (1973).
3. J. Satsuma and N. Yajima, “Initial Value Problem of One-Dimensional Self-Modulation of Nonlinear Waves in Dispersive Media,” *Prog. Theor. Phys. Suppl.* **55**:284–306 (1980).
4. P. V. Mamyshev and S. V. Chernikov, “Ultrashort Pulse Propagation in Optical Fibers,” *Opt. Lett.* **15**:1076–1078 (1990).

5. R. H. Stolen, in *Optical Fiber Telecommunications*, S. E. Miller and H. E. Chynoweth (eds.), Academic Press, New York, 1979, Chap. 5.
6. L. F. Mollenauer, J. P. Gordon, and P. V. Mamyshev, "Solitons in High Bit Rate, Long Distance Transmission," in *Optical Fiber Telecommunications III*, Academic Press, New York, 1997, Chap. 12.
7. L. F. Mollenauer, J. P. Gordon, and M. N. Islam, "Soliton Propagation in Long Fibers with Periodically Compensated Loss," *IEEE J. Quantum Electron.* **QE-22**:157 (1986).
8. L. F. Mollenauer, M. J. Neubelt, S. G. Evangelides, J. P. Gordon, J. R. Simpson, and L. G. Cohen, "Experimental Study of Soliton Transmission Over More Than 10,000 km in Dispersion Shifted Fiber," *Opt. Lett.* **15**:1203 (1990).
9. L. F. Mollenauer, S. G. Evangelides, and H. A. Haus, "Long Distance Soliton Propagation Using Lumped Amplifiers and Dispersion Shifted Fiber," *J. Lightwave Technol.* **9**:194 (1991).
10. K. J. Blow and N. J. Doran, "Average Soliton Dynamics and the Operation of Soliton Systems with Lumped Amplifiers," *Photonics Tech. Lett.* **3**:369 (1991).
11. A. Hasegawa and Y. Kodama, "Guiding-Center Soliton in Optical Fibers," *Opt. Lett.* **15**:1443 (1990).
12. G. P. Gordon, "Dispersive Perturbations of Solitons of the Nonlinear Schroedinger Equation," *JOSA B* **9**:91–97 (1992).
13. A. Hasegawa and Y. Kodama, "Signal Transmission by Optical Solitons in Monomode Fiber," *Proc. IEEE* **69**:1145 (1981).
14. K. J. Blow and N. J. Doran, "Solitons in Optical Communications," *IEEE J. Quantum Electron.*, **QE-19**:1883 (1982).
15. P. B. Hansen, H. A. Haus, T. C. Damen, J. Shah, P. V. Mamyshev, and R. H. Stolen, "Application of Soliton Spreading in Optical Transmission," Dig. ECOC, Vol. 3, Paper WeC3.4, pp. 3.109–3.112, Oslo, Norway, September 1996.
16. K. Tajima, "Compensation of Soliton Broadening in Nonlinear Optical Fibers with Loss," *Opt. Lett.* **12**:54 (1987).
17. H. H. Kuehl, "Solitons on an Axially Nonuniform Optical Fiber," *J. Opt. Soc. Am. B* **5**:709–713 (1988).
18. E. M. Dianov, L. M. Ivanov, P. V. Mamyshev, and A. M. Prokhorov, "High-Quality Femtosecond Fundamental Soliton Compression in Optical Fibers with Varying Dispersion," *Topical Meeting on Nonlinear Guided-Wave Phenomena: Physics and Applications*, 1989, Technical Digest Series, vol. 2, OSA, Washington, D.C., 1989, pp. 157–160, paper FA-5.
19. P. V. Mamyshev, "Generation and Compression of Femtosecond Solitons in Optical Fibers," *Bull. Acad. Sci. USSR, Phys. Ser.*, **55**(2):374–381 (1991) [*Izv. Acad. Nauk, Ser. Phys.* **55**(2):374–381 (1991)].
20. S. V. Chernikov and P. V. Mamyshev, "Femtosecond Soliton Propagation in Fibers with Slowly Decreasing Dispersion," *J. Opt. Soc. Am. B* **8**(8):1633–1641 (1991).
21. P. V. Mamyshev, S. V. Chernikov, and E. M. Dianov, "Generation of Fundamental Soliton Trains for High-Bit-Rate Optical Fiber Communication Lines," *IEEE J. of Quantum Electron.* **27**(10):2347–2355 (1991).
22. V. A. Bogatyrev, M. M. Bubnov, E. M. Dianov, et al., "Single-Mode Fiber with Chromatic Dispersion Varying along the Length," *IEEE J. of Lightwave Technology* **LT-9**(5):561–566 (1991).
23. V. I. Karpman and V. V. Solov'ev, "A Perturbation Approach to the Two-Soliton System," *Physica D* **3**:487–502 (1981).
24. J. P. Gordon, "Interaction Forces among Solitons in Optical Fibers," *Opt. Lett.* **8**:596–598 (1983).
25. J. P. Gordon and L. F. Mollenauer, "Effects of Fiber Nonlinearities and Amplifier Spacing on Ultra Long Distance Transmission," *J. Lightwave Technol.* **9**:170 (1991).
26. J. P. Gordon and H. A. Haus, "Random Walk of Coherently Amplified Solitons in Optical Fiber," *Opt. Lett.* **11**:665 (1986).
27. K. Smith and L. F. Mollenauer, "Experimental Observation of Soliton Interaction over Long Fiber Paths: Discovery of a Long-Range Interaction," *Opt. Lett.* **14**:1284 (1989).
28. E. M. Dianov, A. V. Luchnikov, A. N. Pilipetskii, and A. N. Starodumov, "Electrostriction Mechanism of Soliton Interaction in Optical Fibers," *Opt. Lett.* **15**:314 (1990).
29. E. M. Dianov, A. V. Luchnikov, A. N. Pilipetskii, and A. M. Prokhorov, "Long-Range Interaction of Solitons in Ultra-Long Communication Systems," *Soviet Lightwave Communications* **1**:235 (1991).
30. E. M. Dianov, A. V. Luchnikov, A. N. Pilipetskii, and A. M. Prokhorov, "Long-Range Interaction of Picosecond Solitons Through Excitation of Acoustic Waves in Optical Fibers," *Appl. Phys. B* **54**:175 (1992).

31. A. Mecozzi, J. D. Moores, H. A. Haus, and Y. Lai, "Soliton Transmission Control," *Opt. Lett.* **16**:1841 (1991).
32. Y. Kodama and A. Hasegawa, "Generation of Asymptotically Stable Optical Solitons and Suppression of the Gordon-Haus Effect," *Opt. Lett.* **17**:31 (1992).
33. M. Nakazawa, E. Yamada, H. Kubota, and K. Suzuki, "10 Gbit/s Soliton Transmission over One Million Kilometers," *Electron. Lett.* **27**:1270 (1991).
34. T. Widdowson and A. D. Ellis, "20 Gbit/s Soliton Transmission over 125 Mm," *Electron. Lett.* **30**:1866 (1994).
35. L. F. Mollenauer, J. P. Gordon, and S. G. Evangelides, "The Sliding-Frequency Guiding Filter: An Improved Form of Soliton Jitter Control," *Opt. Lett.* **17**:1575 (1992).
36. P. V. Mamyshev and L. F. Mollenauer, "Stability of Soliton Propagation with Sliding Frequency Guiding Filters," *Opt. Lett.* **19**:2083 (1994).
37. L. F. Mollenauer, P. V. Mamyshev, and M. J. Neubelt, "Measurement of Timing Jitter in Soliton Transmission at 10 Gbits/s and Achievement of 375 Gbits/s-Mm, Error-Free, at 12.5 and 15 Gbits/s," *Opt. Lett.* **19**:704 (1994).
38. D. LeGuen, F. Fave, R. Boittin, J. Debeau, F. Devaux, M. Henry, C. Thebault, and T. Georges, "Demonstration of Sliding-Filter-Controlled Soliton Transmission at 20 Gbit/s over 14 Mm," *Electron. Lett.* **31**:301 (1995).
39. P. V. Mamyshev and L. F. Mollenauer, "NRZ-to-Soliton Data Conversion by a Filtered Transmission Line," in *Optical Fiber Communication Conference OFC-95*, Vol. 8, 1995 OSA Technical Digest Series, OSA, Washington, D.C., 1995, Paper FB2, pp. 302–303.
40. P. V. Mamyshev and L. F. Mollenauer, "WDM Channel Energy Self-Equalization in a Soliton Transmission Line Using Guiding Filters," *Opt. Lett.* **21**(20):1658–1660 (1996).
41. L. F. Mollenauer, S. G. Evangelides, and J. P. Gordon, "Wavelength Division Multiplexing with Solitons in Ultra Long Distance Transmission Using Lumped Amplifiers," *J. Lightwave Technol.* **9**:362 (1991).
42. P. A. Andrekson, N. A. Olsson, J. R. Simpson, T. Tanbun-ek, R. A. Logan, P. C. Becker, and K. W. Wecht, *Electron. Lett.* **26**:1499 (1990).
43. P. V. Mamyshev, and L. F. Mollenauer, "Pseudo-Phase-Matched Four-Wave Mixing in Soliton WDM Transmission," *Opt. Lett.* **21**:396 (1996).
44. L. F. Mollenauer, P. V. Mamyshev, and M. J. Neubelt, "Demonstration of Soliton WDM Transmission at 6 and 7 $\times$ 10 GBit/s, Error-Free over Transoceanic Distances," *Electron. Lett.* **32**:471 (1996).
45. L. F. Mollenauer, P. V. Mamyshev, and M. J. Neubelt, "Demonstration of Soliton WDM Transmission at up to 8 $\times$ 10 GBit/s, Error-Free over Transoceanic Distances," OFC-96, Postdeadline paper PD-22.
46. M. Suzuki, I. Morita, N. Edagawa, S. Yamamoto, H. Taga, and S. Akiba, "Reduction of Gordon-Haus Timing Jitter by Periodic Dispersion Compensation in Soliton Transmission," *Electron. Lett.* **31**:2027–2029 (1995).
47. N. J. Smith, N. J. Doran, F. M. Knox, and W. Forysiak, "Energy-Scaling Characteristics of Solitons in Strongly Dispersion-Managed Fibers," *Opt. Lett.* **21**:1981–1983 (1996).
48. I. Gabitov and S. K. Turitsyn, "Averaged Pulse Dynamics in a Cascaded Transmission System with Passive Dispersion Compensation," *Opt. Lett.* **21**:327–329 (1996).
49. N. J. Smith, W. Forysiak, and N. J. Doran, "Reduced Gordon-Haus Jitter Due to Enhanced Power Solitons in Strongly Dispersion Managed Systems," *Electron. Lett.* **32**:2085–2086 (1996).
50. V. S. Grigoryan, T. Yu, E. A. Golovchenko, C. R. Menyuk, and A. N. Pilipetskii, "Dispersion-Managed Soliton Dynamics," *Opt. Lett.* **21**:1609–1611 (1996).
51. G. Carter, J. M. Jacob, C. R. Menyuk, E. A. Golovchenko, and A. N. Pilipetskii, "Timing Jitter Reduction for a Dispersion-Managed Soliton System: Experimental Evidence," *Opt. Lett.* **22**:513–515 (1997).
52. S. K. Turitsyn, V. K. Mezentsev and E. G. Shapiro, "Dispersion-Managed Solitons and Optimization of the Dispersion Management," *Opt. Fiber Tech.* **4**:384–452 (1998).
53. J. P. Gordon and L. F. Mollenauer, "Scheme for Characterization of Dispersion-Managed Solitons," *Opt. Lett.* **24**:223–225 (1999).
54. P. V. Mamyshev and N. A. Mamysheva, "Pulse-Overlapped Dispersion-Managed Data Transmission and Intra-Channel Four-Wave Mixing," *Opt. Lett.* **24**:1454–1456 (1999).
55. D. Le Guen, S. Del Burgo, M. L. Moulinaud, D. Grot, M. Henry, F. Favre, and T. Georges, "Narrow Band 1.02 Tbit/s (51 $\times$ 20 gbit/s) Soliton DWDM Transmission over 1000 km of Standard Fiber with 100 km Amplifier Spans," OFC-99, postdeadline paper PD-4.
56. S. Wabnitz, *Opt. Lett.* **21**:638–640 (1996).
57. E. A. Golovchenko, A. N. Pilipetskii, and C. R. Menyuk, *Opt. Lett.* **22**:1156–1158 (1997).

- 58. A. M. Niculae, W. Forysiak, A. G. Gloag, J. H. B. Nijhof, and N. J. Doran, "Soliton Collisions with Wavelength-Division Multiplexed Systems with Strong Dispersion Management," *Opt. Lett.* **23**:1354–1356 (1998).
- 59. P. V. Mamyshev and L. F. Mollenauer, "Soliton Collisions in Wavelength-Division-Multiplexed Dispersion-Managed Systems," *Opt. Lett.* **24**:448–450 (1999).
- 60. L. F. Mollenauer, P. V. Mamyshev, and J. P. Gordon, "Effect of Guiding Filters on the Behavior of Dispersion-Managed Solitons," *Opt. Lett.* **24**:220–222 (1999).
- 61. M. Matsumoto, *Opt. Lett.* **23**:1901–1903 (1998).
- 62. M. Matsumoto, *Electron. Lett.* **33**:1718 (1997).
- 63. L. F. Mollenauer, P. V. Mamyshev, J. Gripp, M. J. Neubelt, N. Mamysheva, L. Gruner-Nielsen, and T. Veng, "Demonstration of Massive WDM over Transoceanic Distances Using Dispersion Managed Solitons," *Opt. Lett.* **25**:704–706 (2000).

FIBER-OPTIC COMMUNICATION STANDARDS

Casimer DeCusatis

*IBM Corporation
Poughkeepsie, New York*

23.1 INTRODUCTION

This chapter presents a brief overview of several major industry standards for optical communications, including the following:

- ESCON/SBCON (Enterprise System Connection/Serial Byte Connection)
- FDDI (Fiber Distributed Data Interface)
- Fibre Channel Standard
- ATM (Asynchronous Transfer Mode)/SONET (Synchronous Optical Network)
- Ethernet (including Gigabit, 10 Gigabit, and other variants)
- InfiniBand

23.2 ESCON

The Enterprise System Connection (ESCON)* architecture was introduced on the IBM System/390 family of mainframe computers in 1990 as an alternative high-speed I/O channel attachment.^{1,2} The ESCON interface specifications were adopted in 1996 by the ANSI X3T1 committee as the serial byte connection (SBCON) standard.³

The ESCON/SBCON channel is a bidirectional, point-to-point 1300-nm fiber-optic data link with a maximum data rate of 17 Mbytes/s (200 Mbit/s). ESCON supports a maximum unrepeat distance of 3 km using 62.5 μm multimode fiber and LED transmitters with an 8-dB link budget, or a maximum unrepeat distance of 20 km using single-mode fiber and laser transmitters with a 14-dB link budget. The laser channels are also known as the ESCON extended distance feature (XDF). Physical connection is provided by an ESCON duplex connector, illustrated in Fig. 1. Recently, the single-mode

*ESCON is a registered trademark of IBM Corporation, 1991.

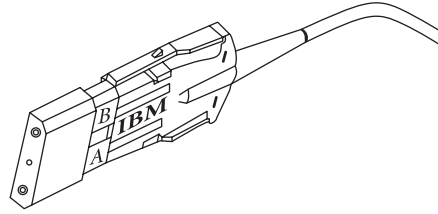


FIGURE 1 ESCON duplex fiber-optic connector.

ESCON links have adopted the SC duplex connector as standardized by Fibre Channel. With the use of repeaters or switches, an ESCON link can be extended up to 3 to 5 times these distances, and using wavelength division multiplexing (WDM) they can be extended even further, up to 100 km or more. However, performance of the attached devices in channel-to-channel applications typically falls off quickly at longer distances due to the longer round-trip latency of the link, making this approach suitable only for applications that can tolerate a lower effective throughput, such as remote backup of data for disaster recovery. There are some applications which can run over an ESCON physical layer without experiencing this performance degradation, such as virtual tape servers (VTS). ESCON devices and CPUs may communicate directly through a channel-to-channel attachment, but more commonly attach to a central nonblocking dynamic crosspoint switch. The resulting network topology is similar to a star-wired ring, which provides both efficient bandwidth utilization and reduced cabling requirements. The switching function is provided by an ESCON director, a nonblocking circuit switch. Although ESCON uses 8B/10B encoded data, it is not a packet-switching network; instead, the data frame header includes a request for connection that is established by the director for the duration of the data transfer. An ESCON data frame includes a header, payload of up to 1028 bytes of data, and a trailer. The header consists of a 2-character start-of-frame delimiter, 2-byte destination address, 2-byte source address, and 1 byte of link control information. The trailer is a 2-byte cyclic redundancy check (CRC) for errors, and a three-character end-of-frame delimiter. ESCON uses a DC-balanced 8B/10B coding scheme developed by IBM.

23.3 FDDI

The fiber distributed data interface (FDDI) was among the first open networking standards to specify optical fiber. It was an outgrowth of the ANSI X3T9.5 committee proposal in 1982 for a high-speed token passing ring as a back-end interface for storage devices. While interest in this application waned, FDDI found new applications as the backbone for local area networks (LANs). The FDDI standard was approved in 1992 as ISO standards IS 9314/1-2 and DIS 9314-3; it follows the architectural concepts of IEEE standard 802 (although it is controlled by ANSI, not IEEE, and therefore has a different numbering sequence) and is among the family of standards (including token ring and ethernet) that are compatible with a common IEEE 802.2 interface. FDDI is a family of four specifications, namely, the physical layer (PHY), physical media dependent (PMD), media access control (MAC), and station management (SMT). These four specifications correspond to sublayers of the data link and physical layer of the OSI reference model; as before, we will concentrate on the physical layer implementation.

The FDDI network is a 100-Mbit/s token passing ring, with dual counterrotating rings for fault tolerance. The dual rings are independent fiber-optic cables; the primary ring is used for data transmission, and the secondary ring is a backup in case a node or link on the primary ring fails. Bypass switches are also supported to reroute traffic around a damaged area of the network and prevent the ring from fragmenting in case of multiple node failures. The actual data rate is 125 Mbit/s, but this is reduced to an effective data rate of 100 Mbit/s by using a 4B/5B coding scheme. This high speed

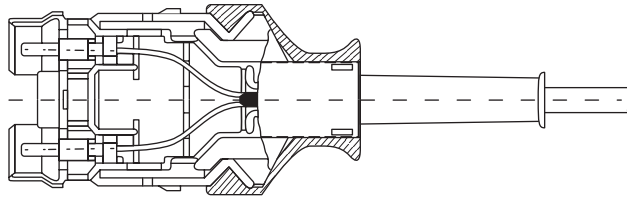


FIGURE 2 FDDI duplex fiber-optic connector.

allows FDDI to be used as a backbone to encapsulate lower speed 4, 10, and 16 Mbit/s LAN protocols; existing ethernet, token ring, or other LANs can be linked to an FDDI network via a bridge or router. Although FDDI data flows in a logical ring, a more typical physical layout is a star configuration with all nodes connected to a central hub or concentrator rather than to the backbone itself. There are two types of FDDI nodes, either *dual attach* (connected to both rings) or *single attach*; a network supports up to 500 dual-attached nodes, 1000 single-attached nodes, or an equivalent mix of the two types. FDDI specifies 1300-nm LED transmitters operating over 62.5 μm multimode fiber as the reference media, although the standard also provides for the attachment of 50, 100, 140, and 185 μm fiber. Using 62.5 μm fiber, a maximum distance of 2 km between nodes is supported with an 11-dB link budget; since each node acts like a repeater with its own phase-lock loop to prevent jitter accumulation, the entire FDDI ring can be as large as 100 km. However, an FDDI link can fail due to either excessive attenuation or dispersion; for example, insertion of a bypass switch increases the link length and may cause dispersion errors even if the loss budget is within specifications. For most other applications, this does not occur because the dispersion penalty is included in the link budget calculations or the receiver sensitivity measurements. The physical interface is provided by a special media interface connector (MIC), illustrated in Fig. 2. The connector has a set of three color-coded keys which are interchangeable depending on the type of network connection;¹ this is intended to prevent installation errors and assist in cable management.

An FDDI data frame is variable in length and contains up to 4500 8-bit bytes, or octets, including a preamble, start of frame, frame control, destination address, data payload, CRC error check, and frame status/end of frame. Each node has an MAC sublayer that reviews all the data frames looking for its own destination address. When it finds a packet destined for its node, that frame is copied into local memory; a copy bit is turned on in the packet; and it is then sent on to the next node on the ring. When the packet returns to the station that originally sent it, the originator assumes that the packet was received if the copy bit is on; the originator will then delete the packet from the ring. As in the IEEE 802.5 token ring protocol, a special type of packet called a *token* circulates in one direction around the ring, and a node can only transmit data when it holds the token. Each node observes a token retention time limit, and also keeps track of the elapsed time since it last received the token; nodes may be given the token in equal turns, or they can be given priority by receiving it more often or holding it longer after they receive it. This allows devices having different data requirements to be served appropriately.

Because of the flexibility built into the FDDI standard, many changes to the base standard have been proposed to allow interoperability with other standards, reduce costs, or extend FDDI into the MAN or WAN. These include a single-mode PMD layer for channel extensions up to 20 to 50 km. An alternative PMD provides for FDDI transmission over copper wire, either shielded or unshielded twisted pairs; this is known as *copper distributed data interface*, or CDDI. A new PMD was also developed to adapt FDDI data packets for transfer over a SONET link by stuffing approximately 30 Mbit/s into each frame to make up for the data rate mismatch (we will discuss SONET as an ATM physical layer in a later section). An enhancement called FDDI-II uses time-division multiplexing to divide the bandwidth between voice and data; it accommodates isochronous, circuit-switched traffic as well as existing packet traffic. An option known as *low cost* (LC) FDDI uses the more common SC duplex connector instead of the more expensive MIC connectors, and a lower-cost transceiver with a 9-pin footprint similar to the single-mode ESCON parts.

23.4 FIBRE CHANNEL STANDARD

Development of the ANSI Fibre Channel (FC) Standard began in 1988 under the X3T9.3 Working Group, as an outgrowth of the Intelligent Physical Protocol Enhanced Physical Project. The motivation for this work was to develop a scaleable standard for the attachment of both networking and I/O devices, using the same drivers, ports, and adapters over a single channel at the highest speeds currently achievable. The standard applies to both copper and fiber-optic media, and uses the English spelling *fibre* to denote both types of physical layers. In an effort to simplify equipment design, FC provides the means for a large number of existing upper-level protocols (ULPs), such as IP, SCSI, and HIPPI, to operate over a variety of physical media. Different ULPs are mapped to FC constructs, encapsulated in FC frames, and transported across a network; this process remains transparent to the attached devices. The standard consists of five hierarchical layers,⁴ namely a physical layer, an encode/decode layer which has adopted the DC-balanced 8B/10B code, a framing protocol layer, a common services layer (at this time, no functions have been formally defined for this layer), and a protocol-mapping layer to encapsulate ULPs into FC. Physical layer specifications for 1, 2, 4, and 8 Gbit/s links have been defined (refer to the ANSI standard for the most recent specifications). If the two link endpoints have different data rate capabilities, the links will auto-negotiate to the highest available rate between either 1, 2, and 4 Gbit/s rates or between 2, 4, and 8 Gbit/s rates. Note that the 10 Gbit/s data rate specifies a 64B/66B encoding scheme, rather than 8B/10B, and consequently is not backward compatible with lower data rates; this rate is typically reserved for inter-switch links (ISLs). The second layer defines the Fibre Channel data frame; frame size depends upon the implementation and is variable up to 2148 bytes long. Each frame consists of a 4-byte start-of-frame delimiter, a 24-byte header, a 2112-byte payload containing from 0 to 64 bytes of optional headers and 0 to 2048 bytes of data, a 4-byte CRC, and a 4-byte end-of-frame delimiter. In October 1994, the Fibre Channel physical and signaling interface standard FC-PH was approved as ANSI standard X3.230-1994.

Logically, Fibre Channel is a bidirectional point-to-point serial data link. Physically, there are many different media options (see Table 1) and three basic network topologies. The simplest, *default topology*, is a point-to-point direct link between two devices, such as a CPU and a device controller. The second, *Fibre Channel Arbitrated Loop* (FC-AL), connects between 2 and 126 devices in a loop configuration. Hubs or switches are not required, and there is no dedicated loop controller; all nodes on the loop share the bandwidth and arbitrate for temporary control of the loop at any given time. Each node has equal opportunity to gain control of the loop and establish a communications path; once the node relinquishes control, a fairness algorithm ensures that the same node cannot win control of the loop again until all other nodes have had a turn. As networks become larger, they may grow into the third topology, an *interconnected switchable network* or *fabric*, in which all network management functions are taken over by a switching point, rather than each node. An analogy for a switched fabric is the telephone network; users specify an address (phone number) for a device with which they want to communicate, and the network provides them with an interconnection path. In theory there is no limit to the number of nodes in a fabric; practically, there are only about 16 million unique addresses. Fibre Channel also defines three classes of connection service, which offer options such as guaranteed delivery of messages in the order they were sent and acknowledgment of received messages.

As shown in Table 1, FC provides for both single-mode and multimode fiber-optic data links using longwave (1300-nm) lasers and LEDs as well as short-wave (780 to 850 nm) lasers. The physical connection is provided by an SC duplex connector defined in the standard (see Fig. 3), which is keyed to prevent misplugging of a multimode cable into a single-mode receptacle. This connector design has since been adopted by other standards, including ATM, low-cost FDDI, and single-mode ESCON. The requirement for international class 1 laser safety is addressed using open fiber control (OFC) on some types of multimode links with shortwave lasers. This technique automatically senses when a full duplex link is interrupted, and turns off the laser transmitters on both ends to preserve laser safety. The lasers then transmit low-duty cycle optical pulses until the link is reestablished; a handshake sequence then automatically reactivates the transmitters.

TABLE 1 Examples of the Fiber Channel Standard Physical Layer

Media Type	Data Rate (Mbytes/s)	Maximum Distance	Signaling Rate (Mbaud)	Transmitter
SMF	800	10 km	8500.0	LW laser
	400	10 or 4 km	4250.0	LW laser
	200	10 km	2125.0	LW laser
	100	10 km	1062.5	LW laser
	50	10 km	1062.5	LW laser
	25	10 km	1062.5	LW laser
50- μ m multimode fiber	800	10 km	8500.0	SW laser
	400	10 or 4 km	4250.0	SW laser
	200	10 km	2125.0	SW laser
	100	500 m	1062.5	SW laser
	50	1 km	531.25	SW laser
	25	2 km	265.625	SW laser
62.5- μ m multimode fiber	12.5	10 km	132.8125	LW LED
	100	300 m	1062.5	SW laser
	50	600 m	531.25	SW laser
	25	1 km	265.625	LW LED
105- Ω type 1 shielded twisted pair electrical	12.5	2 km	132.8125	LW LED
	25	50 m	265.125	ECL
75 Ω mini coax	12.5	100 m	132.8125	ECL
	100	10 m	1062.5	ECL
	50	20 m	531.25	ECL
	25	30 m	265.625	ECL
75 Ω video coax	12.5	40 m	132.8125	ECL
	100	25 m	1062.5	ECL
	50	50 m	531.25	ECL
	25	75 m	265.625	ECL
150 Ω twinax or STP	12.5	100 m	132.8125	ECL
	100	30 m	1062.5	ECL
	50	60 m	531.25	ECL
	25	100 m	265.625	ECL

LW = long wavelength, SW = short wavelength, ECL = emitter-coupled logic.

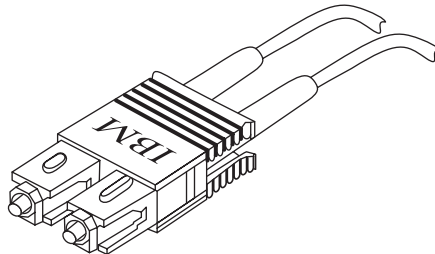


FIGURE 3 Single-mode SC duplex fiber-optic connector, per ANSI FC Standard specifications, with one narrow key and one wide key. Multimode SC duplex connectors use two wide keys.

23.5 ATM/SONET

Developed by the ATM Forum, this protocol promised to provide a common transport media for voice, data, video, and other types of multimedia. ATM is a high-level protocol that can run over many different physical layers including copper; part of ATM's promise to merge voice and data traffic on a single network comes from plans to run ATM over the synchronous optical network (SONET) transmission hierarchy developed for the telecommunications industry. SONET is really a family of standards defined by ANSI T1.105-1988 and T1.106-1988, as well as by several CCITT recommendations.⁵⁻⁸ Several different data rates are defined as multiples of 51.84 Mbit/s, known as OC-1. The numerical part of the OC-level designation indicates a multiple of this fundamental data rate, thus 155 Mbit/s is called OC-3. The standard provides for incremental data rates including OC-3, OC-9, OC-12, OC-18, OC-24, OC-36, and OC-48 (2.48832 Gbit/s). Both single-mode links with laser sources and multimode links with LED sources are defined for OC-1 through OC-12; only single-mode laser links are defined for OC-18 and beyond. SONET also contains provisions to carry sub-OC-1 data rates, called *virtual tributaries*, which support telecom data rates including DS-1 (1.544 Mbit/s), DS-2 (6.312 Mbit/s), and 3.152 Mbit/s (DS1C). The basic SONET data frame is an array of nine rows with 90 bytes per row, known as a synchronous-transport signal level 1 (STS-1) frame. In an OC-1 system, an STS-1 frame is transmitted once every 125 μ s (810 bytes per 125 μ s yields 51.84 Mbit/s). The first three columns provide overhead functions such as identification, framing, error checking, and a pointer which identifies the start of the 87-byte data payload. The payload floats in the STS-1 frame, and may be split across two consecutive frames. Higher speeds can be obtained either by concatenation of N frames into an STS- N c frame (the "c" stands for *concatenated*) or by byte-interleaved multiplexing of N frames into an STS- N frame.

ATM technology incorporates elements of both circuit and packet switching. All data is broken down into a 53-byte cell, which may be viewed as a short fixed-length packet. Five bytes make up the header, providing a 48-byte payload. The header information contains routing information (cell addresses) in the form of virtual path and channel identifiers; a field to identify the payload type; an error check on the header information; and other flow control information. Cells are generated asynchronously; as the data source provides enough information to fill a cell, it is placed in the next available cell slot. There is no fixed relationship between the cells and a master clock, as in conventional time-division multiplexing schemes; the flow of cells is driven by the bandwidth needs of the source. ATM provides bandwidth on demand; for example, in a client-server application the data may come in bursts; several data sources could share a common link by multiplexing during the idle intervals. Thus, the ATM adaptation layer allows for both constant and variable bit rate services. The combination of transmission options is sometimes described as a *pleiosynchronous network*, meaning that it combines some features of multiplexing operations without requiring a fully synchronous implementation. Note that the fixed cell length allows the use of synchronous multiplexing and switching techniques, while the generation of cells on demand allows flexible use of the link bandwidth for different types of data, characteristic of packet switching. Higher-level protocols may be required in an ATM network to ensure that multiplexed cells arrive in the correct order, or to check the data payload for errors (given the typical high reliability and low BER of modern fiber-optic technology, it was considered unnecessary overhead to replicate data error checks at each node of an ATM network). If an intermediate node in an ATM network detects an error in the cell header, cells may be discarded without notification to either end user. Although cell loss priority may be defined in the ATM header, for some applications the adoption of unacknowledged transmission may be a concern.

ATM data rates were intended to match SONET rates of 51, 155, and 622 Mbit/s; an FDDI-compliant data rate of 100 Mbit/s was added, in order to facilitate emulation of different types of LAN traffic over ATM. In order to provide a low-cost copper option and compatibility with 16-Mbit/s token ring LANs to the desktop, a 25-Mbit/s speed has also been approved. For premises wiring applications, ATM specifies the SC duplex connector, color coded beige for multimode links and blue for single-mode links. At 155 Mbit/s, multimode ATM links support a maximum distance of 3 km while single-mode links support up to 20 km.

23.6 ETHERNET

Ethernet was originally a local area network (LAN) communication standard developed for copper interconnections on a common data bus; it is an IEEE standard 802.3.⁹ The basic principle used in Ethernet is carrier sense multiple access with collision detection (CSMA/CD). Ethernet LANs may be configured as a bus, often wired radially through a central hub. A device attached to the LAN that intends to transmit data must first sense whether another device is transmitting. If another device is already sending, then it must wait until the LAN is available; thus, the intention is that only one device will be using the LAN to send data at a given time. When one device is sending, all other attached devices receive the data and check to see if it is addressed to them; if it is not, then the data is discarded. If two devices attempt to send data at the same time (e.g., both devices may begin transmission at the same time after determining that the LAN is available; there is a gap between when one device starts to send and before another potential sender can detect that the LAN is in use), then a collision occurs. Using CSMA/CD as the media access control protocol, when a collision is detected attached devices will detect the collision and must wait for different lengths of time before attempting retransmission. Since it is not always certain that data will reach its destination without errors or that the sending device will know about lost data, each station on the LAN must operate an end-to-end protocol for error recovery and data integrity. Data frames begin with an 8-byte preamble used for determining start-of-frame and synchronization, and a header consisting of a 6-byte destination address, 6-byte source address, and 2-byte length field. User data may vary from 46 to 1500 bytes, with data shorter than the minimum length padded to fit the frame; the user data is followed by a 2-byte CRC error check. Thus, an Ethernet frame may range from 70 to 1524 bytes.

The original Ethernet standard, known also as 10Base-T (10 Mbit/s over unshielded twisted pair copper wires) was primarily a copper standard, although a specification using 850-nm LEDs was also available. Subsequent standardization efforts increased this data rate to 100 Mbit/s over the same copper media (100Base-T), while once again offering an alternative fiber specification (100Base-FX). Recently, the standard has continued to evolve with the development of Gigabit Ethernet (1000Base-FX), which operates over fiber as the primary medium. Standardized as IEEE 802.3z, Gigabit Ethernet includes changes to the MAC layer in addition to a completely new physical layer operating at 1.25 Gbit/s. Switches rather than hubs predominate, since at higher data rates throughput per end user and total network cost are both optimized by using switched rather than shared media. The minimum frame size has increased to 512 bytes; frames shorter than this are padded with idle characters (carrier extension). The maximum frame size remains unchanged, although devices may now transmit multiple frames in bursts rather than single frames for improved efficiency. The physical layer will use standard 8B/10B data encoding. The standard allows several different physical connector types for fiber, including the SC duplex and various small-form-factor connectors about the size of a standard RJ-45 jack, although the LC duplex has become the most commonly used variant. Early transceivers were packaged as gigabit interface converters, or GBICs, which allows different optical or copper transceivers to be plugged onto the same host card. This has been replaced by small form factor pluggable transceivers (either SFP or SFP+). Some variants of the standard allow operation of long-wave (1300 nm) laser sources over both single-mode and multimode fiber. When a transmitter is optimized for a single-mode launch condition, it will underfill the multimode fiber; this causes some modes to be excited and propagate at different speeds than others, and the resulting differential mode delay significantly degrades link performance. One solution involves the use of special optical cables known as *optical mode conditioners* with offset ferrules to simulate an equilibrium mode launch condition into multimode fiber.

Ethernet continues to evolve as one of the predominant protocols for data center networking. Standards are currently being defined for both 40 Gbit/s and 100 Gbit/s Ethernet links. Other emerging standards allow transport of Fibre Channel over Ethernet, in an effort to converge two common types of optical networks. In an effort to make the Ethernet links more robust, other new standards known as Converged Enhanced Ethernet (CEE) are under development. The CEE standards add features such as new types of fine grained flow control, enhanced quality of service, and lossless transmission.

TABLE 2 Examples of the InfiniBand Physical Layer

Media Type	Per Lane Data Rate (Gbits/s)	Number of Lanes	Unidirectional Signaling Rate (Mbytes)	Transmitter
50- μ m multimode fiber	2.5	1X	250	VCSEL
		4X	10,000	VCSEL
		8X	20,000	VCSEL
		12X	30,000	VCSEL
50- μ m multimode fiber	5.0	1X	500	VCSEL
		4X	20,000	VCSEL
		8X	40,000	VCSEL
		12X	60,000	VCSEL
50- μ m multimode fiber	10.0	1X	10,000	VCSEL
SMF	10.0	1X	10,000	LW laser

LW = long wavelength, VCSEL = vertical cavity surface emitting laser.

23.7 INFINIBAND

The InfiniBand standards was developed by the InfiniBand Trade Association (IBTA) in an attempt to converge multiple protocol networks. Currently, it is most widely used for low-latency, high-performance applications in data communication. InfiniBand specifies 8B/10B encoded data, and both serial and parallel optical links, some of which are illustrated in Table 2.¹⁰ These are referred to by the number of lanes in the physical layer interface; for example, a 4X link employs four optical fibers in each direction of a bidirectional link. The industry standard multifiber push-on (MPO) connector is specified for parallel optical links, while the SC duplex connector is commonly used for single fiber links (note that the 10 Gbit/s serial IB link physical layer is very similar to the 10-Gbit Ethernet link specification). InfiniBand is a switched point-to-point protocol, although some data communication applications employ the InfiniBand physical layer only, and are therefore not compatible with InfiniBand switches (e.g., the Parallel Sysplex IB links developed for IBM mainframes). Although InfiniBand links are not designed for operation at distances beyond 10 km, with sufficiently large receive buffers or other flow control management techniques they can be extended to much longer distances. This can be done using protocol independent wavelength multiplexing, or by encapsulating InfiniBand into another protocol such as SONET.

23.8 REFERENCES

1. D. Stigliani, "Enterprise Systems Connection Fiber Optic Link," Chap. 13, in *Handbook of Optoelectronics for Fiber Optic Data Communications*, C. DeCusatis, R. Lasky, D. Clement, and E. Mass (eds.), Academic Press, San Diego, California (1997).
2. "ESCON I/O Interface Physical Layer Document" *IBM Document Number SA23-0394*, 3rd ed. IBM Corporation, Mechanicsburg, Pennsylvania (1995).
3. Draft ANSI Standard X3T11/95-469 (rev. 2.2) "ANSI Single Byte Command Code Sets Connection Architecture (SBCON)," ANSI, Washington, DC (1996).
4. ANSI X3.230-1994 (rev. 4.3), "Fibre Channel—Physical and Signaling Interface (FC-PH)," ANSI X3.272-199x (rev. 4.5), "Fibre Channel—Arbitrated Loop (FC-AL)," (June 1995); ANSI X3.269-199x, (rev. 012), "Fiber Channel Protocol for SCSI (FCP)," ANSI, Washington, DC (May 30, 1995).
5. ANSI T1.105-1988, "Digital Hierarchy Optical Rates and Format Specification," ANSI, Washington, DC (1988).
6. CCITT Recommendation G.707, "Synchronous Digital Hierarchy Bit Rates," CCITT (1991).

7. CCITT Recommendation G.708, "Network Node Interfaces for the Synchronous Digital Hierarchy," CCITT, Geneva, Switzerland (1991).
8. CCITT Recommendation G.709, "Synchronous Multiplexing Structure," CCITT, Geneva, Switzerland (1991).
9. IEEE 802.3z, "Draft Supplement to Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications: Media Access Control (MAC) Parameters, Physical Layer, Repeater and Management Parameters for 1000 Mb/s Operation," IEEE, Piscataway, New Jersey (June 1997).
10. A. Ghiasi, "InfiniBand," in *Handbook of Fiber Optic Data Communication*, 3rd ed., C. DeCusatis (ed.), Academic Press, San Diego (2008).

This page intentionally left blank

OPTICAL FIBER SENSORS

Richard O. Claus

*Virginia Tech
Blacksburg, Virginia*

Ignacio Matias and Francisco Arregui

*Public University Navarra
Pamplona, Spain*

24.1 INTRODUCTION

Optical fiber sensors are a broad topic. The objective of this chapter is to briefly summarize the fundamental properties of representative types of optical fiber sensors and how they operate. Four different types of sensors are evaluated systematically on the basis of performance criteria such as resolution, dynamic range, cross-sensitivity to multiple ambient perturbations, fabrication, and demodulation processes. The optical fiber sensing methods that will be investigated include well-established technologies such as fiber Bragg grating (FBG)–based sensors, and rapidly evolving measurement techniques such as those involving long-period gratings (LPGs). Additionally, two popular versions of Fabry-Perot interferometric sensors (intrinsic and extrinsic) are evaluated.

The outline of this chapter is as follows. The principles of operation and fabrication processes of each of the four sensors are discussed separately. The sensitivity of the sensors to displacement and simultaneous perturbations such as temperature is analyzed. The overall complexity and performance of a sensing technique depends heavily on the signal demodulation process. Thus, the detection schemes for all four sensors are discussed and compared on the basis of their complexity. Finally, a theoretical analysis of the cross-sensitivities of the four sensing schemes is presented and their performance is compared.

Measurements of a wide range of physical measurands by optical fiber sensors have been investigated for more than 20 years. Displacement measurements using optical fiber sensors are typical of these, and both embedded and surface-mounted configurations have been reported by researchers in the past.¹ Fiber-optic sensors are small in size, are immune to electromagnetic interference, and can be easily integrated with existing optical fiber communication links. Such sensors can typically be easily multiplexed, resulting in distributed networks that can be used for health monitoring of integrated, high-performance materials and structures.

Optical fiber sensors of displacement are perhaps the most basic of all fiber sensor types because they may be configured to measure many other related environmental factors. They should possess certain important characteristics. First, they should either be insensitive to ambient fluctuations in temperature and pressure or should employ demodulation techniques that compensate for changes in the output signal due to these additional perturbations. In an embedded configuration, the sensors for axial strain measurements should have minimum cross-sensitivity to other strain states. The sensor signal should itself be simple and easy to demodulate. Nonlinearities in the output require

expensive decoding procedures or necessitate precalibration and sensor-to-sensor incompatibility. The sensor should ideally provide an absolute and real-time displacement or strain measurement in a form that can be easily processed. For environments where large strain magnitudes are expected, the sensor should have a large dynamic range while at the same time maintaining the desired sensitivity. We now discuss each of the four sensing schemes individually and present their relative advantages and shortcomings.

24.2 EXTRINSIC FABRY-PEROT INTERFEROMETRIC SENSORS

The extrinsic Fabry-Perot interferometric (EFPI) sensor, proposed by a number of groups and authors, is one of the most popular fiber-optic sensors used for applications in health monitoring of smart materials and structures.^{2,3} As the name suggests, the EFPI is an interferometric sensor in which the detected intensity is modulated by the parameter under measurement. The simplest configuration of an EFPI is shown in Fig. 1.

The EFPI system consists of a single-mode laser diode that illuminates a Fabry-Perot cavity through a fused biconical tapered coupler. The cavity is formed between an input single-mode fiber and a reflecting target element that may be a fiber. Since the cavity is external to the lead-in/lead-out fiber, the EFPI sensor is independent of transverse strain and small ambient temperature fluctuations. The input fiber and the reflecting fiber are typically aligned using a hollow core tube as shown in Fig. 10. For optical fibers with uncoated ends, Fresnel reflection of approximately 4 percent results at the glass-to-air and air-to-glass interfaces that define the cavity. The first reflection at the glass-air interface R_1 , called the *reference reflection*, is independent of the applied perturbation. The second reflection at the air-glass interface R_2 , termed the *sensing reflection*, is dependent on the length of the cavity d , which in turn is modulated by the applied perturbation. These two reflections interfere

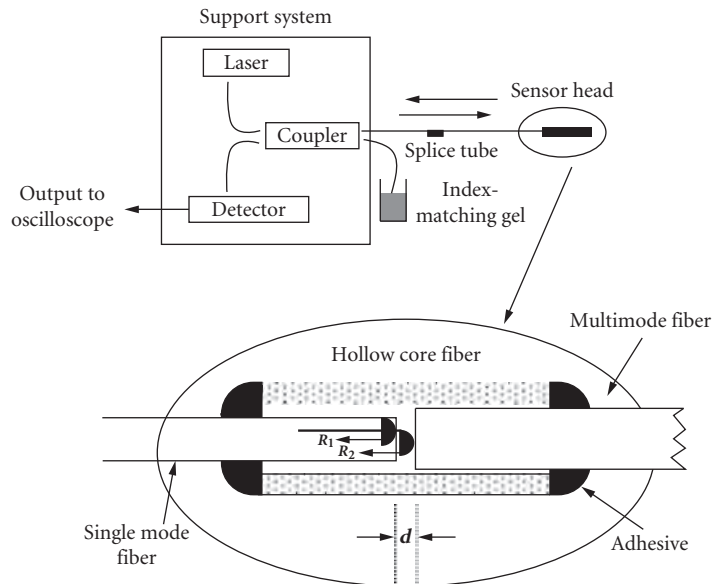


FIGURE 1 Extrinsic Fabry-Perot interferometric sensor and system.

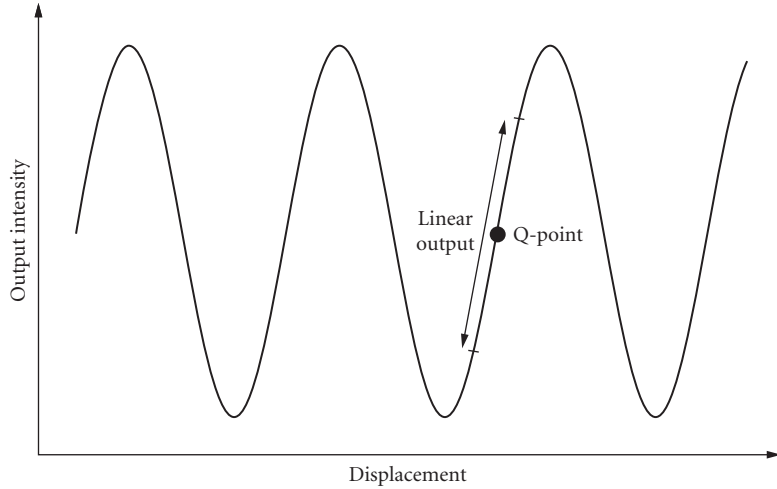


FIGURE 2 EFPI transfer function curve.

(provided $2d < L_c$, the coherence length of the light source) and the intensity I at the detector varies as a function of the cavity length,

$$I = I_0 \cos\left(\frac{4\pi}{\lambda} d\right) \quad (1)$$

where I_0 is the maximum value of the output intensity and λ is the center wavelength of the light source, here assumed to be a laser diode.

The typical intensity-versus-displacement transfer function curve [Eq. (1)] for an EFPI sensor is shown in Fig. 2. Small perturbations that result in operation around the quiescent or Q point of the sensor lead to an approximately linear variation in output intensity versus applied displacement. For larger displacements, the output signal is not a linear function of the input signal, and the output signal may vary over several sinusoidal periods. In this case, a *fringe* in the output signal is defined as the change in intensity from a maximum to a maximum, or from a minimum to a minimum, so each fringe corresponds to a change in the cavity length by half of the operating wavelength λ . The change in the cavity length Δd is then employed to calculate the strain using the expression

$$\varepsilon = \frac{\Delta d}{L} \quad (2)$$

where L is defined as the gauge length of the sensor and is typically the distance between two points where the input and reflecting fibers are bonded to the hollow-core support tube.

The EFPI sensor has been used for the analysis of materials and structures.^{1,3} The relatively low temperature sensitivity of the sensor element, due to the opposite directional expansion of the fiber and tube elements, makes it attractive for the measurement of strain and displacement in environments where the temperature is not anticipated to change over a wide range. The EFPI sensor is capable of measuring subangstrom displacements with strain resolution better than $1 \mu\epsilon$ and a dynamic range greater than $10,000 \mu\epsilon$. Moreover, the large bandwidth simplifies the measurement of highly cyclical strain. The sensor also allows single-ended operation and is hence suitable for applications where ingress to and egress from the sensor location are important. The sensor requires simple and inexpensive fabrication equipment and an assembly time of a few minutes. Additionally, since the cavity is external to the fibers, transverse strain components that tend to influence the response of similar intrinsic sensors through Poisson-effect cross-coupling have negligible effect on the EFPI sensor output.

24.3 INTRINSIC FABRY-PEROT INTERFEROMETRIC SENSORS

The intrinsic Fabry-Perot interferometric (IFPI) sensor is similar in operation to its extrinsic counterpart, but significant differences exist in the configurations of the two sensors.⁴ The basic IFPI sensor is shown in Fig. 3. An optically isolated laser diode is used as the optical source to one of the input arms of a bidirectional 2×2 coupler. The Fabry-Perot cavity is formed internally by fusing a small length of single-mode fiber to one of the output legs of the coupler. As shown in Fig. 3, the reference (R) and sensing (S) reflections interfere at the detector to again provide a sinusoidal intensity variation versus cavity path length modulation. The cavity can also be implemented by introducing two Fresnel or other reflectors along the length of a single fiber. The photosensitivity effect in germanosilicate fibers has been used in the past to fabricate broadband grating-based reflector elements to define such an IFPI cavity.⁵ Since the cavity is formed within an optical fiber, changes in the refractive index of the fiber due to the applied perturbation can significantly alter the phase of the sensing signal S . Thus the intrinsic cavity results in the sensor being sensitive to ambient temperature fluctuations and all states of strain.

The IFPI sensor, like all other interferometric signals, has a nonlinear output that complicates the measurement of large-magnitude strain. This can again be overcome by operating the sensor in the linear regime around the Q point of the sinusoidal transfer function curve. The main limitation of the IFPI strain sensor is that the photoelastic-effect-induced change in index of refraction results in a nonlinear relationship between the applied perturbation and the change in cavity length. For most IFPI sensors, the change in the propagation constant of the fundamental mode dominates the change in cavity length. Thus IFPIs are highly susceptible to temperature changes and transverse strain components.⁶ In embedded applications, the sensitivity to all of the strain components can result in complex signal output. The process of fabricating an IFPI strain sensor is more complicated than that for the EFPI sensor since the sensing cavity of the IFPI sensor must be formed within the optical fiber by some special procedure. The strain resolution of IFPI sensors is approximately $1 \mu\epsilon$ with an operating range greater than $10,000 \mu\epsilon$. IFPI sensors also suffer from drift in the output signal due to variations in the polarization state of the input light.

Thus the preliminary analysis shows that the extrinsic version of the Fabry-Perot optical fiber sensor seems to have an overall advantage over its intrinsic counterpart. The extrinsic sensor has negligible cross-sensitivity to temperature and transverse strain. Although the strain sensitivities, dynamic ranges, and bandwidths of the two sensors are comparable, the IFPIs can be expensive and cumbersome to fabricate due to the intrinsic nature of the sensing cavity.

The extrinsic and intrinsic Fabry-Perot interferometric sensors possess nonlinear sinusoidal outputs that complicate signal processing at the detector. Although intensity-based sensors have a simple output variation, they suffer from limited sensitivity to strain or other perturbations of interest. Grating-based sensors have recently become popular as transducers that provide wavelength-encoded output signals that can typically be easily demodulated to derive information about the perturbation under investigation. We next discuss the advantages and drawbacks of Bragg grating sensing technology. The basic operating mechanism of Bragg grating-based strain sensors is then reviewed

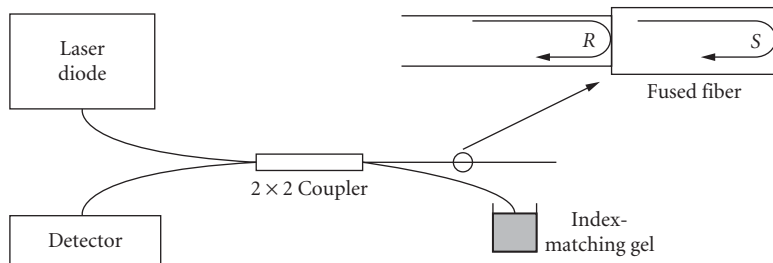


FIGURE 3 The intrinsic Fabry-Perot interferometric (IFPI) sensor.

and the expressions for strain resolution are obtained. These sensors are then compared to the recently developed long-period grating devices in terms of fabrication process, cross-sensitivity to multiple measurands, and simplicity of signal demodulation.

24.4 FIBER BRAGG GRATING SENSORS

The phenomenon of photosensitivity was discovered by Hill and coworkers in 1978.⁷ It was found that permanent refractive index changes could be induced in optical fibers by exposing the germanium-doped core of a fiber to intense light at 488 or 514 nm. Hill found that a sinusoidal modulation of index of refraction in the core created by the spatial variation of such an index-modifying beam gives rise to refractive index grating that can be used to couple the energy in the fundamental guided mode to various guided and lossy modes. Later Meltz et al.⁸ suggested that photosensitivity is more efficient if the fiber is side-exposed to a writing beam at wavelengths close to the absorption wavelength (242 nm) of the germanium defects in the fiber. The side-writing process simplified the fabrication of Bragg gratings, and these devices have recently emerged as highly versatile components for optical fiber communication and sensing systems. Recently, loading of the fibers with hydrogen prior to writing has been used to produce order-of-magnitude larger changes in index in germanosilicate fibers.⁹

Principle of Operation

Bragg gratings in optical fibers are based on a phase-matching condition between propagating optical modes. This phase-matching condition is given by

$$k_g + k_c = k_B \quad (3)$$

where k_g , k_c , and k_B are, respectively, the wave vectors of the coupled guided mode, the resulting coupling mode, and the grating. For a first-order interaction, $k_B = 2\pi/\Lambda$, where Λ is the spatial period of the grating. In terms of propagation constants, this condition reduces to the general form of interaction for mode coupling due to a periodic perturbation

$$\Delta\beta = \frac{2\pi}{\Lambda} \quad (4)$$

where $\Delta\beta$ is the difference in the propagation constants of the two modes involved in mode coupling, where both modes are assumed to travel in the same direction.

Fiber Bragg gratings (FBGs) involve the coupling of the forward-propagating fundamental LP_{01} in a single-mode fiber to the reverse-propagating LP_{01} mode.¹⁰ Here, consider a single-mode fiber with β_{01} and $-\beta_{01}$ as the propagation constants of the forward- and reverse-propagating fundamental LP_{01} modes. To satisfy the phase-matching condition,

$$\Delta\beta = \beta_{01} - (-\beta_{01}) = \frac{2\pi}{\Lambda} \quad (5)$$

where $\beta_{01} = 2\pi n_{\text{eff}}/\lambda$, n_{eff} is the effective index of the fundamental mode, and λ is the free-space wavelength of the source. Equation (5) reduces to¹⁰

$$\lambda_B = 2\Lambda n_{\text{eff}} \quad (6)$$

where λ_B is termed the *Bragg wavelength*—the wavelength at which the forward-propagating LP_{01} mode couples to the reverse-propagating LP_{01} mode. Such coupling is wavelength dependent, since the propagation constants of the two modes are a function of the wavelength. Hence, if an FBG element is interrogated using a broadband optical source, the wavelength at which phase matching occurs is back-reflected. This back-reflected wavelength is a function of the grating period Λ and the

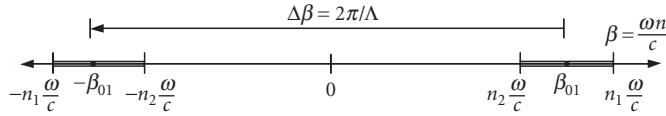


FIGURE 4 Mode-coupling mechanism in fiber Bragg gratings. The large value of $\Delta\beta$ in FBGs requires a small value of the grating periodicity Λ . The hatched regions represent the guided modes in the forward ($\beta > 0$) and reverse ($\beta < 0$) directions.

effective index n_{eff} of the fundamental mode as shown in Eq. (6). Since strain and temperature effects can modulate both of these parameters, the Bragg wavelength is modulated by both of these external perturbations. The resulting spectral shifts are utilized to implement FBGs for sensing applications.

Figure 4 shows the mode-coupling mechanism in fiber Bragg gratings using a β -plot. Since the difference in propagation constants ($\Delta\beta$) between the modes involved in coupling is large, we see from Eq. (4) that only a small period, Λ , is needed to induce this mode coupling. Typically for optical fiber communication system applications the value of λ_B is approximately $1.5\ \mu\text{m}$. From Eq. (6), Λ is thus approximately $0.5\ \mu\text{m}$ for $n_{\text{eff}} = 1.5$, the approximate index of refraction of the glass in a fiber. Due to the small period, on the order of $1\ \mu\text{m}$, FBGs are typically classified as short-period gratings (SPGs).

Bragg Grating Sensor Fabrication

Fiber Bragg gratings have commonly been manufactured using two side-exposure techniques, namely *interferometric* and *phase mask* methods. The interferometric method, shown in Fig. 5, uses an ultraviolet (UV) writing beam at 244 or 248 nm, split into two parts of approximately the same intensity by a beam splitter.⁸ The two beams are focused on a portion of the Ge-doped fiber, whose protective coating has been removed using cylindrical lenses. The period of the resulting interference pattern, and hence the period of the Bragg grating element to be written, is varied by altering the mutual angle θ . A limitation of this method is that any relative vibration of the pairs of mirrors and lenses can lead to the degradation of the quality of the fringe pattern and the fabricated grating; thus the entire system has a stringent stability requirement. To overcome this drawback, Kashyap¹⁰ proposed a novel interferometer technique in which the path difference between the interfering UV beams is produced by propagation through a right-angled prism, as shown in Fig. 6. This geometry is inherently stable because both beams are perturbed similarly by any prism vibration.

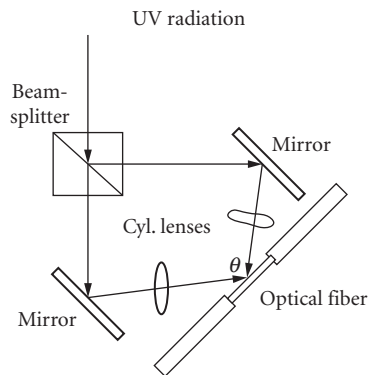


FIGURE 5 Fabrication of Bragg gratings using interferometric scheme.

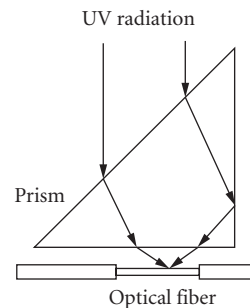


FIGURE 6 Bragg grating fabrication using prism method.

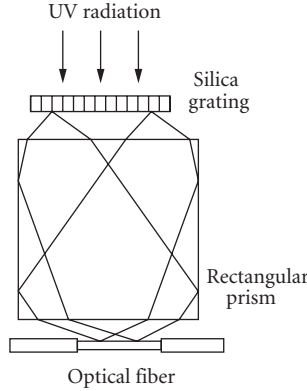


FIGURE 7 Phase mask method of fabricating Bragg gratings.

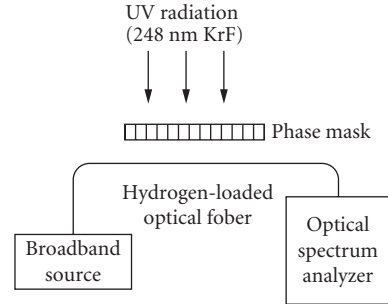


FIGURE 8 Setup to write Bragg gratings in germanosilicate fibers.

The phase mask technique has gained popularity as an efficient holographic side-writing procedure for grating fabrication.¹¹ In this method, shown in Fig. 7, an incident UV beam is diffracted into -1 , 0 , and $+1$ orders by a relief grating typically generated on a silica plate by electron beam exposure and plasma etching. The two first diffraction orders undergo total internal reflection at the glass-air interface of a rectangular prism and interfere at the location of the fiber placed directly behind the mask. This technique is wavelength specific, since the period of the resulting two-beam interference pattern is uniquely determined by the diffraction angle of -1 and $+1$ orders and thus the properties of the phase mask. Obviously, different phase masks are required for the fabrication of gratings at different Bragg wavelengths. A setup for actively monitoring the growth of a grating in the transmission mode during fabrication is shown in Fig. 8.

Bragg Grating Sensors

From Eq. (6) we see that a change in the value of n_{eff} and/or Λ can cause the Bragg wavelength λ to shift. This fractional change in the resonance wavelength $\Delta\lambda/\lambda$ is given by

$$\frac{\Delta\lambda}{\lambda} = \frac{\Delta\Lambda}{\Lambda} + \frac{\Delta n_{\text{eff}}}{n_{\text{eff}}} \quad (7)$$

where $\Delta\Lambda/\Lambda$ and $\Delta n_{\text{eff}}/n_{\text{eff}}$ are the fractional changes in the period and the effective index, respectively. The relative magnitudes of the two changes depend on the type of perturbation to which the grating is subjected. For most applications the effect due to change in effective index is the dominating mechanism.

An axial strain ε in the grating changes the grating period and the effective index and results in a shift in the Bragg wavelength, given by

$$\frac{1}{\lambda} \frac{\Delta\lambda}{\varepsilon} = \frac{1}{\Lambda} \frac{\Delta\Lambda}{\varepsilon} + \frac{1}{n_{\text{eff}}} \frac{\Delta n_{\text{eff}}}{\varepsilon} \quad (8)$$

The first term on the right side of Eq. (8) is unity, while the second term has its origin in the photoelastic effect. An axial strain in the fiber serves to change the refractive index of both the core and the cladding. This results in a variation in the value of the effective index of glass. The photoelastic or strain-optic coefficient is approximately -0.27 . Thus, the variations in n_{eff} and Λ due to strain have contrasting effects on the Bragg peak. The fractional change in the Bragg wavelength due to axial strain is 0.73ε , or 73 percent of the applied strain. At 1550 and 1300 nm, the shifts in the resonance wavelength are 11 nm/% ε and 9 nm/% ε , respectively. An FBG at 1500 nm shifts by 1.6 nm for every 100°C rise in temperature.⁷

Limitations of Bragg Grating Strain Sensors

The primary limitation of Bragg grating sensors is the complex and expensive fabrication technique. Although side-writing is commonly being used to manufacture these gratings, the requirement of expensive phase masks increases the cost of the sensing system. In the interferometric technique, stability of the setup is a critical factor in obtaining high-quality gratings. Since index changes of the order of 10^{-3} are required to fabricate these gratings, laser pulses of high energy levels are necessary.

The second primary limitation of Bragg gratings is their limited bandwidth. The typical value of the full width at half-maximum (FWHM) is between 0.1 and 1 nm. Although higher bandwidths can be obtained by chirping the index or period along the grating length, this adds to the cost of the grating fabrication. The limited bandwidth requires high-resolution spectrum analysis to monitor the grating spectrum. Kersey and Berkoff¹² have proposed an unbalanced Mach-Zehnder interferometer to detect the perturbation-induced wavelength shift. Two unequal arms of the Mach-Zehnder interferometer are excited by the back reflection from a Bragg grating sensor element. Any change in the input optical wavelength modulates the phase difference between the two arms and results in a time-varying sinusoidal intensity at the output. This interference signal can be related to the shift in the Bragg peak and the magnitude of the perturbation can be obtained. Recently, modal interferometers have also been proposed to demodulate the output of a Bragg grating sensor.¹³ The unbalanced interferometers are also susceptible to external perturbations and hence need to be isolated from the parameter under investigation. Moreover, the nonlinear output may require fringe counting, which can be complicated and expensive. Additionally, a change in the perturbation polarity at the maxima or minima of the transfer function curve will not be detected by this demodulation scheme. To overcome this limitation, two unbalanced interferometers may be employed for dynamic measurements.

Cross-sensitivity to temperature leads to erroneous displacement measurements in applications where the ambient temperature has a temporal variation. So a reference grating used to measure temperature change may be utilized to compensate for the output of the strain sensor. Recently, temperature-independent sensing has been demonstrated using chirped gratings written in tapered optical fibers.¹⁴

Finally, the sensitivity of fiber Bragg grating strain sensors may not be adequate for certain applications. This sensitivity of the sensor depends on the minimum detectable wavelength shift at the receiver. Although excellent wavelength resolution can be obtained with unbalanced interferometric detection techniques, standard spectrum analysis systems typically provide a resolution of 0.1 nm. At 1300 nm, this minimum detectable change in wavelength corresponds to a strain resolution of $111 \mu\epsilon$. Hence, in applications where strains smaller than $100 \mu\epsilon$ are anticipated, Bragg grating sensors may not be practical. The dynamic range of strain measurement can be as much as $15,000 \mu\epsilon$.

24.5 LONG-PERIOD GRATING SENSORS

This section discusses the use of novel long-period gratings (LPGs) as displacement-sensing devices. We analyze the principle of operation of these gratings, their fabrication process, typical experimental evaluation, their demodulation process, and their cross-sensitivity to ambient temperature.

Principle of Operation

Long-period gratings that couple the fundamental guided mode to different guided modes have been demonstrated.^{15,16} Gratings with longer periodicities that involve coupling of a guided mode to forward-propagating cladding modes were recently proposed by Vengsarkar et al.^{17,18} As discussed previously, fiber gratings satisfy the Bragg phase-matching condition between the guided and cladding or radiation modes or another guided mode. This wavelength-dependent phase-matching condition is given by

$$\beta_{01} - \beta = \Delta\beta = \frac{2\pi}{\Lambda} \quad (9)$$

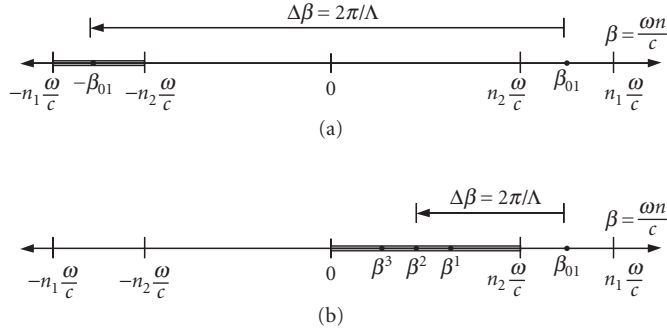


FIGURE 9 Depiction of mode coupling in (a) Bragg gratings and (b) long-period gratings. The differential propagation constant $\Delta\beta$ determines the grating periodicity.

where Λ is the period of the grating and β_{01} and β are the propagation constant of the fundamental guided mode and the mode to which coupling occurs, respectively.

For conventional fiber Bragg gratings, the coupling of the forward-propagating LP_{01} mode occurs to the reverse-propagating LP_{01} mode ($\beta = -\beta_{01}$). Since $\Delta\beta$ is large in this case, as shown in Fig. 9a, the grating periodicity is small, typically on the order of $1\ \mu\text{m}$. Unblazed long-period gratings couple the fundamental mode to the discrete and circularly symmetric, forward-propagating cladding modes ($\beta = \beta^n$), resulting in smaller values of $\Delta\beta$, as shown in Fig. 9b, and hence periodicities ranging in the hundreds of micrometers.¹⁷ The cladding modes attenuate rapidly as they propagate along the length of the fiber, due to the lossy cladding-coating interface and bends in the fiber. Since $\Delta\beta$ is discrete and a function of the wavelength, this coupling to the cladding modes is highly selective, leading to a wavelength-dependent loss. As a result, any modulation of the core and cladding guiding properties modifies the spectral response of long-period gratings, and this phenomenon can be utilized for sensing purposes. Moreover, since the cladding modes interact with the fiber jacket or any other material surrounding the cladding, changes in the index of refraction or other properties of these effective coatings materials can also be detected.

LPG Fabrication Procedure

To fabricate long-period gratings, hydrogen-loaded (3.4 mole %), germanosilicate fibers may be exposed to 248-nm UV radiation from a KrF excimer laser through a chrome-plated amplitude mask possessing a periodic rectangular transmittance function. Figure 10 shows a typical setup used

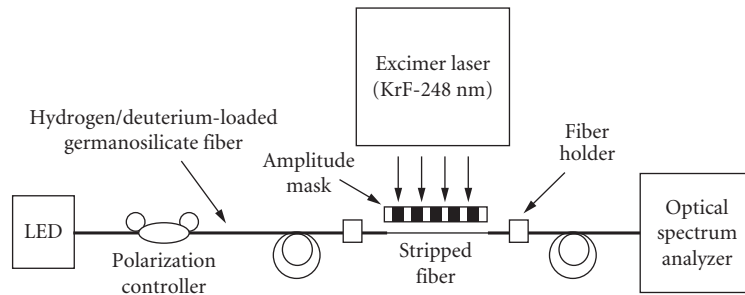


FIGURE 10 Setup to fabricate long-period gratings using an amplitude mask.

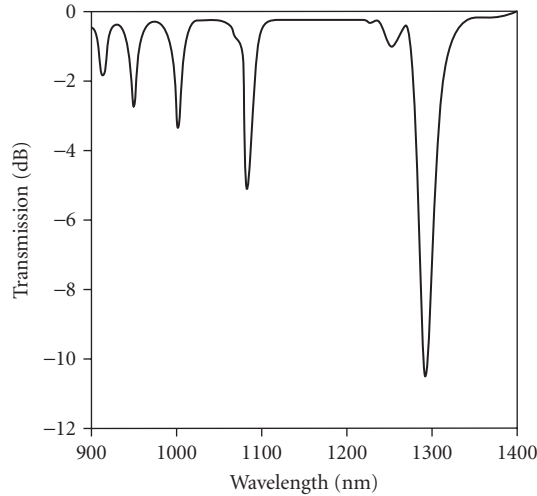


FIGURE 11 Transmission spectrum of a long-period grating written in Corning FLEXCOR fiber with period $\Lambda = 198 \mu\text{m}$. The discrete, spiky loss bands correspond to the coupling of the fundamental guided mode to discrete cladding modes.

to fabricate such gratings. The laser is pulsed at approximately 20 Hz with a pulse duration of several nanoseconds. The typical writing times for an energy of $100 \text{ mJ}/\text{cm}^2/\text{pulse}$ and a 2.5-cm exposed length vary between 6 and 15 min for different fibers. The coupling wavelength λ_p shifts to higher values during exposure due to the photoinduced enhancement of the refractive index of the fiber core and the resulting increase in β_{01} . After writing, the gratings are annealed at 150°C for several hours to remove the unreacted hydrogen. This high-temperature annealing causes λ_p to move to shorter wavelengths due to the decay of UV-induced defects and the diffusion of molecular hydrogen from the fiber. Figure 11 depicts the typical transmittance of a grating. Various attenuation bands correspond to coupling to discrete cladding modes of different orders. A number of gratings can be fabricated at the same time by placing more than one fiber behind the amplitude mask. Due to the relatively long spatial periods, the stability requirements during the writing process are not so severe as those for short-period Bragg gratings.

For coupling to the highest-order cladding mode, the maximum isolation (loss in transmission intensity) is typically in the 5- to 20-dB range on wavelengths depending on fiber parameters, duration of UV exposure, and mask periodicity. The desired fundamental coupling wavelength can easily be varied by using inexpensive amplitude masks of different periodicities. The insertion loss, polarization mode dispersion, backreflection, and polarization-dependent loss of a typical grating are 0.2 dB, 0.01 ps, -80 dB , and 0.02 dB, respectively. The negligible polarization sensitivity and backreflection of these devices eliminates the need for expensive polarizers and isolators.

We now look at representative experiments that have been performed and discussed to examine the displacement sensitivity of long-period gratings written in different fibers.^{19,20} For example, gratings have been fabricated in four different types of fibers—standard dispersion-shifted fiber (DSF), standard 1550-nm fiber, and conventional 980- and 1050-nm single-mode fibers. For the sake of brevity, these will be referred to as fibers A, B, C, and D, respectively. The strain sensitivity of gratings written in different fibers was determined by axially straining the gratings between two longitudinally separated translation stages. The shift in the peak loss wavelength of the grating in fiber D as a function of the applied strain is depicted in Fig. 12 along with that for a Bragg grating (about $9 \text{ nm}/\% \text{ } \epsilon$ at 1300 nm).⁷ The strain coefficients of wavelength shift β for fibers A, B, C, and D are shown in Table 1.

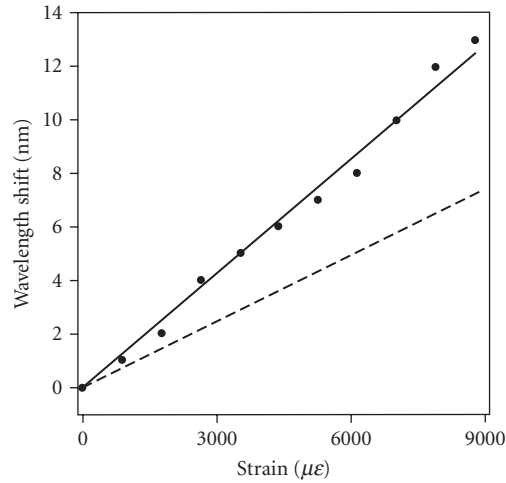


FIGURE 12 Shift in the highest order resonance band with strain for a long-period grating written in fiber D (circles). Also depicted is the shift for a conventional Bragg grating (dashed line).

TABLE 1 Strain Sensitivity of Long-Period Gratings Written in Four Different Types of Fibers

Type of Fiber	Strain Sensitivity (nm/% ϵ)
A—standard dispersion-shifted fiber (DSF)	-7.27
B—standard 1550-nm communication fiber	4.73
C—conventional 980-nm single-mode fiber	4.29
D—conventional 1060-nm single-mode fiber	15.21

Values correspond to the shift in the highest order resonance wavelength.

Fiber D has a coefficient of 15.2 nm/% ϵ , which gives it a strain-induced shift that is 50 percent larger than that for a conventional Bragg grating. The strain resolution of this fiber for a 0.1-nm detectable wavelength shift is 65.75 $\mu\epsilon$.

The demodulation scheme of a sensor determines the overall simplicity and sensitivity of the sensing system. Short-period Bragg grating sensors were shown to possess signal processing techniques that are complex and expensive to implement. We now present a simple demodulation method to extract information from long-period gratings. The wide bandwidth of the resonance bands enables the wavelength shift due to the external perturbation to be converted into an intensity variation that can be easily detected.

Figure 13 shows the shift induced by strain in a grating written in fiber C. The increase in the loss at 1317 nm is about 1.6 dB. A laser diode centered at 1317 nm was used as the optical source, and the change in transmitted intensity was monitored as a function of applied strain. The transmitted intensity is plotted in Fig. 14 for three different trials. The repeatability of the experiment demonstrates the feasibility of using this simple scheme to utilize the high sensitivity of long-period gratings. The transmission of a laser diode centered on the slope of the grating spectrum on either side of the resonance wavelength can be used as a measure of the applied perturbation. A simple detector and amplifier combination at the output can be used to determine the transmission through the detector.

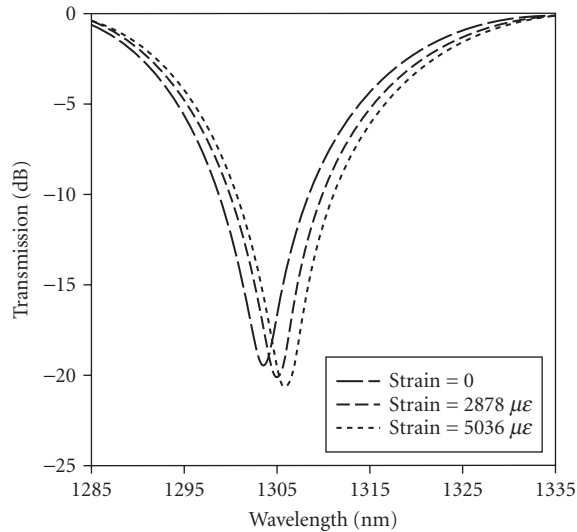


FIGURE 13 Strain-induced shift in a long-period grating fabricated in fiber C. The loss at 1317 nm increases by 1.6 dB due to the applied strain (5036 $\mu\epsilon$).

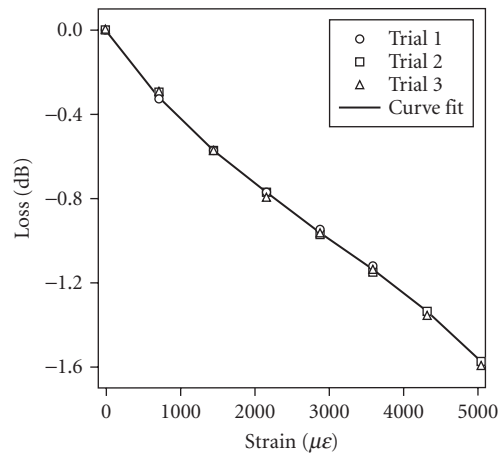


FIGURE 14 The change in the grating transmission at 1317 nm as a function of strain for three different trials. The increase in loss by 1.6 dB at 5036 $\mu\epsilon$ provides evidence of the feasibility of the simple setup used to measure strain.

On the other hand, a broadband source can also be used to interrogate the grating. At the output an optical bandpass filter can be used to transmit only a fixed bandwidth of the signal to the detector. The bandpass filter should again be centered on either side of the peak loss band of the resonance band. These schemes are easy to implement, and unlike the case for conventional Bragg gratings, complex and expensive interferometric demodulation schemes are not necessary.²⁰

TABLE 2 Temperature Sensitivity of Long-Period Gratings Written in Four Different Types of Fibers

Type of Fiber	Temperature Sensitivity (nm/°C)
A—standard dispersion-shifted fiber (DSF)	0.062
B—standard 1550-nm communication fiber	0.058
C—conventional 980-nm single mode fiber	0.154
D—conventional 1060 nm single mode fiber	0.111

Values correspond to the shift in the highest order resonance wavelength.

Temperature Sensitivity of Long-Period Gratings

Gratings written in different fibers were also tested for their cross-sensitivity to temperature.²⁰ The temperature coefficients of wavelength shift for different fibers are shown in Table 2. The temperature sensitivity of a fiber Bragg grating is 0.014 nm/°C. Hence the temperature sensitivity of a long-period grating is typically an order of magnitude higher than that of a Bragg grating. This large cross-sensitivity to ambient temperature can degrade the strain sensing performance of the system unless the output signal is adequately compensated. Multiparameter sensing using long-period gratings has been proposed to obtain precise strain measurements in environments with temperature fluctuations.¹⁹

In summary, long-period grating sensors are highly versatile. These sensors can easily be used in conjunction with simple and inexpensive detection techniques. Experimental results prove that these methods can be used effectively without sacrificing the enhanced resolution of the sensors. Long-period grating sensors are insensitive to input polarization and do not require coherent optical sources. Cross-sensitivity to temperature is a major concern while using these gratings for strain measurements.

24.6 COMPARISON OF SENSING SCHEMES

Based on these results, interferometric sensors have a high sensitivity and bandwidth but are limited by nonlinearity in their output signals. Conversely, intrinsic sensors are susceptible to ambient temperature changes, while grating-based sensors are simpler to multiplex. Each may be used in specific applications.

24.7 CONCLUSION

We have briefly summarized the performance of four different interferometric and grating-based sensors as representative of the very wide range of possible optical fiber sensor instrumentation and approaches. This analysis was based on the sensor head fabrication and cost, signal processing, cross-sensitivity to temperature, resolution, and operating range. Relative merits and demerits of the various sensing schemes were discussed.

24.8 REFERENCES

1. R. O. Claus, M. F. Gunther, A. Wang, and K. A. Murphy, "Extrinsic Fabry-Perot Sensor for Strain and Crack Opening Displacement Measurements from Minus 200 to 900°C," *Smart Mat. Struct.* **1**:237–242 (1992).
2. K. A. Murphy, M. F. Gunther, A. M. Vengsarkar, and R. O. Claus, "Fabry-Perot Fiber Optic Sensors in Full-Scale Fatigue Testing on an F-15 Aircraft," *App. Opt.* **31**:431–433 (1991).
3. V. Bhatia, C. A. Schmid, K. A. Murphy, R. O. Claus, T. A. Tran, J. A. Greene, and M. S. Miller, "Optical Fiber Sensing Technique for Edge-Induced and Internal Delamination Detection in Composites," *J. Smart Mat. Struct.* **4** (1995).

4. C. E. Lee and H. F. Taylor, "Fiber-Optic Fabry-Perot Temperature Sensor Using a Low-Coherence Light Source," *J. Lightwave Technol.* **9**:129–134 (1991).
5. J. A. Greene, T. A. Tran, K. A. Murphy, A. J. Plante, V. Bhatia, M. B. Sen, and R. O. Claus, "Photo Induced Fresnel Reflectors for Point-Wise and Distributed Sensing Applications," in *Proceedings of the Conference on Smart Structures and Materials*, SPIE'95, paper 2444-05, February 1995.
6. J. Sirkis, "Phase-Strain-Temperature Model for Structurally Embedded Interferometric Optical Fiber Strain Sensors with Applications," *Fiber Opt. Smart Struct. Skins IV*, SPIE **1588** (1991).
7. K. O. Hill, Y. Fuiji, D. C. Johnson, and B. S. Kawasaki, "Photosensitivity in Optical Fiber Waveguides: Applications to Reflection Filter Fabrication," *Appl. Phys. Lett.* **32**:647 (1978).
8. G. Meltz, W. W. Morey, and W. H. Glenn, "Formation of Bragg Gratings in Optical Fibers by Transverse Holographic Method," *Opt. Lett.* **14**:823 (1989).
9. P. J. Lemaire, A. M. Vengsarkar, W. A. Reed, V. Mizrahi, and K. S. Kranz, "Refractive Index Changes in Optical Fibers Sensitized with Molecular Hydrogen," in *Proceedings of the Conference on Optical Fiber Communications*, OFC'94, Technical Digest, paper TuL1, 1994, p. 47.
10. R. Kashyap, "Photosensitive Optical Fibers: Devices and Applications," *Opt. Fiber Technol.* **1**:17–34 (1994).
11. D. Z. Anderson, V. Mizrahi, T. Ergodan, and A. E. White, "Phase-Mask Method for Volume Manufacturing of Fiber Phase Gratings," in *Proceedings of the Conference on Optical Fiber Communication*, post-deadline paper PD16, 1993, p. 68.
12. A. D. Kersey and T. A. Berkoff, "Fiber-Optic Bragg-Grating Differential-Temperature Sensor," *IEEE Phot. Technol. Lett.* **4**:1183–1185 (1992).
13. V. Bhatia, M. B. Sen, K. A. Murphy, A. Wang, R. O. Claus, M. E. Jones, J. L. Grace, and J. A. Greene, "Demodulation of Wavelength-Encoded Optical Fiber Sensor Signals Using Fiber Modal Interferometers," *SPIE Photon. East*, Philadelphia, Pa, paper 2594–09, October 1995.
14. M. G. Xu, L. Dong, L. Reekie, J. A. Tucknott, and J. L. Cruz, "Chirped Fiber Gratings for Temperature-Independent Strain Sensing," in *Proceedings of the First OSA Topical Meeting on Photosensitivity and Quadratic Nonlinearity in Glass Waveguides: Fundamentals and Applications*, paper PMB2, 1995.
15. K. O. Hill, B. Malo, K. Vineberg, F. Bilodeau, D. Johnson, and I. Skinner, "Efficient Mode Conversion in Telecommunication Fiber Using Externally Written Gratings," *Electron. Lett.* **26**:1270–1272 (1990).
16. F. Bilodeau, K. O. Hill, B. Malo, D. Johnson, and I. Skinner, "Efficient Narrowband $LP_{01} \leftrightarrow LP_{02}$ Mode Converters Fabricated in Photosensitive Fiber: Spectral Response," *Electron. Lett.* **27**:682–684 (1991).
17. A. M. Vengsarkar, P. J. Lemaire, J. B. Judkins, V. Bhatia, J. E. Sipe, and T. E. Ergodan, "Long-Period Fiber Gratings as Band-Rejection Filters," in *Proceedings of Conference on Optical Fiber Communications*, OFC'95, post-deadline paper, PD4-2, 1995.
18. A. M. Vengsarkar, P. J. Lemaire, J. B. Judkins, V. Bhatia, J. E. Sipe, and T. E. Ergodan, "Long-Period Fiber Gratings as Band-Rejection Filters," *J. Lightwave Technol.* **14**(1):58–65 (1996).
19. V. Bhatia, M. B. Burford, K. A. Murphy, and A. M. Vengsarkar, "Long-Period Fiber Grating Sensors," in *Proceedings of the Conference on Optical Fiber Communication*, paper ThP1, February 1996.
20. V. Bhatia and A. M. Vengsarkar, "Optical Fiber Long-Period Grating Sensors," *Opt. Lett.* **21**:692–694 (1996).

24.9 FURTHER READING

- Bhatia, V., M. J. de Vries, K. A. Murphy, R. O. Claus, T. A. Tran, and J. A. Greene, "Extrinsic Fabry-Perot Interferometers for Absolute Measurements," *Fiberoptic Prod. News* **9**:12–3 (December 1994).
- Bhatia, V., M. B. Sen, K. A. Murphy, and R. O. Claus, "Wavelength-Trackable White Light Interferometry for Highly Sensitive Strain and Temperature Measurements," *Electron. Lett.*, 1995, submitted.
- Butter, C. D., and G. B. Hocker, "Fiber Optics Strain Gage," *Appl. Opt.* **17**:2867–2869 (1978).
- Sirkis, J. S. and H. W. Haslach, "Interferometric Strain Measurement by Arbitrarily Configured, Surface Mounted, Optical Fiber," *J. Lightwave Technol.*, **8**:1497–1503 (1990).

HIGH-POWER FIBER LASERS AND AMPLIFIERS

Timothy S. McComb, Martin C. Richardson, and Michael Bass

*CREOL, The College of Optics and Photonics
University of Central Florida
Orlando, Florida*

25.1 GLOSSARY

Symbols

a	fiber core radius
F	filling factor of air holes in a PCF
G	signal gain in a fiber
$g(z)$	gain as a function of length
k_0	wavenumber in vacuum
M^2	beam quality factor; a value of 1 indicates a diffraction-limited beam
n	refractive index
N	population density of an energy level
P	power of laser emission
R	mirror reflectivity
V_{eff}	effective V parameter for a PCF
w_L	mode field radius
α	propagation loss
η_{pump}	pump overlap with doped region
η_{signal}	signal overlap with doped region
λ	wavelength
Λ	pitch or spacing of air holes in a PCF
ν	frequency
σ	emission or absorption cross section
τ	upper-state lifetime

Abbreviations and Definitions

Air-cladding	Region of air holes connected by thin glass bridges to the cladding region of a photonic crystal fiber forming an air-glass boundary used to guide pump radiation with high numerical aperture
AO	Acousto-optical
AR	Antireflective
ASE	Amplified spontaneous emission
CCC	Chirally coupled core, a type of fiber with a small satellite core chirally wrapped around the signal core; the small core couples higher-order modes out from the central core allowing larger mode areas
Cladding	The region of an optical fiber that surrounds the core; in conventional fibers this region is lower refractive index than the core to allow total internal reflection guidance
Core	The central region in a fiber where signal light is guided in most cases by total internal reflection
CPA	Chirped pulse amplification, a technique used to amplify ultrashort pulses whereby the pulse is temporally stretched before being amplified and recompressed after amplification in order to avoid high peak powers in the amplifier
DFB	Distributed feedback laser
Dichroic	An optical element that exhibits desired properties at two separate wavelengths (for instance, a mirror that is HR at one wavelength and HT at another)
EO	Electro-optical
FBG	Fiber Bragg grating
GG IAG	Gain-guided index antiguided fibers are fibers with core index less than that of the cladding but which can have gain in the core to compensate for losses
GMRF	Guided-mode resonance filter, a device consisting of a subwavelength grating on top of a waveguide layer used to form a narrow band reflectivity
HOM	Higher-order mode, any mode of a fiber of higher order than the fundamental mode
HR	High reflectivity
HT	High transmission
LMA	Large mode area; any of a number of techniques or technologies used to increase the mode field diameter of a fiber intended for use in an amplifier or laser system where high power operation requires larger mode field diameters to avoid nonlinear effects
MCVD	Modified chemical vapor deposition; a technique for fiber preform fabrication involving depositing chemical “soots” on the inside of a glass tube and subsequently collapsing the tube
MFA	Mode field adaptor; device used to change the mode field diameter of a fiber to match that of a second fiber
MFD	Mode field diameter; diameter of the distribution of radiation within an optical fiber usually at $1/e^2$ value of power
MOPA	Master oscillator power amplifier; system involving a low-power laser source (the oscillator) amplified by one or many amplifier chains
Multicore fiber	Fibers possessing several cores within a single cladding with each core designed to separately guide light and with these cores arrayed in such a pattern as to produce desired beam profile in the far field
NA	Numerical aperture of a fiber; a function of square root of the difference of the squares of the refractive index of core and cladding
OVD	Outside vapor deposition; a technique for fiber preform manufacture
PCF	Photonic crystal fiber; one of several fiber types with a latticelike structure of different refractive indices to create guidance in a fiber

PM	Polarization-maintaining fiber; fiber designed with one of a variety of stress-inducing structures to introduce birefringence in the fiber core
Pump cladding	The region of a double-clad fiber within which pump radiation is guided so it can cross through the doped fiber core, surrounded by a lower index glass or polymer layer
SBS	Stimulated Brillouin scattering
SESAM	Semiconductor saturable absorber mirror; a mirror with built-in saturable absorption that is often used to mode-lock fiber laser systems
SMET	Single-mode excitation technique; a technique where light from a laser source is launched into a fiber with appropriate care to launch only the fundamental mode of the fiber, even if the fiber itself is multimode
SPM	Self-phase modulation
SRS	Stimulated Raman scattering
TEC	Thermally expanded core; a technique for heating a fiber core causing the core dopants to diffuse, thus expanding it
TFB	Tapered fiber bundle; device combining multiple pump fibers into a bundle to deliver pump radiation to a fiber laser or amplifier
USP	Ultrashort pulse
VAD	Vapor axial deposition; a technique for fiber preform manufacture
VBG	Volume Bragg grating
V-parameter	Parameter that indicates guidance properties of a fiber; a value less than 2.405 indicates single-mode guidance; V is a function of wavelength, NA, and core radius
ZBLAN	A fluoride glass named for its chemical composition containing ZnF_2 , BaF_2 , LaF_3 , AlF_3 , and NaF

25.2 INTRODUCTION

Introductory Remarks

Although fiber lasers were first demonstrated at the dawn of the laser age,¹ it is only recently that fiber lasers have risen in visibility on the landscape of laser development. This resurgence arose as a consequence of transformational changes in pump technology, and fiber design and fabrication techniques. Thus nowadays a fiber laser can be thought of as a device for the conversion of light from low brightness laser diodes to high brightness, highly coherent laser light. Their rise in power and brightness capabilities has been so significant that they are now beginning to invade the applications space once dominated by solid-state lasers. As these devices permeate many fields of laser applications in manufacturing medicine and defense, there are growing demands and constraints placed on the fiber laser's spectrum, beam quality, and pulse duration. These demands can only be met with solid-state lasers at the expense of efficiency, complexity, and cost. However, they are in many cases a natural consequence, or relatively simple modification of modern fiber lasers.

As industrial, medical, and defense applications of the high-power lasers force increasing levels of electrical efficiency, beam quality, and ruggedness with commensurate reductions in cost, complexity, and footprint, fiber lasers will increasingly meet these needs. In this chapter we summarize the basic principles of fiber lasers, the latest developments in design and fabrication, their different modalities in output characteristics, and their potential for future growth in output power and overall utility.

A Brief History of Fiber Lasers

Although most of the developments in fiber lasers did not occur until the onset of the fiber optics telecommunications and the development of high-power optical diodes, it should not be forgotten

that the basic concept of the fiber laser, that of a doped fiber core surrounded by optically transparent cladding was devised by Snitzer in the early 1960s at the very birth of the laser age.¹⁻³ The current rapid growth in high-power fiber laser achievements owes its origins to (1) the development of high quality silica fibers and (2) the development of high-power diode laser technology.

The rapid growth in telecommunications and the large investments in improving optical fiber technology spurred new interest in fiber lasers in the 1970s. Multimode core-pumped fiber lasers were demonstrated, pumped by both diode lasers and bulk lasers.^{4,5} Silica became the standard host material for most fiber lasers, a direct extension of fiber optics telecommunications technology. Despite the improving optical quality of silica fibers, the first single-mode fiber laser was not demonstrated until the mid-1980s. This laser was still of low power, due to the unavailability of high brightness pump sources for the directly core-pumped scheme.⁶ The 1980s also saw the first tunable and Q-switched fiber lasers.^{7,8} At that time fiber laser development was largely driven by the potential of (erbium) Er-doped fiber at approximately 1.5 μm ,^{9,10} to serve as low-loss, low-dispersion, optical amplifiers¹¹ transformed the telecommunications industry.

High laser power single-mode fiber lasers beyond approximately 1 W of output power, at this time were limited by available single-mode diode pump sources required for the core-pumping scheme. It was difficult to efficiently couple light from then-available diode sources into single-mode fiber cores due to the diode's elliptical beam output shape and limitations on single transverse mode diode output powers. The most significant advance to overcoming this problem has become the basis for all high-power fiber lasers manufactured today, the so called "double clad fiber" was first proposed in 1988 by Snitzer.¹² Shown schematically in Fig. 1, the introduction of a second, undoped inner cladding of much larger diameter (typically > 100 μm) than the doped core allowed for effective coupling of much higher pump powers. So long as this inner cladding layer had a refractive index less than that of the core, there was effective coupling of the pump light into the core region. Similarly the outer cladding layer must have a refractive index less than that of the inner cladding region, in order to limit the loss of pump light from the fiber.

This innovation led to a period of rapid increase in fiber laser output power, limited only by pump diode power.¹²⁻¹⁵ The double clad fiber enabled scaling to 110 W average powers and approximately 100 μJ peak powers in purely single-mode fiber cores. At these power levels, however, further increases in the power of fiber lasers were impeded by a set of new problems associated with nonlinear optical effects and fiber damage in the very small single-mode cores.^{16,17} New concepts for increasing the size of the fiber laser core, now commonly referred to as large-mode-area (LMA) fibers were demonstrated in Refs. 17 to 19. LMA technologies have given birth to the state-of-the-art high-power fiber lasers that exist today, enabling continuous wave (CW) lasers to reach powers of more than 3 kW and pulsed fiber laser to reach more than 4 mJ output power in nanosecond pulses with diffraction-limited beam quality.^{20,21}

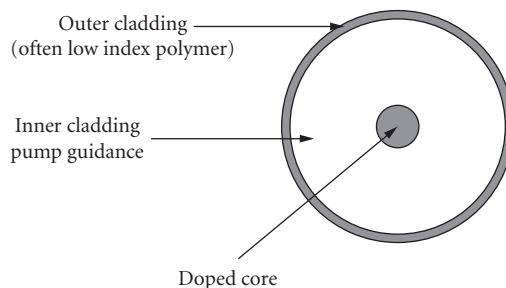


FIGURE 1 Simple schematic of a double clad fiber. The refractive index profile should have the following index relation: $n_{\text{core}} > n_{\text{inner cladding}} > n_{\text{outer cladding}}$.

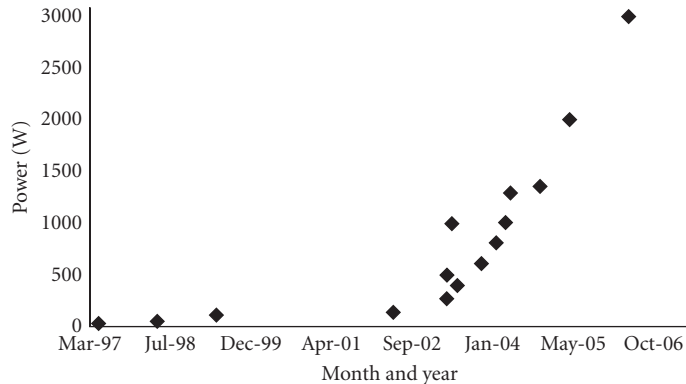


FIGURE 2 Plot of highest achieved CW output powers in fiber lasers operating at 1 μm wavelength dating from 1997 to the present day based on Refs. 16 and 21 to 34.

Rapid Growth of Fiber Lasers

Beginning with the development of the first double clad fiber laser and continuing with LMA fiber laser technologies, the output powers of fiber lasers has undergone near-exponential growth in the last 10 years. A convenient benchmark for such growth can be seen in the output power of CW Yb fiber lasers beginning around 1997. Figure 2 shows the highest reported output powers of Yb CW fiber lasers over a period of years.

The LMA fiber technologies described here are critical to the future growth of high-power fiber lasers. In the future, the development and refinement of LMA technologies will lead to further growth in output powers, with 10 kW being a reasonable goal in the not-too-distant future. Indeed, recently a paper outlining the limitations in output power from single-mode Yb-based fiber lasers has approximated the upper limit on output power at 36 kW from a single fiber based on today's technology and reasonable assumptions for future growth.³⁵

Comparison of Fiber Lasers to Bulk Lasers

High-power fiber lasers possess several advantages when compared to bulk solid-state lasers. These include their compact, simple construction, simplified thermal management, extremely high beam quality, and high optical-to-optical efficiency. These beneficial properties are forces driving the use of fiber lasers for high-power applications.

Fiber laser systems offer a clear advantage in terms of their size and performance in harsh environments. This follows from the fact that fiber lasers are often completely monolithic. That is, they are comprised of an unbroken chain of all-fiber-based components with no need for realignment and no potential for contamination. Though some very high-power kilowatt-class fiber-based systems have not yet been made completely monolithic and still require some free-space components, the fact that a majority of the cavity exists "in fiber" still provides a stability benefit compared to bulk solid-state lasers. In fact the highest power systems demonstrated have been completely monolithic systems.^{21,32}

Fiber lasers also have a distinct advantage over bulk lasers in terms of thermal management resulting from the much larger surface area of fiber lasers compared to bulk lasers.^{36,37} This enables the heat deposition resulting from the quantum defect of the pump light to be distributed over the full length of a fiber. Most fiber lasers can operate with only minimal attention to thermal management, though cooling and attention to thermal issues in fiber lasers cannot be completely ignored for reasons including polymer coating integrity and laser efficiency as suggested by studies done in Refs. 38 to 40.

Often, high-power applications call for the power to be delivered at long distances from the laser output itself or for light to be focused to extremely small spot sizes leading to a requirement for diffraction-limited beams. In a fiber laser, the transverse modes are defined by the fiber itself. Correct fiber design ensures that no other modes besides the lowest-order transverse mode are allowed to exist in the waveguide, the core of the fiber, leading to excellent output beam quality. Bulk lasers are often victims of optical distortions due to thermal lensing and birefringence in the gain media caused by temperature gradients and the temperature sensitive nature of the refractive index. In general, fiber lasers are mostly immune to thermal gradient induced optical distortion because of more efficient heat removal and the wave-guiding nature of the fiber.

25.3 Fiber Laser Limitations

Despite the many advantages of fiber lasers, fiber-based systems also have some limitations that require further research and development to mitigate. These limitations all stem from the small core sizes of fibers.

Optical Damage

The most obvious limitation is the damage threshold of the fiber core material due to the laser high-power density in the relatively small fiber core area. The bulk damage threshold of silica is extremely high ($\sim 600 \text{ GW/cm}^2$ at $\sim 1000 \text{ nm}$ wavelength) though tightly focused pulses can damage the bulk material. However, damage most easily occurs when light exiting the fiber reaches the surface damage threshold, which in silica is approximately 40 GW/cm^2 .⁴¹ The most obvious way to mitigate this damage threshold is to increase the core size; however, due to the need to maintain beam quality this technique has limitations. An alternative method for damage mitigation in fiber amplifiers is end capping, a process which involves splicing a coreless short section of fiber onto the end of a fiber, allowing the expansion of the beam before reaching the glass-air interface.

Nonlinear Effects

A second issue in high-power fiber lasers is the result of detrimental nonlinear effects. Such effects are based on third-order nonlinearities in the glass such as self phase modulation, stimulated Raman scattering (SRS), stimulated Brillouin scattering (SBS), and self focusing. Details on the origins and background of such effects can be found in Chap. 10, “Nonlinear Effects in Optical Fibers,” in this volume and Chap. 15, “Stimulated Raman and Brillouin Scattering” in Vol. IV. The impact and severity of these nonlinear effects varies depending on the laser type. Narrow linewidth lasers suffer from unwanted spectral broadening at high powers caused by stimulated Brillouin scattering. Pulsed lasers can experience the effects of Raman scattering. Ultrashort pulse lasers can experience pulse distortions based on self phase modulation. In addition, high average powers cause a hard limitation due to the onset of self focusing in bulk glass leading eventually to catastrophic damage. With the exception of self-focusing, which is core size independent; other nonlinear effects can again be mitigated by simply increasing core diameter. In addition, other techniques such as acoustic design of fiber and thermal control of a fiber can be used to reduce such effects as SBS.^{42–44}

Energy Storage

A limitation in high power fiber lasers with high energy pulses is the small gain volume of the doped core. Even in a very long fiber the total volume of gain medium is small. Fiber lasers that must have high pulse energies find a limitation in terms of the capability to store sufficient energy in the fiber core.

In addition, the leading edge of a pulse can “steal” or saturate the gain in fiber amplifiers and the pulse itself can become temporally distorted. Such issues can be mitigated by of course increasing the fiber core diameter, leading to more gain medium volume and, consequently, higher energy storage. In addition, pulse deformation can be circumvented by using an input pulse designed to compensate for the deformation with a shape designed to accommodate for the nonuniform temporal gain.

CW Damage Threshold

The damage threshold of extremely high-power CW fiber lasers is reviewed in Ref. 35. There each of the previously discussed damage mechanisms, as well as additional damage considerations are considered in order to make an estimate of maximum achievable power from a single-mode single-fiber device operating at a wavelength near 1000 nm.³⁵ Using this reference as a basis the reader can also make some extrapolations about the damage threshold of fiber lasers at longer, eyesafe wavelengths. In many cases it appears that such longer-wavelength fiber laser (most notably at 1.5 and 2 μm) may have improved power handling capabilities in many situations compared to their 1- μm counterparts.

25.4 Fiber Laser Fundamentals

Fiber Laser Operation

The fundamentals of fiber laser operation and amplification are the same as those in any laser system. Equations describing laser operation of a bulk laser can be adapted to fiber lasers by taking into account the wave-guiding nature of the fiber and using the appropriate parameters. A detailed description of laser operation as a whole can be found in Chap. 16, “Lasers” and Chap. 23, “Quantum Theory of the Laser,” in Vol. II of this *Handbook*. Included here are only a few equations useful to the design and operation of fiber lasers.

It is convenient to describe the operation of a fiber gain medium by a set of rate equations. The following define the operation of a simple fiber laser system that, depending on the sign selected, can have pump and signal light traveling in either direction in the fiber,

$$\begin{aligned}\pm \frac{\partial P_{\text{pump}}^{\pm}}{\partial z} &= \eta_{\text{pump}} (\sigma_{10}(\lambda_{\text{pump}}) N_1 - \sigma_{01}(\lambda_{\text{pump}}) N_0) P_{\text{pump}}^{\pm} \\ \pm \frac{\partial P_{\text{signal}}^{\pm}}{\partial z} &= \eta_{\text{signal}} (\sigma_{10}(\lambda_{\text{signal}}) N_1 - \sigma_{01}(\lambda_{\text{signal}}) N_0) P_{\text{signal}}^{\pm}\end{aligned}\quad (1)$$

Here the $\pm$ indicates direction of propagation, N_i is the energy of a given level i , σ_{ij} is the emission or absorption cross section from level i to level j , P is power of a given signal, η_{pump} is the pump overlap with the doped region, and η_{signal} is the signal overlap with the doped region.^{45,46} This model does not take amplified spontaneous emission (ASE) or temporal effects into account. Equation (1) can be solved by applying slightly more sophisticated modeling discussed in Refs. 45 and 46. Solving the coupled differential Eq. (1) numerically, one can obtain the change in pump and laser signal power over distance along the fiber and in time. In addition, by using the appropriate boundary conditions these equations can be used to model amplifiers in the co-, counter- and bidirectional propagation of pump light and also oscillators.

Considering the signal power in the fiber, the total gain in a fiber is the integral of gain along the fiber length where the gain at any given point along the fiber is given by

$$g(z) = [\sigma_{10}(\lambda_{\text{signal}}) N_1(z) - \sigma_{01}(\lambda_{\text{signal}}) N_0(z)] \quad (2)$$

Thus, the signal gain in a fiber can be expressed as

$$G = \exp \left\{ \frac{[\sigma_{10}(\lambda_{\text{signal}}) + \sigma_{01}(\lambda_{\text{signal}})] \eta_q \tau_{10} P_{\text{absorbed}}}{h \nu_{\text{pump}} A_{\text{core}}} - \sigma_{01}(\sigma_{\text{signal}}) N_0(z) L - \alpha_L L \right\} \quad (3)$$

with α_L the core propagation loss, and all other terms as defined earlier.⁴⁷ The final expression to be defined is the laser saturation power given by⁴⁷

$$P_{\text{satsignal}} = \frac{h \nu_{\text{signal}} A_{\text{core}}}{[\sigma_{10}(\lambda_{\text{signal}}) + \sigma_{01}(\lambda_{\text{signal}})] \tau_{10}} \quad (4)$$

This saturation power can be used to describe the change in the small signal gain as the power is increased. In addition, in order for the amplifier to have efficient energy extraction, the signal input must be on the order of the saturation power.

Another important relationship for fiber lasers describes the relative output power from either end of a fiber. Knowledge of the optimum amount of feedback required into one end of the fiber is useful to ensure a majority of output from the other end. To determine the exact output power performance of a fiber laser, either solutions to coupled differential equations for signal round trip propagation in a fiber or a detailed Rigrod analysis⁴⁸ must be computed. However a simplified expression can be written if certain assumptions are made, and the ratio of output power from each fiber end as a function of their reflectivity can be written as follows:

$$\frac{P_1}{P_2} = \frac{1 - R_1}{1 - R_2} \sqrt{\frac{R_2}{R_1}} \quad (5)$$

where $R_{1 \text{ or } 2}$ is the reflectivity of end 1 or 2. Clearly if the reflectivity of one end is far greater than the other, the output power will be predominantly from the lower reflectivity end.^{46,48} Since most simple fiber lasers use approximately 4 percent Fresnel reflection at one end as an output coupler (and even those that employ fiber Bragg grating output coupler use reflectivities typically in the range of 4 to 15%), only relatively low feedback (much less than 90%) is required on the opposing end to ensure that a majority of power leaves from the low reflectivity end.

Important Fiber Equations

A second set of equations related to fiber lasers govern the guidance of light in the fibers themselves. The fundamentals of light guidance in optical fibers can be found in Part 4, "Fiber Optics," Vol. V and Ref. 49. Since LMA fiber lasers are currently the most promising for substantial output powers the following discussion deals with guided light propagation in this type of fiber laser.

The fiber V parameter is defined as

$$V = k_0 a \sqrt{n_1^2 - n_2^2} \quad (6)$$

where is k_0 the wavenumber and a the core radius and with square root of the difference of the squares of n_1 and n_2 alternately expressed as the fiber NA (numerical aperture). The V parameter can be used to determine the guiding properties of a fiber with a given core and numerical aperture. A fiber core with a V parameter of less than 2.405 can sustain only a single lowest-order mode. Most LMA fibers have a V parameter closer to 3 or 4, with large core radii a and small numerical apertures NA. Weak guidance of higher-order modes in LMA fibers allows them to be stripped out by using techniques that provide preferential loss to higher-order modes discussed later and hence enables single-mode propagation in a multimode fiber. The control of the NA and hence the V parameter is therefore a principal factor in LMA fiber laser design.

It is important to note that the true mode area of a fiber does not correspond to the actual core size of the fiber, due to the multimode nature of LMA fiber cores fundamental mode diameter (also known as mode field diameter) may be smaller than the actual core diameter. In the case of truly single-mode fibers the mode field diameter may actually be larger than core diameter. Hence, when quoting mode area (the important factor in determining damage threshold and nonlinearities in fiber lasers) one should use the actual size of the lowest-order mode, not the core size. This is especially important when dealing with fiber splicing or mode field adaptation between dissimilar fibers where splice or device losses must be minimized. This mode diameter can be approximated by the equation

$$w_L \approx a \left(0.65 + \frac{1.619}{V^{3/2}} + \frac{2.879}{V^6} \right) \quad (7)$$

where w_L is the mode radius, V is the V parameter and a is core radius, it should be noted that this is valid for the mode field radius of the fundamental mode in the cases where V is more than 2.405.⁵⁰

We now consider briefly the highly multimode regime. In this case hundreds to thousands of fiber modes can exist in different mode families. The actual number of modes in any fiber is proportional to V^2 , so for fibers with very large core and/or NA, a very large number of modes can propagate.⁵¹ It is a property of light in fibers that as one goes to higher mode number, the location of energy in the higher-order modes moves outward from the center of the fiber. As a consequence, when light is propagating in a highly multimode fiber (for instance, the pump cladding of a double clad fiber), there is a significant portion of energy that may never cross into the core. This can be both a positive effect in terms of evanescent pump coupling and a negative effect in terms of helical modes not being absorbed in the core of double clad fibers. As consequence, one must be aware of not only the number of modes but their shape within the fiber when making design considerations in a fiber laser system.

25.5 FIBER LASER ARCHITECTURES

Fiber laser architectures have evolved as improvements in pump coupling technologies and laser diode brightness, and fiber-based components have driven higher power, and more efficient, compact, stable, and robust platforms. Originally so-called “free-space coupling” using conventional optical elements to image the light output from diode lasers or diode laser bars (lenses, prisms, etc.) was the only effective fiber laser architecture for high-power systems and it still offers advantages for developmental systems. However, there are inherent advantages to “all-fiber” or “monolithic” systems in simplicity, efficiency, stability, compactness, and cost. Here the relevant diode and fiber technologies are described before free-space pumping architectures are laid out, and the latest all-fiber approaches are summarized.

Pumping Techniques

After the fiber itself, the most critical element of a fiber laser system is the pump delivery technique. Since diode lasers are used to pump fiber lasers, this section describes how low-brightness diode laser light is coupled into the core of a fiber laser or amplifier. Efficient fiber lasers rely on effective launch of high average diode laser pump power into the fiber. Unabsorbed or improperly launched pump light results in waste heat that can cause thermal damage to polymer coatings on the fiber, to the pump combiners, to other devices in the system or to the fiber itself. It is also a waste of expensive diode laser pump power. As a result in any fiber laser system, whether free space coupled or monolithic, the design of the pumping system is critical. Fiber-pumping design can be divided into issues of core-clad configuration, choice of diode laser source, and actual pump power delivery method into the fiber.

Pump Launch Schemes By far the simplest and most widely used pump scheme is end pumping by imaging the diode laser delivery fiber (or diode output facet itself) onto the end of the fiber laser or

amplifier. Two lenses are used, designed for minimal aberrations, resulting in small pump spot size for efficient coupling. The pump light must be delivered into a spot less than the diameter of the fiber and within the fiber pump cladding's numerical aperture. The output of bare diode bars or stacks can be shaped to take their long, narrow distributions and make them approximately circular while conserving brightness. Two methods for this both involve using fast and slow axis micro lenses on bars and then using either stacked glass plates to slice the beam into sections and stack the sections on top of each other,⁵² or by using two mirrors to restack the different parts of the beam.⁵³ These methods allow for the use of bare diode bars or stacks rather than fiber-coupled diodes, thus providing lower cost alternatives compared to fiber-coupled bars, and potentially higher available powers than fiber-coupled methods. Experiments have demonstrated the use of diode bars to directly launch more than 1 kW into a fiber end.²⁷

Though end-pumped schemes are capable of providing kilowatt-level incident pump powers, other methods allowing more uniform distribution of pump light along the length of a fiber have been investigated. One approach uses various techniques to inject pump light along the length of a fiber by either microprisms or V-groves etched into the side of the fiber.^{54,55} This approach allows pump power to be distributed along a very large length of fiber and requires minimal alignment, compared to free-space end-pumping techniques, but its implementation is more difficult and time consuming with the need for multiple etches or other fiber preparation. Moreover, in the event of fiber damage, the entire array must be replaced, rather than simply the fiber as is the case in end-pumped configurations.

Directly side pumping a fiber core has also been attempted, however, obtaining sufficient absorption over a few micron fiber core is challenging. Some fibers have been created with flattened sides allowing them to be coiled tightly in a spiral, thus permitting pump light from diode bars to enter from the side and pass through multiple fiber cores in the spiral, allowing for higher efficiency pumping with relatively simple to use diode bars.²³ While allowing minimal alignment, simple packaging and easy thermal management, difficulties with this technique arise in manufacture, assembly, and repair of such fibers.

Free-space end-pumping schemes work very well but have the disadvantage of being limited to only being able to pump the two ends of the fiber, limiting the amount of pump power to that available in a single diode bar, stack (or two in polarization combined cases), or fiber-coupled device. In addition, launching very high pump power into small fiber ends may cause thermal damage to the fiber. To avoid this, fibers may need to be end capped or have actively cooled ends, especially when using nonsilica glass fibers which have lower melting points.

Monolithic or "all-fiber" fiber lasers best realize fiber's potential advantages over bulk solid-state lasers. There are several main techniques to achieve monolithic pumping, each with particular advantages and disadvantages. Telecommunication lasers utilize waveguide-based "Y" couplers or wavelength multiplexers to achieve pump and signal combination; however, these devices are optimized for low powers and for operation of core-pumped lasers with single-mode pump components not capable of the power levels needed for high-power systems. Alternative techniques must be used in high-power double-clad fibers.

The most straightforward technique for coupling pump light is simply splicing a single pump fiber to the end of a gain fiber (of course with a fiber Bragg reflector in between to form a resonator). This method is low loss and very simple to implement, however, the main downfall of this technique is that only one fiber can be spliced to the laser, and hence power delivery is limited to the power available in a reasonably small (100 to 800 μm) delivery fiber. To achieve higher pump powers, devices with multiple input ports are required. One of the most common current methods for launching pump light from multiple sources is by way of tapered fiber bundle (TFB). TFB's are simple in concept as seen in Fig. 3, involving bundling a number of undoped, coreless pump delivery fibers around a central fiber (possibly with a core in amplifier configurations), fusing the whole bundle together with high heat, and tapering it down to an appropriate diameter to allow it to be spliced into a system.⁵⁶⁻⁵⁸

Several methods for TFB design and manufacture and their evolution in time are pointed out in Refs. 56 to 58. TFBs are capable of handling upward of 1 kW of optical power with the addition of proper thermal management techniques as demonstrated in Ref. 59.

A concept similar to TFBs, the nonfused fiber coupler, involves angle polishing a fiber to an appropriate angle and butting it to a flattened portion of a second fiber.⁶⁰ The benefit of this technique is that no fusion splicing is required; the fibers must only be polished and butted together. Again the concept of such fibers is simple and the number of inputs is scalable and power can be distributed along the whole fiber length, however, the implementation is difficult as precise angle polishing is required.

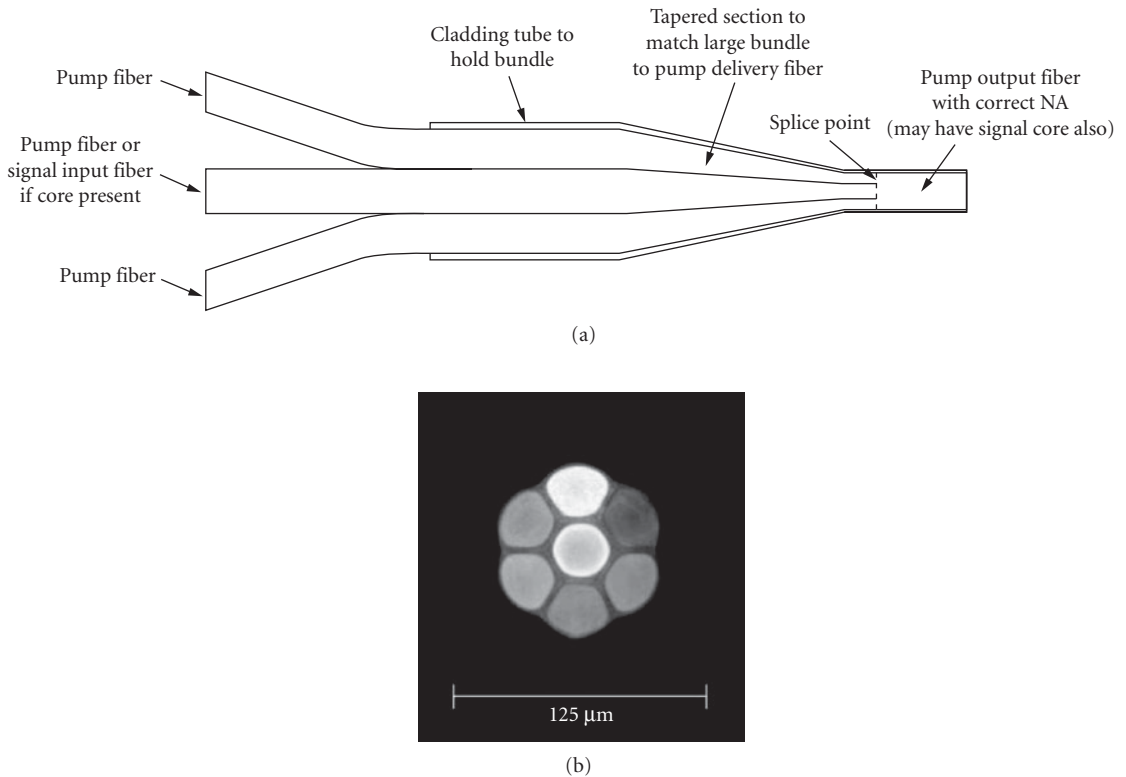


FIGURE 3 (a) Schematic of tapered fiber bundle. (b) End view of actual tapered fiber bundle.

TFBs are excellent options for most free-space fiber systems. However, in many cases lasers need to be very short, or of soft glass materials that cannot be spliced to silica glass TFBs efficiently. It may also be desirable to distribute pump light over the length of the fiber more uniformly to reduce “spot heating” effects. Thus an alternative all-fiber technique is evanescent coupling, also called GTWave technology. This technique consists of placing two fibers in optical contact by either stripping polymer layers and manufacturing by hand, or by pulling two fibers together into optical contact and covering them in one polymer coating.^{61–63} This technique’s ability to work for short, highly doped fibers, or distribute pump light very evenly makes it an attractive technique in some applications.^{62,64,65} The method also does not require any special equipment, only bare fibers and some method to hold them together such as heat-shrink tubing (though custom fiber drawing techniques can also be used), in contrast to the splicers and cleavers needed to make and implement TFBs. Evanescent coupling can also work in situations where TFBs cannot be made or purchased, such as with soft glass fibers or very large mode area fibers. In addition when fabricated as all-fiber systems pulled together on a draw tower this technology is used in several high-power commercial systems.⁶⁵

As development of high-power fiber lasers continues, even higher-power handling “all-fiber” components will become available and this will result in the monolithic fiber laser system increasing in power and beginning to function at power levels comparable to high-power free-space systems currently available. Currently the development of pump-combining devices and techniques continues as new methods are developed or existing methods improved or combined together to provide greater power handling, efficiency, and reliability of pump components. This section is only a brief introduction to the basic categories of pump components and techniques.

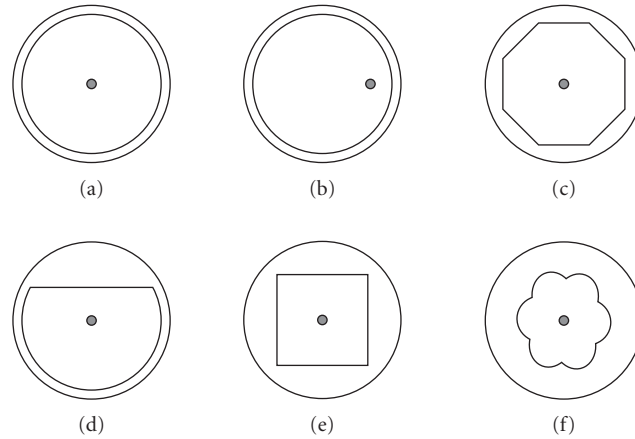


FIGURE 4 Different fiber core geometries for enhanced pump absorption: (a) unmodified fiber; (b) offset core fiber; (c) octagonal (or otherwise polygonal) fiber; (d) “D” shaped fiber; (e) square or rectangular fiber; and (f) “flower” shaped fiber.⁶⁶

Pump Cladding Design Pumping in double clad optical fibers relies on the absorption of pump light by a doped core as it propagates in an outer cladding. However, the circular symmetry of conventional fiber poses a challenge to pump absorption due to excitation of helical pump modes that allow light to propagate down the fiber without crossing through the core. In the double clad pumping structure this is an issue because pump light will never have the opportunity to be absorbed by the doped core. The amount of pump light in these higher-order modes is significant and must be dealt with for efficient laser operation. The simplest technique is to coil the fiber though many fiber types cannot be bent to sufficiently small diameters. A polarization maintaining (PM) fiber has stress rods built into it which have the effect of breaking the symmetry of the fiber to reduce helical modes in cladding pumped fibers. For particular applications PM fibers are not always available. As a result other methods have been developed.⁶⁶ The most common way to reduce higher-order modes is to build a design into the fiber which breaks the circular symmetry by using any number of different cladding shapes such as shown in Fig. 4.

While all these designs are effective in causing higher pump absorption, some have more useful features than others as noted in Ref. 66. Overall, choice of shape is a matter of preference and exactly what purpose the fiber laser will have. It involves making compromises between effectiveness of design, ease of fabrication, and ability of the fiber to be spliced to conventional round fibers. It is, nevertheless, an important choice in the final design of a high-power fiber laser.

Choice of Diode Pump Choice of the appropriate type of diode pump for a particular fiber laser application is critical to optimizing the performance of the laser system. There are two choices to consider when determining what type of diode to use in the construction of a fiber laser. One can use single emitters and combine them to reach the desired powers using pump combiners or other techniques outlined earlier or one can use diode bars or stacks of bars for pumping. A more detailed discussion of this topic can be found in Ref. 59. Diode bars tend to simplify a system, as they are available packaged and deliver high powers, while a larger number of single emitters must first be spliced together via a tapered fiber bundle or other method before high powers can be achieved. Single emitters may be lower in cost per watt and also may be more efficient, meaning lower cooling loads and often need only air cooling. However, the cost savings of single emitter units versus bars may be offset by the time and effort required to splice and wire individual single emitters together.⁵⁹ Fiber-coupled bars tend to be less sensitive to thermal fluctuations causing wavelength shifts, with bars wavelength only fluctuating a few nanometers over temperatures, while single emitters can change by 10 to 20 nm.⁵⁹ Optical damage can occur in single emitters since the fiber pigtailed to them leads directly to the diode facet. Because of

the free-space optics inherent in fiber-coupled bars it is more difficult for stray back-reflections to cause damage and easier to integrate dichroic elements to mitigate the problem.⁵⁹ Single emitters have the advantage of being able to be electrically modulated far more rapidly than diode bars and so are superior in systems where pump diodes must be pulsed, especially at high repetition rates. Single emitters also tend to offer slightly longer lifetimes than diode bars.⁵⁹ Overall, there is no clear choice between diode bars and single emitters. The choice depends on the particular laser application and is usually an engineering decision that weighs all factors involved in using a particular technology.

Free-Space Fiber Laser Designs

Free-space fiber systems can take two basic forms: resonators and amplifiers. Laser resonators are common when moderate- to high-power systems are desired with the maximum simplicity and compactness. High-power resonators offer the ability to use a minimal number of components to achieve desired laser performance. However, they are often limited by difficulties in achieving high powers in particular spectral or temporal modalities. When an oscillator cannot fulfill a particular need, multistage amplifier systems with low-power master oscillators, providing the pulse or CW linewidth characteristics, and one or several power amplifier stages, providing the energy, are turned to. Known as master oscillator power amplifiers (MOPAs), such systems have higher potential for performance but are more complex and expensive to construct due to the need for additional components and complexities. The basic architectures of both types of systems are discussed in the following sections.

Free-Space Oscillator Architectures The most basic of all free-space cavities can be formed by use of a diode pump source and polished or cleaved fiber.^{36,37} The resonator is formed between approximately 4 percent Fresnel reflections (in the case of silica fibers with refractive index of ~ 1.45) of the two cleaved or polished fiber facets. The high gain of the fiber laser allows it to overcome very high losses and oscillate despite the small amount of feedback. (In fact, one of the main difficulties of making high-power fiber amplifiers is keeping the fibers from “parasitically lasing” since with high enough pump powers even angled fiber ends can give enough feedback to promote oscillation.) This simple cavity is not particularly suited for many applications due to its low efficiency, high threshold, and laser emission being split almost evenly between both ends. However, this configuration is useful for simply testing the free-running characteristics of prototype fibers and is also useful for providing initial alignment of a more complicated cavity.

The next step in sophistication of fiber laser resonators involves placing a mirror at one end of the cavity as shown in several forms in Fig. 5.^{36,37}

The simplest form of the single mirror resonator; Fig. 5a uses a dichroic mirror; a mirror that is highly reflective at the laser wavelength and highly transmitting at the pump wavelength to enhance the cavity Q, lower laser threshold, and provide output from only one end of the resonator. This technique works well though at high power there is a potential for mirror damage due to the high intensity incident on the mirror.

Damage to mirrors can be avoided by separating them from the end of the fiber and placing a lens in between the mirror and fiber facet as in Fig. 5b and 5c. The scheme in Fig. 5b is clearly the simpler of the two schemes with pump light and laser light both traveling through the same lens. For many fiber lasers systems, especially those based on erbium or thulium where the pump and laser wavelength are significantly separated, chromatic aberration in the lens causes the pump and signal to focus at different points. As a result it is difficult to perfectly align the laser beam to complete the resonator in Fig. 5b and efficiently couple the pump light with a single lens (unless the pump delivery fiber is significantly smaller in diameter than the double clad laser fiber). Cavities based on Fig. 5c are also useful when adding additional elements to the resonator such as Q-switches. It should be noted that fiber lasers usually can operate with only Fresnel feedback from their output ends; however, sometimes a higher reflectivity output coupler must be added to the resonator to provide sufficient feedback to allow laser oscillation. Other more exotic and less common resonators for fiber lasers are free-space ring type resonators which are often used in mode-locked fiber lasers as in Refs. 67 and 68. Rings are often more difficult to keep stably aligned and require additional elements such as isolators and wave plates to ensure unidirectional operation.

Basic Free-Space MOPA System Designs Amplifiers for high-power lasers differ greatly in concept and design from those intended for telecommunications applications. Communications amplifiers

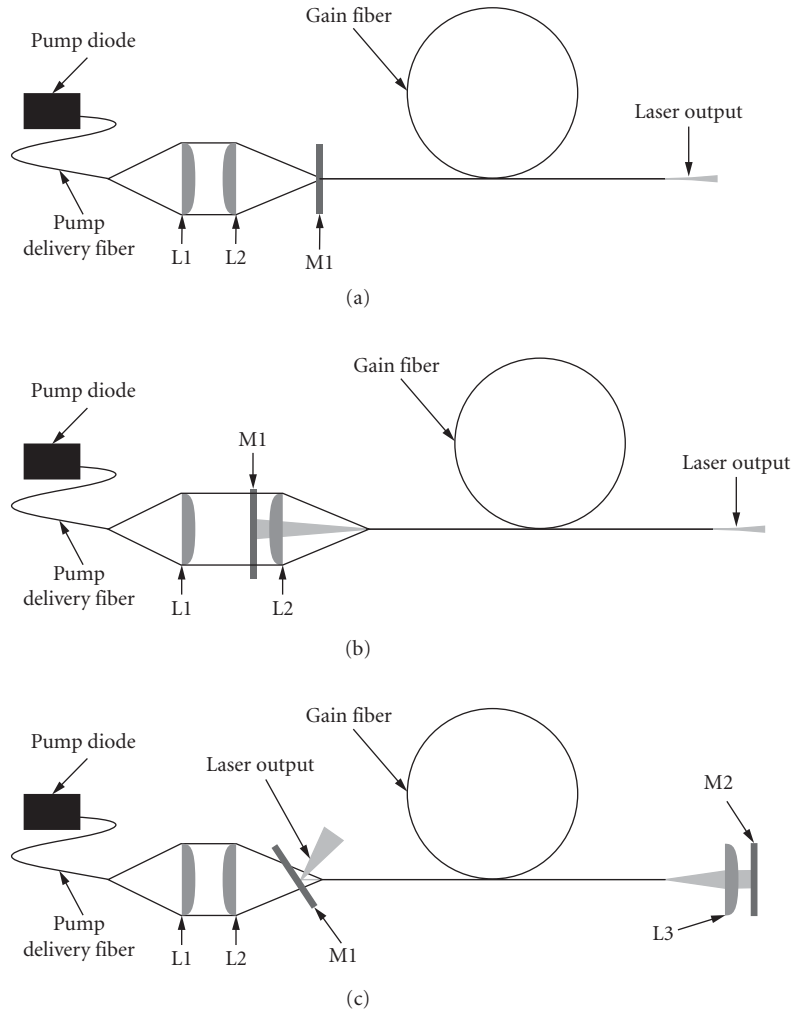


FIGURE 5 Note in all images L1 and L2 are pump coupling optics, L3 is a collimating lens for the signal, and M1 and M2 are dichroic mirrors. (a) Simple butted mirror cavity. (b) External resonator for lasers with pump and signal close in wavelength, mirror outside cavity provided feedback through lens. (c) External feedback cavity on opposite end from pump to compensate for large wavelength difference between pump and signal. Bulk cavity elements like Q-switches elements can be placed between L3 and M2.

operate in the small-signal regime, amplifying small signals from noise with high fidelity and thus demand very high gains. High-power lasers demand efficiency from amplifiers with a desire for maximum extraction of energy. Master oscillator or power amplifier designs (MOPA) are often used to avoid difficulties associated with the maintenance of particular laser parameters in resonators at high powers (i.e., narrow linewidths or short pulse durations).

There are two basic configurations for free-space MOPAs designated as copropagating and counter-propagating MOPAs. These two schemes should produce identical results as they both involve a fixed gain per unit length for an input signal. In practice, however, the two schemes are different in their efficiency at high power and their ASE characteristics. The essential difference between the two

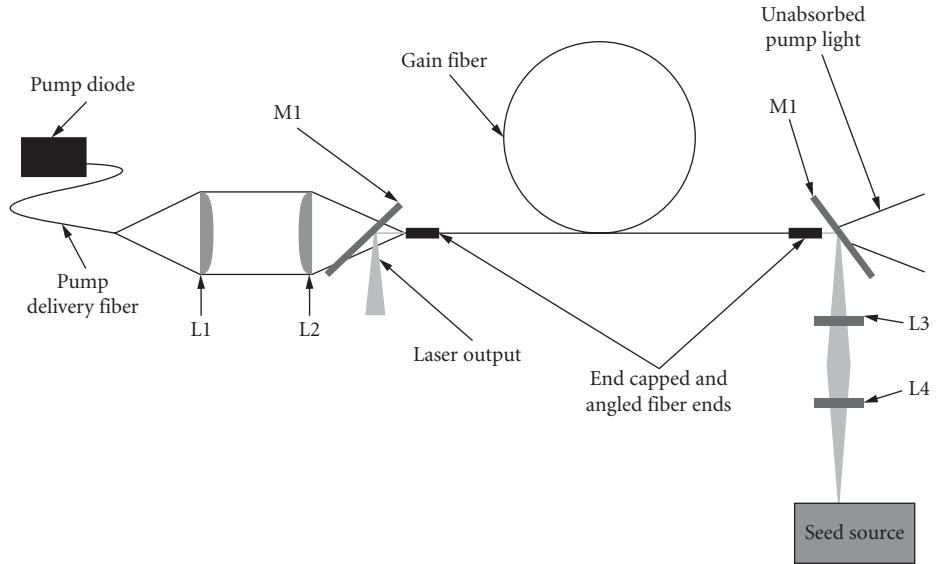


FIGURE 6 Counter-propagating MOPA scheme. Lenses L1 and L2 are matched pump delivery lenses. M1 is a dichroic mirror and L3 and L4 are signal delivery lenses.

schemes is the direction of the pump light relative to the signal light. The counter-propagating MOPA scheme is the most commonly used configuration in high-power amplifiers as it has higher gain when operating in saturation compared to a copropagating scheme.^{10,45,69} A counter-propagating MOPA is shown in Fig. 6.

Aside from the direction of pump propagation relative to the signal the counter-propagating MOPA is identical in construction to a copropagating MOPA. Copropagating MOPAs are usually only used in high-power lasers when one wants to be conservative and protect pump diodes from leaked high-power signals or when one wants improved signal to noise ratio for amplifying very small signals.^{10,45,69} Counter-propagating amplifiers have higher gain and efficiency when generating high powers while copropagating amplifiers, though lower in overall output, tend to have less ASE and better signal to noise ratios on the output direction. Explanations for this phenomenon lie in the fact that gain is not uniform along a fiber laser and hence a signal sees higher gain either earlier or later in its propagation (depending on pump direction) as discussed qualitatively and analytically in Refs. 10, 45, and 69. Amplifier gain and output power can be increased by setting up either a bidirectionally pumped amplifier or a double-passed amplifier. Unfolding the MOPA through the mirror shows that both cases are essentially identical (though bidirectional amplifiers allow more pump power due to double end pumping), and both cases offer the combination of benefits and issues associated with counter- and copropagating schemes.^{10,45,69}

Angle cleaving or polishing is a critical consideration in fiber amplifier systems due to the fact that fibers have such high gain that they can lase, even without mirrors, based solely on Fresnel reflection feedback. As a consequence one or both ends of any fiber intended for MOPA use must be angled to prevent oscillations. Even with this angling, oscillations may still occur at very high pumping powers. Techniques beyond angle cleaving for reducing these oscillations can become quite complex.⁷⁰

Often one amplifier stage for the low-power seed is insufficient to reach the desired output powers since high-power MOPAs require reasonably powerful seeds to fully saturate leading to efficient energy extraction. As a consequence these systems require many stages of preamplification before the final power stage is reached. Often between stages there will also be devices to act as pulse pickers or gates to reduce ASE between pulses. There will also be optical isolators to keep reflected pulse energy and ASE from moving backward in the system, stealing system gain or reaching significantly high powers to cause optical damage to amplifier components.

All-Fiber Monolithic Systems

For fiber lasers to reach their full potential in terms of stability and ease of alignment the laser resonator or amplifier chain must be completely within a fiber. In actuality, no fiber or laser system can be truly monolithic as some components including fiber-coupled diodes, isolators, modulators, and other components require some measure of free-space propagation. However, the nature of these components is that, though not truly monolithic, they have been packaged in such a way that they are extremely stable and are fiber coupled. Here a general overview of the basic components of monolithic fiber lasers and amplifiers is considered.

Monolithic Fiber Laser Resonators The fiber laser resonator in the “all-fiber” configuration consists of “fiberized versions” of the same components as a free-space fiber laser resonator. Such resonators consist of the pump source and a fiber Bragg grating (FBG) spliced to each end of a double clad gain medium. It should be noted that two FBGs are used. One is a high reflector and one is partially reflecting. The latter can be designed to have a very low reflectance to mimic the effects of Fresnel reflection from a fiber end facet. The grating is preferred to directly using the end facet reflection, to avoid degradation of the laser if the end facet becomes contaminated or to provide specific laser performance in terms of linewidth or extraction efficiency by optimizing output coupler parameters.

An alternative form of fiber laser resonator is the ring resonator, consisting of a loop of fiber spliced back onto itself through a so-called tap coupler or splitter. This coupler splits off some portion of the light as a useful beam and allows the rest of the light to remain in the cavity. Pump light is spliced into the ring by way of a TFB or other pump multiplexing device into the gain fiber portion of the ring. Rings may also contain sections of undoped fiber to complete loops, to provide nonlinear effects needed for mode locking, or to provide dispersion compensation. Optical isolators are used to allow unidirectional oscillation. “all-fiber” rings are not commonly used in high-power resonator systems due to the lack of tap couplers and splitters designed for operation with LMA fibers at high powers. On the other hand, the ring configuration is still a useful cavity for seed lasers that can be directly spliced into fiber MOPA systems.

Monolithic MOPA Configurations Monolithic fiber MOPA systems have potential in terms of their ability to extend the stability characteristics of small low-power lasers to higher powers. The fiber MOPA in Fig. 7 is simple, and can be “chained” together as needed to form multiple stages of amplification.

Pump light is injected by TFBs or similar devices in the optimum direction since pump direction influences amplifier performance. Figure 7 shows a three stage system to demonstrate the location where two different pumping directions might be considered (copropagation early on to minimize

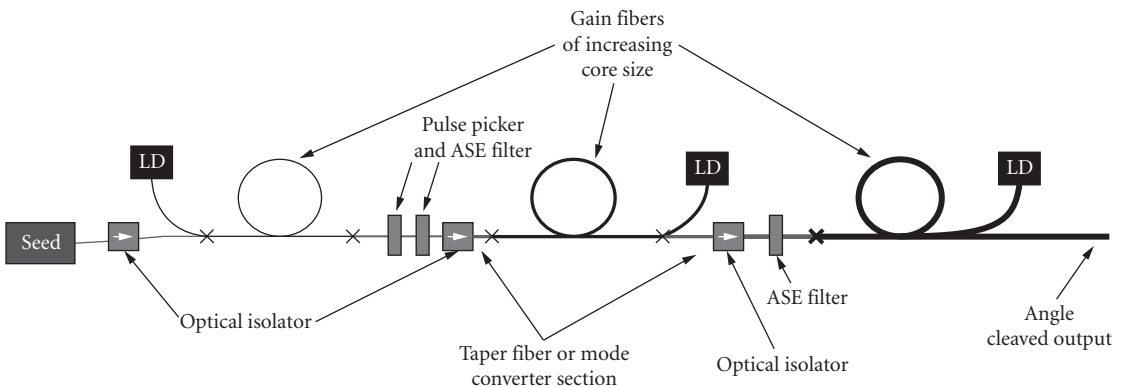


FIGURE 7 All-fiber-spliced MOPA system. LD are varying power pump laser diodes. The seed can be either another fiber laser or simply a fiber coupled laser diode. Sections are fusion spliced together with tapered sections of fiber providing the mode scaling between sections.

forward ASE and counter-propagation in the power stages to be sure of saturation). The only other elements that are included in the MOPA aside from undoped fiber between stages are the in-line isolators, ASE filters, and pulse picker for choosing repetition rate in pulsed systems.

All-fiber MOPAs work well up to reasonably high power levels, however, fibers for handling extreme power levels are not yet practically spliced into MOPAs as often the components such as isolators and pump combiners to deal with such large core sizes are not available.

Mode Field Adaptors The concept of mode matching and expansion can also be used to aid the design of MOPA-based systems. Usually the early stages of a MOPA system comprise small core telecommunications grade fibers while later stages will have larger cores. Sections can be connected together with fiber cores accurately aligned along their centerlines as seen in Fig. 8a. For large jumps in fiber core size a fiber section with a core diameter in between the two being matched may be spliced in as well, the so-called “bridge fiber method.” Often an appropriate bridge fiber is not readily available and thus other techniques have been developed, both similar in their general concept of gradually changing the mode field diameter to match two dissimilar fibers.

One technique involves simply tapering down a section of passive large core fiber matching the LMA section until its core dimensions match that of the smaller core fiber as seen in Fig. 8b.⁵⁶ The second method involves heating a fiber along its length so that the dopants in its core migrate to larger diameters via diffusion forming a mode matched core. This is known as the *thermally expanded core* (TEC) method⁷¹ (Fig. 8c). Choosing the appropriate fiber types by calculating mode field diameter of each using Eq. (7) and altering one or the other to make a reasonable match is the most straightforward way to make MFAs. In some cases, as seen in Eq. (7), the MFD of a fiber is not simply a linear function of the core diameter but rather has a minimum or maximum achievable value. Consequently, a small MFD cannot be increased sufficiently to match a larger MFD or a large MFD cannot be tapered down to a small MFD. A combination of both techniques (or even all three including a bridge fiber) must be

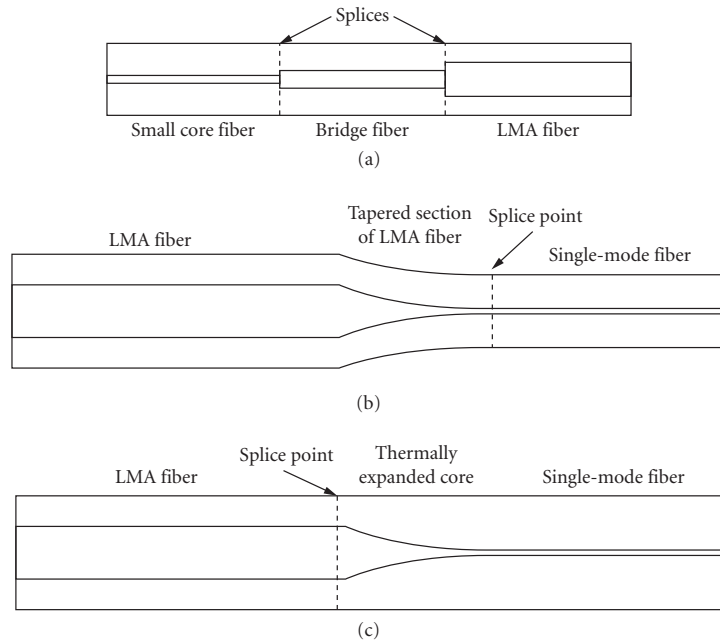


FIGURE 8 Methods for producing mode field adaptors: (a) bridge fiber method; (b) tapered fiber method; and (c) thermally expanded core (TEC) method.

employed to fabricate an appropriate mode field adaptor.⁵⁶ Often such MFAs are included as part of the construction of a TFB used to inject pump light to minimize the number of system components.⁵⁶

Fiber Bragg Gratings Fiber Bragg gratings are Bragg stacked layers of varying index of refraction written directly into the core of an optical fiber. Their operating principle is the same as any dielectric mirror utilizing a stack of quarter-wave layers as discussed in Chap. 17, “Fiber Bragg Gratings” and Ref. 72. Currently FBGs for large-mode area fibers are becoming available and several methods for their manufacture have been investigated including using photosensitive glass and writing holographically⁷² or by writing using direct inscription with femtosecond laser pulses.^{73–75}

Fusion Splicing The key advantage of an all-fiber-based laser system is the fact that the fibers are permanently bonded or “fused” together eliminating the potential for misalignment. Two fiber tips are microaligned in close proximity using a camera or other user-operated or automated feedback systems. The fiber ends are then heated rapidly with an arc discharge or electrical filament to melt the glass and fuse the two ends together. This method allows for minimal losses (0.1 dB) while giving high mechanical strength and stability.⁷⁶ Such a splicing technique was used many years ago for telecommunication fibers but only recently have these techniques been devised to handle LMA fibers requiring increased fusing temperatures and facet uniformity. New LMA splicers coming onto the market are one of the main reasons for the advance of LMA fiber technology as a whole. For a complete treatment of fusion splicing and fiber cleaving the reader is referred to Ref. 76.

25.6 LMA Fiber Designs

The desire for higher power or pulse energy fiber lasers requires the increase of core diameter in order to avoid damage and unwanted nonlinear effects. A main advantage of fiber lasers lies in their inherent beam quality stemming from single transverse mode operation. As a result methods for creating large mode area fibers while preserving beam quality are sought.

Figure 9 shows a schematic of different techniques for achieving LMA single-mode operation and will be referred to throughout this section.

Conventional LMA Techniques

Conventional LMA techniques utilize relatively standard fiber designs to enable high beam quality and only require low-core numerical apertures to keep the guidance of higher-order modes weak so that they can be stripped by introducing preferential losses. The weaker guidance causes higher bend loss for the fundamental mode and very low fiber NAs leads to fabrication challenges.^{19,77} Using the weak guidance concept, several techniques for LMA fibers have been developed.

Early LMA Fibers and Dopant Profiling The first designs for LMA fibers were used in Refs. 18, 77, and 78 to achieve high powers from Q-switched lasers. To improve beam quality in addition to using very small NAs the fiber core was given a tailored dopant and index profile to help provide preferential gain to the lowest-order mode.^{18,78} A theoretical study of tailored gain is given in Ref. 79. Using this technique the core diameter was extended to approximately 21 μm and output powers of up to 0.5 mJ at 500 kHz were achieved with M^2 of approximately 1.3.¹⁸

Coiling and Weak NA Guidance The technique of fiber coiling, first used in Ref. 80, takes advantage of the added bend losses in a weakly guided LMA fiber. Losses in higher-order modes increase with bend radius due to their weaker guidance in low NA fibers and by using tight fiber coiling, higher-order modes can be stripped out with minimal expense in power loss to the fundamental mode. Such fibers’ index profiles resemble those in Fig. 9a, which is essentially a standard fiber design with very low NA (0.1). The theory of how bend loss in LMA fibers has been treated in Refs. 81 and 84, but in general optimal bending is determined experimentally.

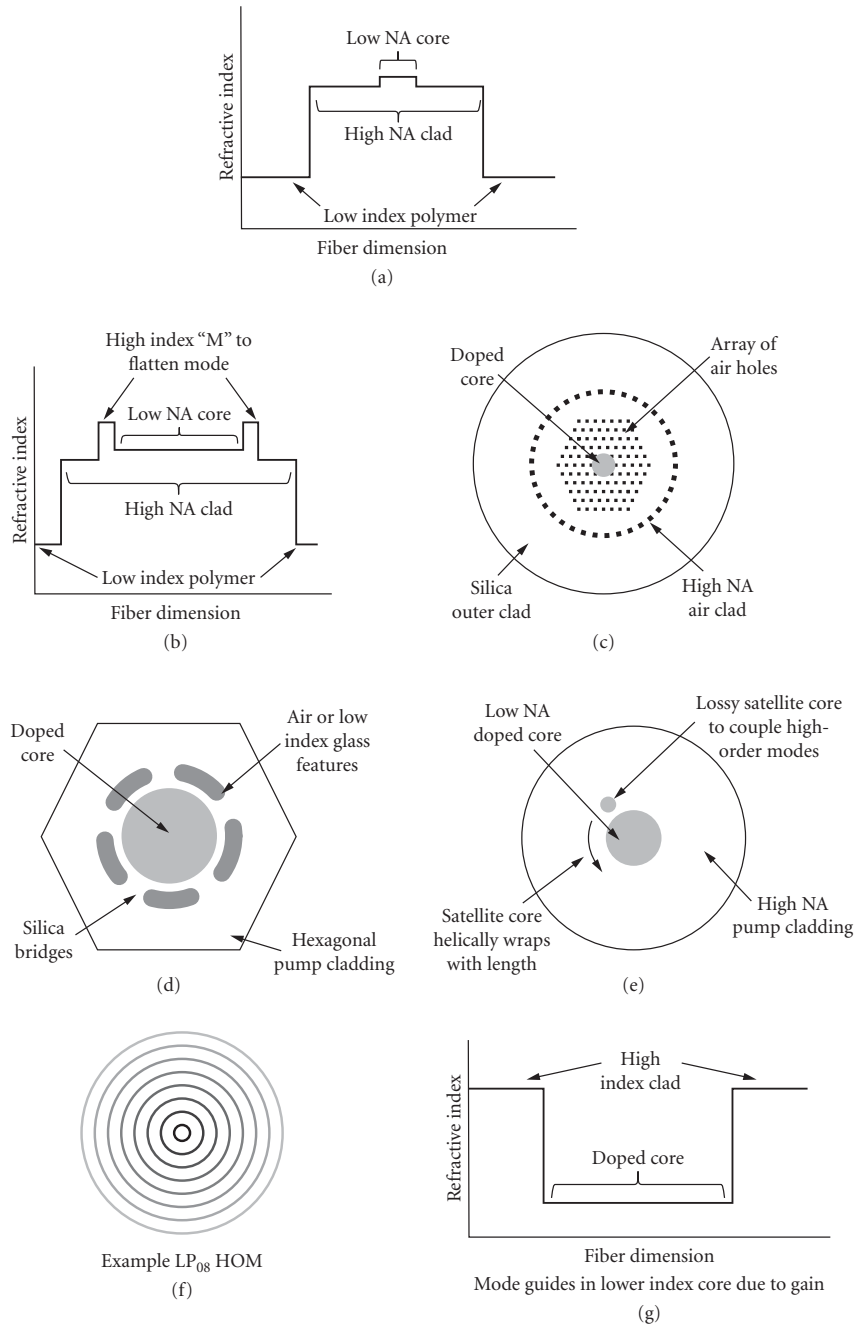


FIGURE 9 Sketches of various types of fiber designs discussed throughout this section: (a) conventional fiber design; (b) large flattened mode or “Batman” fiber; (c) typical photonic crystal fiber design; (d) leakage channel fibers; (e) chirally coupled core fiber; (f) example of the radiation pattern of a higher-order mode; and (g) index profile of a gain-guided index antiguided fiber.

Large Flat-Mode Fibers An extension of the use of specialized index profiles to enhance the performance of LMA type fibers involves the use of “M” shaped (or “Batman”) doping, whereby a slightly higher index step is placed on either side of the fiber core (Fig. 9b).^{85,86} This addition causes “flattening” of the mode field diameter and thus a larger but more flat topped fundamental mode is achieved. The mode area is increased (by up to a factor of 2 or 3) while maintaining near single-mode operation. This design described by Refs. 85 and 86 has been successfully used in several cases, mostly for achieving high peak power ultrashort pulses. By implementing these cores, the threshold for nonlinear interactions in the fiber core was increased by a factor of 2.5.⁸⁵

Single-Mode Excitation A technique that is related to conventional techniques is the single-mode excitation technique (SMET). This method, outlined by Refs. 87, 88, and 89, is most applicable to high-power fiber-based free-space MOPA systems. The mode field diameter of a small-core seed fiber is matched perfectly to that of the fundamental mode of a large-core fiber by using specially selected lenses and careful alignment. Using care, the fundamental mode can then be exclusively launched in a fiber and using this technique fibers with cores as large as 300 μm with mode field diameters (at $1/e^2$ of the peak on axis intensity) as large as 195 μm have been excited⁸⁹ (though not made into lasers) and fiber MOPAs with cores of upwards of 80 μm have been achieved.⁹⁰ SMET can be a very effective technique and has been used in many high peak power fiber lasers including in Refs. 90, 91, and 92.

Inclusion of Single-Mode Sections via Tapering Tapered fiber sections can be used to limit the propagation of higher-order modes in conventional type fibers. A tapered section fiber was demonstrated in Ref. 93, where a small section of multimode fiber was tapered to a diameter that allowed it to become single mode. When inserted into an oscillator (at the output end), the section improved the beam quality, reducing M^2 from 2.6 to 1.4; however, the slope efficiency was reduced from 85 to 67 percent.⁹³ Distributing the mode filtering over the length of the fiber (by such techniques as coiling) is more effective since loss is better distributed.^{93,94} This approach has allowed powers of up to 100 W with near perfect M^2 by using a 27- μm diameter core with an 834- μm clad diameter fiber and tapering it gradually down by factor of 4.8 over approximately 10 m of length.⁹⁵

Photonic Crystal Fibers

Limitations of precision in fabrication of conventional step-index fiber preforms leads to difficulties in obtaining fiber core numerical apertures less than approximately 0.06.^{19,96} In response, a class of fibers called photonic crystal fibers (PCF) has been developed.^{97,98} These fibers utilize an organized structure of refractive index differences (air holes or different high- or low-index glasses) to either tailor the core numerical aperture to sufficiently small values or to create a photonic bandgap where only certain wavelengths of light are allowed to propagate. Versions of these fibers have been developed for their beneficial nonlinear, dispersive, and loss properties in studies since the mid-1990s.^{97,99–101} The first fiber lasers based on PCFs were demonstrated in the early 2000s and had core sizes around 10 to 15 μm .^{102,103} Current state of the art in PCF technology allows core sizes of up to 100 μm ^{20,104} with powers of up to 4.3 mJ pulsed, corresponding to 4.5 MW peak power in approximately nanosecond pulses.

Air-Filled or Holey Fibers Air-filled PCFs are the most common and first demonstrated PCFs for fiber lasers.^{97,103} The air holes function to alter the NA giving it the desired value to maintain single-mode guidance. Figure 9c shows a sketch of a typical PCF design. The V parameter is still the limiting factor in the guidance of single mode beams in a PCF and its value is approximately calculated by

$$V_{\text{eff}} = \frac{2\pi}{\lambda} \Lambda F^{1/2} (n_0^2 - n_a^2)^{1/2} \quad (8)$$

where V_{eff} is the effective V parameter (still < 2.405 for single-mode propagation), Λ is the spacing of the air holes (pitch), F is the filling factor of air to glass, n_0 is the index of the glass, and n_a is the

index of the air (or whatever material fills the holes).¹⁰⁵ This is only an approximate model, giving rule of thumb results, and Ref. 106 covers PCF analysis in further detail. Thus control of the pitch and hole size of the PCF structure allow production of endlessly single mode structures.¹⁰⁷

As most PCFs manufactured for high-power laser purposes are double clad designs, they must also be designed with cladding pumping in mind. Since the largest core PCFs must not be bent sharply they must be able to absorb pump power in relatively short lengths.¹⁰⁸ To achieve this with reasonable doping levels in the core the core-to-cladding ratio must be kept as large as possible which is usually achieved by using an air cladding. The *air cladding* is simply a ring of very thin silica “bridges” or alternatively very large air holes with small connections to the outer glass fiber.¹⁰⁸ Using this technique, air claddings with NAs as large as 0.8 have been realized.¹⁰⁹ The one detrimental effect of the air cladding is that it does not allow for efficient thermal conduction of heat generated in the fiber core out to the cladding. An analysis of heat transport in PCFs is given in Ref. 110. Despite this, the highest average power from a rodlike air clad PCF reported to date is 320 W¹¹¹ and 1.53 kW from a longer, coilable PCF.¹¹² The highest pulse energy is 4.3 mJ.²⁰ A polarization maintaining PCF with 2300 μm^2 mode area was also demonstrated to have 161-W output power in a single polarization.¹¹³

All Solid-Core PCFs The main issue with the use of PCFs, especially in their adoption to all fiber systems lies in their air holes. It is challenging and complicated to form and fabricate a preform and fiber with air holes¹¹⁴ (though extrusion techniques may make this process simpler for instance in Refs. 115 and 116). It is also difficult to splice and cleave a fiber with air holes. There are further difficulties associated with (1) contamination presented by particles working their way into the air holes and (2) finding fiber-based components compatible with PCFs.¹⁰⁴ All-solid PCF fibers simply replace the air holes with a low or high index material to avoid the air-hole issues. In addition the solid defects can be engineered to provide a photonic bandgap which is a more complex way of guidance providing single-mode performance with the potential additional benefits of polarization maintenance and dispersion management. Novel uses for bandgap PCFs such as low-dispersion femtosecond lasers and 900-nm Nd lasers (using a bandgap to suppress 1064 nm operation) have been constructed^{117,118} and other lasers have also been proposed.¹¹⁹

Leakage Channel Fibers So-called leakage channel fibers^{120–122} sketched in Fig. 9d, employ low index of refraction regions surrounding a large core with “bridge regions” of the same index as core glass connecting core and cladding. The result is preferential leakage of the higher-order fiber modes providing more robust lowest-order single-mode operation. The early types of these fibers resembled PCFs as they used air-filled regions. Recently all solid designs have been demonstrated with core diameters as large as 170 μm .¹²⁰ Advantages of such fibers include the ability to be effectively cleaved, spliced, and bent. However, they are more challenging to manufacture than conventional fibers.

Chirally Coupled Core Fibers

A new type of LMA fiber laser is the chirally coupled core (CCC) fiber (Fig. 9e), which utilizes specially designed satellite cores helically wrapped around the central LMA gain core to couple out the higher-order modes into the lossy satellite while maintaining the lowest order mode in the core.¹²³ Designing the helical core or multiple cores with proper pitch and size leads to strong coupling of high-order modes to the intentionally lossy satellite.¹²³ The losses in individual modes can be calculated and optimized for a particular helix pitch and size. Several CCC fibers have been fabricated and tested in two common wavelength regimes, 1060 and 1550 nm exhibiting excellent beam qualities for inner core diameters as large as 50 μm . CCC fibers are attractive also because they do not require small core NAs, can handle strong bending, and can be used for very long lengths to achieve high CW powers with less thermal load. In fabrication CCC fibers require the preform core to be formed by standard modified chemical vapor deposition (MCVD), but a hole must be bored for insertion of the satellite rod. Then during drawing the preform must be spun at a set speed to achieve the chirality of the satellite at the desired pitch. An approximately 40-W fiber laser from a 33- μm core CCC fiber with very large V parameter but high beam quality has been demonstrated at 1064 nm.¹²⁴

Multicore Fibers Fiber lasers with multiple doped cores within one central pump cladding have received some attention in recent years due to their potential to scale mode area by allowing the addition of light from several individual cores. With proper design of the fiber, the fiber laser can have high beam quality in the far field.^{125,126} Multicore fibers are difficult to manufacture due to the need for drilling preforms and adding several cores.

Higher-Order Mode Fibers Another newly emerging class of LMA fiber designs is the higher-order mode fiber (HOM). These rely on using mode-conversion techniques usually based on long period fiber Bragg gratings.^{127,128} HOM beams have Bessel function spatial distribution.¹²⁹ The gratings are specifically designed so that they efficiently couple energy to only one HOM and are then used in tandem to subsequently de-excite the HOM back in to a more useful LP_{01} mode at the fiber output. (though leaving the beam in a HOM may also be useful). Typically the LP_{0X} mode (where X is an integer >1) is excited. A sketch of such a mode is seen in Fig. 9f.

Modes as high as LP_{08} have been launched and through this process mode areas of approximately $3200 \mu\text{m}^2$ achieved.¹²⁸ A review of the theory is described in Ref. 127 in terms of their use in cladding modes of fibers, where this same theory applies to launching of HOMs into large-core multimode fibers as well. An experimental investigation into launching HOMs in custom large core fibers is reported in Ref. 128. Currently HOMs have only been used once in an actual laser resonator;¹³⁰ however, HOM-based modules have also been exploited for their ability to produce anomalous dispersion in silica fiber allowing generation of pulses as short as 60 fs in Yb fiber lasers.^{131,132} An investigation of HOMs ability to reduce SBS in fibers has been carried out in Ref. 133 where it was determined that SBS in HOM fibers can be reduced by using higher-order modes.

Gain-Guiding Index Antiguinding Gain-guiding index antiguiding (GG IAG) fibers are another new class of potential LMA technologies first proposed by Siegman in 2003.¹³⁴ These fibers consist of low refractive index cores surrounded by higher refractive index claddings, hence their so-called antiguiding nature. The core does not support conventionally index-guided modes. Instead it does support these modes in their “leaky” form. These the modes exist in the core, but constantly leak energy to the cladding if the fiber core is not excited. These modes can be thought of as existing due to Fresnel reflections at the core-cladding interface. Though these reflections can be low loss they can never be lossless as are total internal reflections. When the fiber is pumped gain in the core can compensate for the loss in the leaky modes resulting in lossless modes confined in the core. To maintain single-mode operation, gain must be supplied such that it makes up for the loss in the lowest-order mode but not for higher-order modes. The basic theory for GG IAG fibers is laid out in two papers by Siegman.^{134,135} Based on this principle laser oscillators can be designed with their gain and oscillation threshold conditions optimized to attain single-mode operation by using analysis given in Refs. 136 to 138. The first gain guided fiber lasers were demonstrated in flashlamp pumped cavities with core sizes from 100 to $400 \mu\text{m}$ ^{136,137} and M^2 less than 1.5. Diode pumping through the fiber end was subsequently demonstrated.¹³⁸ The mode areas reported by these papers are the largest reported in any fiber laser; however, efficiency and output power scaling of GG IAG fibers using new techniques and pumping schemes must still be addressed before the technology can become competitive with other LMA technologies. Gain-guiding effects have also been seen in conventional fibers. Recently a dramatic change in the guidance properties in highly doped optical fibers was reported at very high pump powers where the gain and refractive index changed due to very strong pumping.¹³⁹

25.7 ACTIVE FIBER DOPANTS

Apart from Raman fiber lasers which are not part of this review, all high-power fiber lasers utilize rare earth ion dopants. Though almost every rare earth ion has been doped into fibers for fiber laser applications high-power diode pump sources are not yet available for some. As seen in Table 1, almost every rare earth ion has been doped into fibers for fiber laser applications; however, high-power diode pump sources are not yet available for some dopants.

TABLE 1 Operating Range, Dopant Ion, and Transitions of Various Rare Earth Ions Fabricated in Fiber Lasers¹⁴⁰

Operating Range (nm)	Dopant (km)	Transition	Type of Host		Type of Transition
			Oxide	Fluoride	
∞ 455	Tm ³⁺	$^1D_1 \rightarrow ^3F_4$		Yes	UC, ST
∞ 480	Tm ³⁺	$^1G_4 \rightarrow ^3H_5$	Yes	Yes	UC, 3L
∞ 490	Pr ³⁺	$^3P_0 \rightarrow ^3H_4$		Yes	UC, 3L
∞ 520	Pr ³⁺	$^3P_1 \rightarrow ^3H_5$		Yes	UC, 4L
∞ 550	Ho ³⁺	$^3S_2, ^1F_4 \rightarrow ^5I_1$	No	Yes	UC, 3L
∞ 550	Er ³⁺	$^4S_{1/2} \rightarrow ^4I_{15/2}$	No	Yes	UC, 3L
601–618	Pr ³⁺	$^3P_0 \rightarrow ^3H_6$		Yes	UC, 4L
631–641	Pr ³⁺	$^3P_0 \rightarrow ^3F_2$		Yes	UC, 4L
∞ 651	Sm ¹⁺	$^4G_{1/2} \rightarrow ^6H_{2/2}$	Yes		4L
707–725	Pr ³⁺	$^3P_0 \rightarrow ^3F_4$		Yes	UC, 4L
∞ 753	Ho ³⁺	$^3S_{12} ^3F_4 \rightarrow ^3I_1$	No	Yes	UC, ST?
803–825	Tm ³⁺	$^3H_4 \rightarrow ^3H_6$	No	Yes	3L
∞ 850	Er ³⁺	$^4S_{5/1} \rightarrow ^4I_{15/2}$	No	Yes	4L
880–886	Pr ³⁺	$^3P_1 \rightarrow ^1G_4$		Yes	4L
902–916	Pr ³⁺	$^3P_1 \rightarrow ^1G_4$		Yes	4L
900–950	Nd ³⁺	$^4F_{3/1} \rightarrow ^4I_{4/2}$	Yes		3L
970–1040	Yb ³⁺	$^5F_{3/2} \rightarrow ^5F_{7/2}$	Yes		3L
980–1000	Er ³⁺	$^4I_{11/12} \rightarrow ^4I_{15/2}$	No	Yes	3L
1000–1150	Nd ³⁺	$^4F_{1/1} \rightarrow ^4I_{11/1}$	Yes	Yes	4L
1060–1110	Pr ³⁺	$^1D_2 \rightarrow ^3F_4$	Yes		4L
1260–1350	Pr ³⁺	$^1G_4 \rightarrow ^1H_5$	No	Yes	4L
1320–1400	Nd ³⁺	$^4F_{1/1} \rightarrow ^4I_{1/2}$	Yes	Yes	4L
∞ 1380	Ho ³⁺	$^3H_4 \rightarrow ^3F_4$	?	Yes	4L
1460–1510	Tm ³⁺	$^3D_2 \rightarrow ^3F_4$	No	Yes	ST
∞ 1510	Tm ³⁺	$^5D_2 \rightarrow ^1G_4$		Yes	UC, 4L
1500–100	Er ³⁺	$^4I_{1/4} \rightarrow ^4I_{15/2}$	Yes	Yes	3L
∞ 1660	Er ³⁺	$^2H_{11/2} \rightarrow ^4I_{0/2}$	No	Yes	4L
∞ 1720	Er ³⁺	$^4S_{1/2} \rightarrow ^4I_{0/2}$	No	Yes	4L
1700–2015	Tm ³⁺	$^3F_4 \rightarrow ^3H_6$	Yes	Yes	3L
2040–2080	Ho ³⁺	$^3I_1 \rightarrow ^5I_8$	Yes	Yes	3L
2250–2400	Tm ³⁺	$^3H_4 \rightarrow ^3H_5$	No	Yes	4L
∞ 2700	Er ³⁺	$^4I_{14/2} \rightarrow ^4I_{13/2}$	No	Yes	ST
∞ 2900	Ho ³⁺	$^5I_6 \rightarrow ^5I_7$	No	Yes	ST

3L = three-level; 4L = four-level; UC = up-conversion; ST = apparent self-terminating.

Discussed here are only those dopants, which are now capable of high powers using efficient high power, high brightness AlGaAs, and InGaAs laser pump diodes. These dopant ions include neodymium, erbium, ytterbium, thulium, and holmium. In contrast to these dopants used in a crystalline matrix for bulk solid-state lasers, required for good thermal conductivity, in the glassy hosts used for fibers they have broad absorption and emission spectral linewidths.¹⁴⁰ The broad absorption bands relax the needs for strict control of diode pump wavelength while the broad-emission spectra allow for broad tunability at high powers and the generation of extremely short pulses.

Neodymium-Doped Fibers

Neodymium (Nd) is perhaps the most common laser dopant of bulk solid-state lasers in both crystalline and glass hosts, and it is not surprising that it was the dopant of choice for many early fiber

lasers.^{1,4,12} Its four-level excitation scheme permitted low lasing thresholds necessary in early fiber lasers when glass composition was poor and available pump powers were minimal. The two most common Nd transitions in fiber lasers are the 1.06- μm transition and the 1.3- μm transition, though Nd can also lase directly to the ground state on an approximately 900-nm transition, pump bands for Nd are in the 800-nm range.

The most prevalent transition used in Nd fiber lasers at 1.06 μm has achieved powers in the 300-W range and has been multiplexed into fiber systems to more than 1 kW.^{23,141} Although wavelength tuning over more than 60 nm with watt-level output powers³⁶ has also been shown, Nd is only efficient over 10 to 20 nm around the peak of the 1060-nm emission band.³⁶ As an amplifier Nd has been used at 1.06 and 1.3 μm . The latter wavelength has seen minimal use in high-power lasers due to its larger quantum defect, excited state absorption, and as a consequence of ASE at 900 and 1060 nm stealing gain from the 1.3- μm laser system.^{36,140} In recent years Nd has fallen out of favor for high-power fiber lasers since Yb³⁺ fiber lasers operate in about the same wavelength region and with more efficiency.

Erbium Fiber Lasers

Erbium is well known as the gain medium for most telecommunications amplifiers operating in the 1.5- μm range. It has also been used in high-power fiber laser systems producing output powers as high as 297 W (in Er:Yb codoped lasers).¹⁴² Erbium possesses several potential operating wavelengths, but with available high-power pump sources around 1400 and 980 nm (980 nm is most commonly used when Er is codoped with Yb), only the 1550-nm transitions generate high powers. At these pump wavelengths the absorption cross section is small making a double clad scheme impractical for direct pumping of erbium. To obtain high-power-efficient lasing, erbium is most often sensitized by codoping with Yb, allowing for much higher pump light absorption and, consequently, high power lasing. However, codoping leads to problems, as emission of 1.06- μm light from the Yb ions can take place at high pump powers.¹⁴² The development of high-power diode lasers at 1480 nm may allow Er-doped fibers to be directly pumped with minimal quantum defect to very high power levels and efficiencies akin to those achieved at 1.06 μm with Yb-doped fibers. High-power Er fiber lasers pumped in the 1480-nm region will have many applications because of their relative eye safety. Erbium has a large emission bandwidth with tunability demonstrated from 1533 to 1600 nm.³⁶ Erbium has also been doped in a fluoride host fiber (ZBLAN) giving rise to laser transitions in the 2.7- to 2.9- μm region (silica is not transparent here due to OH⁻ absorption), with powers of approximately 10 W. However, the low melting point of the fluoride glasses limits the potential for very high-power operation.^{143,144} In addition, tunability from 2.7 to 2.83 μm has been achieved with a power more than 2 W.¹⁴⁵ The large quantum defect from its 980-nm pump bands is another challenge though such fiber lasers are still one of the most direct ways to access the 2.7- to 2.9- μm range.

Ytterbium-Doped Fibers

The ytterbium (Yb) ion is by far the most commonly used ion in fiber lasers. Yb worked its way into the fiber laser mainstream for several reasons including its lower quantum defect, the ability to dope high levels of Yb into silica fibers with lower tendency toward concentration quenching, and no excited state absorption or upconversion at longer wavelengths.^{36,47} These benefits outweigh the three-level operation of Yb, which leads to higher laser thresholds than Nd.^{36,47} Yb has only two main energy levels, and is able to lase because these two levels are split in energy due the Stark effect.⁴⁷

The broad (~80 nm) absorption spectrum of Yb is peaked at 915 and 976 nm permitting wide flexibility in the choice of pump source wavelengths. With the low quantum defect there is substantial overlap between emission and absorption. Though the peak of the Yb emission is at 1030 nm, ground state absorption to the upper laser level causes longer fibers to “red-shift” their peak gain wavelength.³⁶ This length-dependent gain limits the tuning potential of Yb lasers. Longer fiber lengths are required to efficiently absorb pump power resulting in larger peak gain wavelength “red-shifts” and narrowing of

the tuning range.³⁶ Tunability has been demonstrated for over 150 nm of bandwidth³⁶ and the minimal quantum defect in Yb has made it the medium of choice for the operation in multi-kilowatts regime.²¹

Thulium-Doped Fiber Lasers

Thulium-doped fiber lasers are of interest because of (eye-safer) emission at approximately 2000 nm. This is desirable at high powers for defense applications and remote sensing. In addition, many applications call specifically for light in the 2- μ m regime including difference frequency generation with 1 μ m to create other IR wavelengths (3 to 5 μ m), difference frequency generation of two closely spaced 2- μ m beams to create terahertz radiation, and light detection and ranging (LIDAR).³⁶ Approximately 1.9- to 2- μ m output also allows thulium (Tm) fiber lasers to be used to pump holmium fiber lasers to reach further into the IR.

Tm is a three-level laser system which terminates on the ground state leading to high laser threshold pump powers similar to those in both Er and Yb. In addition, termination at the ground state makes Tm lasers extremely temperature sensitive. Tm lasers are usually actively cooled for efficient operation. Tm has the potential advantage of having several available pump bands at wavelengths that can be reached by high-power sources. The 1200-nm absorption line was recently explored with direct diode pumping;¹⁴⁶ however, sufficiently high-power diodes at this wavelength are not yet available. Tm lasers can be pumped at 1060 nm, though with very weak absorption. Similar to Er, Tm can also be codoped with Yb;¹⁴⁷ however, this leads to similar 1- μ m emission issues as in the Yb:Er codoped laser and also results in a huge reduction in potential efficiency. The two most commonly used pump wavelengths for Tm are approximately 1550 and 790 nm. The 1550-nm band is pumped by multiplexed Er Yb-codoped fiber lasers. Though this allows for very high power pumping this transition is relatively inefficient. It first requires the construction of a number of Er:Yb fiber lasers. Though pump to signal quantum defect is reasonably low in the Tm laser using Er:Yb laser pumping the overall efficiency suffers due to the Er:Yb lasers. Fiber laser pumping of thulium allows direct core pumping in situations where very short fiber lengths are required. Many commercial systems use cladding or core-pumped schemes with 1550-nm pumping.¹⁴⁸

Another promising scheme for Tm pumping is the use of 790-nm laser diodes. At first glance this process might seem extremely inefficient since there is a large quantum defect between signal and pump (maximum of ~40 percent efficiency). However, there is an additional process that can be exploited. This is called “cross-relaxation” or two-for-one pumping.¹⁴⁹ and allows the transformation of one pump photon into (theoretically) two laser photons. The result is a maximum efficiency of approximately 80 percent.¹⁴⁹ Experimental efficiencies of up to 68 percent (optical to optical power) have been reported¹⁵⁰ with 60 percent easily achieved at high powers.¹⁵¹ The challenge for cross-relaxation pump process is that it is dependent on temperature, dopant-concentration, and fiber composition.^{149,152}

Fiber laser emission from Tm is useful with an extremely large potential emission bandwidth stretching from 1700 nm to beyond 2100 nm.¹⁵³ This wide bandwidth gives Tm potential for application in both ultrashort pulse lasers and highly tunable mid-IR sources. Active tuning of Tm has been demonstrated over 230 nm.³⁶ The fiber length affects tuning range due to the same three-level reabsorption processes discussed for Yb lasers.

Tm fiber lasers have been demonstrated with single-mode CW powers of 268 and 415 W for 790- and 1550-nm pumping, respectively.^{148,151} In addition, an 885-W system with slightly multimode beam quality was also reported.¹⁵⁴ One additional benefit of Tm lasing is that the longer 2- μ m wavelength allows larger core sizes with better beam qualities, lower nonlinearities, and higher damage thresholds since all these properties improve with longer wavelength.

Holmium-Doped Fibers

Ho-doped fibers are one of the only current option to achieve wavelengths longer than 2 μ m. Doped in the proper fiber material (silica loses transparency beyond ~2.1 μ m), holmium (Ho) fiber lasers have reached watt-level output powers.^{155–157} One of the most prominent absorption regions for Ho is around 1.9- to 2- μ m region, where high-power pumping can be provided by Tm fiber lasers.

Tm pumping allows for extremely efficient conversion since the laser wavelength of Ho is approximately 2.1 μm and the quantum defect is minimal. This technique provides a way for stretching the Tm bandwidth to longer wavelengths in the near IR. Other Ho pump bands include the 1160-nm band which can be pumped directly by Yb fiber lasers or as demonstrated by Refs. 156 and 157 with direct laser diodes. In addition, Ho can be sensitized with ions such as Yb or Tm, taking advantage of their broad pump bands. Using this approach a Tm:Ho laser has achieved 83-W output power at 2.1 μm .^{158,159}

25.8 FIBER FABRICATION AND MATERIALS

High-power fiber lasers most commonly employ silica-based fibers because of its strength and thermal stability. However, other materials must be considered for fiber lasers because they offer different merits in terms of transparency, ability to be doped, laser parameters, and manufacturability.

Fiber Fabrication

Fiber fabrication can be divided into two main stages: fiber pulling and preform manufacture. The pulling stage is the actual “fiber making” stage and is common to any preform and fiber material. Preform manufacture is dependent on many fabrication methods. Further details on many of the steps of fiber fabrication can be found in Part 2, “Fabrication,” in Vol. II.

Fiber Pulling The process of fiber pulling is rather similar for all fiber materials with the exception of the operating temperature required, whether a polymer coating is deposited, and the final pulling diameter. This stage of manufacture involves taking a fiber preform and lowering it into an oven at the top of a fiber draw tower. The oven, if set to the proper temperature (varying from $\sim 800^\circ\text{C}$ for soft glasses to $\sim 2000^\circ\text{C}$ for silica) causes the preform to heat and eventually a globule of glass will drop down with a solid glass “string” attached.^{11,140,160} This “string” is the beginning of fiber itself and by wrapping this strand around a mandrel at a constant speed, the diameter of the fiber can be controlled precisely and long lengths of fiber can be formed in large spools.^{11,140,160} Coatings can also be applied during this stage and subsequently hardened (this is an important step for double clad fibers, as low-index polymers allow guidance of the pump light in the fiber cladding).

Preform Manufacture The heart of fiber manufacture and the current and future progress in high-power fiber lasers depend upon the development of “fiber pullable,” defect-free, low-loss preforms. It is one of the main challenges of fiber laser development today.^{11,140,160} The four main preform development techniques are summarized below.

Modified chemical vapor deposition Chemical vapor deposition (CVD) is widely used for the rapid fabrication of large preforms with very accurate index steps and compositions. It is implemented in several approaches, including traditional modified chemical vapor deposition method (MCVD), outside vapor deposition (OVD), and vapor-axial deposition (VAD).^{11,140,160} All three methods involve depositing a soot of chemical oxides onto some kind of rotating silica substrate mounted in a lathe by heating various gasses flowing over the substrate. The soot deposition builds up the desired core and clad layers and the preform tube is collapsed to form the actual preform. OVD applies the soot to the outside of a silica rod, VAD applies the soot to the end of a silica rod which acts as a “seed,” while the most common MCVD introduces the soot to the inside of a silica tube which is subsequently collapsed.^{11,140,160} CVD is extremely flexible in terms of the index profiles and dopants; however, there are challenges in making very large uniform cores for high-power fiber lasers.

Core drilling and preform machining MCVD is an excellent way to obtain radially symmetric doping profiles; however, often extra features that are not radially symmetric are desired such as in polarization-maintaining fibers or fibers with odd-shaped claddings to promote pump absorption.

Holes can be drilled into the preform and subsequently filled with glass or left to be air filled as in PCFs. The preform can also be given flat surfaces to provide the nonuniform pump claddings. In principle the core-drilling technique is quite simple, however, the difficulty with this method is twofold; first, obtaining pullable glass for this type of preform can be challenging (as usually this method is used for nonsilica glasses) and second, drilling holes in glass without breaking it is not a simple task; it takes a good deal of time, care, and skill to make such a preform not to mention the expense.

The benefits of core drilling lie in its flexibility to be feasible with any glass (especially, soft glasses for which MCVD is not available). It also allows the core to be inserted with precise doping characteristics since it is prepared as bulk glass separately with precise control over its composition. In addition, core drilling allows the use of smaller samples of glass which is critical for many types of glasses that are expensive or difficult to make in large quantities. The core index profile is also very uniform since it is one solid piece of glass, hence fibers with very large cores and high dopant concentrations can be more effectively manufactured.

Stack and draw This method is commonly used to manufacture photonic crystal fibers, which require complex structures that would be too expensive or risky to manufacture using hole-drilling techniques.¹⁶¹ The basic premise is to use assorted rods (or cane) and tubes of the desired glass material and stack them in the desired pattern inside a larger glass tube.¹⁶¹ A core region can be added by way of introducing a doped core rod into the pattern of rods or tubes.

This method is useful for PCF manufacture, though, as with the core drilling approach, it depends on finding fiber-pullable glass components that are the appropriate size, particularly, in glasses other than silica. Furthermore, fusing of the rods and tubes together into a solid preform without collapsing the tubes is not trivial and requires much practice.

Extrusion In the extrusion technique glass is pressed through a die to obtain a preform with the desired air holes or glass dopants.¹¹⁶ The technique has a great potential in the realm of PCFs, as arbitrary shapes can be readily generated with a suitable die. A doped core region can be included by simply using two types of glass in an extrusion (not dissimilar to the way different colors are obtained in one tube of toothpaste).¹¹⁶ To date this method has only been used with so-called “soft glasses” having low melting temperatures.

Fiber Materials Though silica fiber is the most dominant material for high-power fiber lasers, several other materials find niches for specific applications. Varying the doped host material may provide benefits in one of three important areas, its laser properties such as upper-state lifetime and emission cross section, its doping properties and potential for higher doping, and finally transparency considerations when operating at mid-IR wavelengths. These different materials and their potential benefits will be discussed in subsequent sections. Table 2 which contains glass data for several glass types will be referred to throughout this section.

TABLE 2 Sample Properties of Different Types of Glasses Used for Fibers^{162–172}

Glass Type	T_x (°C)	Bulk Damage (GW/cm ²)	Thermal Conductivity (W/mK)	dn/dt (10 ⁻⁵ /°C)	CTE (10 ⁻⁷)	Trans. (μm)	Young's Modulus (GPa)	Knoop Hardness	n_2 (10 ⁻²⁰ m ² /W)
Silica	1175	600	1.3	11.9	0.55	0.3–2.1	72	600	3.4
Phosphate	366	25	0.84	-4.5	104	0.4–2	71.23	418	1.2
Germanate	741	—	0.55	1.2	63.4	0.5–3.9	85.77	560	—
Tellurite	482	10	1.25	—	—	0.5–5	54.5	—	30
Chalcogenide	180	6	0.37	9.3	21.4	0.6–8	15.9	109	400
ZBLAN	385	0.025	0.628	-14.75	17.2	0.5–5	52.7	225	2

Note that data even among one glass type is scattered among different compositions, hence these values may only be taken as approximate for comparison sake.

Silica fiber Nearly all fibers used in telecommunications are based on silica and as a result the manufacture, splicing, cleaving, and polishing of this fiber has been optimized. As a result, adapting its use for high-power fiber lasers has been a natural transition. Silica is the only fiber commonly manufactured by the MCVD technique which is the fastest, simplest, and cheapest way to manufacture fiber preforms.¹¹ Silica's high damage threshold and melting point of approximately 2000°C make it especially suitable for high-power operation. A final benefit of silica is its physical durability to mechanical and thermal stresses. Silica is able to keep cleaved or polished surface well, is stable under vibration and strong enough to coil tightly, making silica fibers suitable for environmentally taxing packaged fiber laser applications.

As useful as it is in many applications silica does also have some cons including a positive refractive index change with temperature, a slightly higher n_2 than some glasses, a limited transparency window in the mid-IR due to its high phonon energy of 1100 cm⁻¹, as seen in Table 2.¹⁴⁰ Sometimes other host glasses provide superior laser parameters compared to silica.^{173,174} A final, and perhaps most important, limitation on silica is the relatively low dopant concentrations it can handle (a few wt. % before the onset of clustering and other detrimental effects).^{140,152} This maximum doping threshold makes it difficult to form highly doped fibers that may be advantageous for dopant concentration-dependent processes such as up conversion and cross relaxation. Highly doped short fibers, not practical in silica, have applications in ultrashort pulse amplification and single-frequency generation.

Phosphates, germanates, and tellurites Glasses such as phosphates with open glass structures are capable of handling far higher dopant concentrations compared to silica.^{140,152} Fiber lasers have been demonstrated with doping percentages as high as 10 times that allowable in silica^{175,176} enabling very short fiber lengths so that these fibers are suitable for high peak power amplification, narrow line-width generation, and for use in core designs that limit the length of the fiber due to bending losses.

Fiber end melting in end-pumped configurations is a significant challenge to these types of fibers as pump powers in the range of 20 to 100 W can cause catastrophic melting. However, using other more evenly distributed pump schemes heat has been more uniformly distributed and higher powers have been achieved.⁶⁴ Phosphate fiber lasers have achieved as high as 20-W output power with Yb doping¹⁷⁶ and 4-kW peak power in single frequency Q-switched systems,¹⁷⁷ germanate fiber lasers with pulse energies of 0.25 mJ and output powers of 104 W have also been reported.^{150,178}

Some of the soft glasses possess better laser characteristics for some dopant ions in terms of cross section and lifetime (there are many studies for different dopants and fiber types, however, see for example,¹⁷⁴ for tellurite fibers). They are unfortunately more difficult to manufacture in terms of cost of materials and required time for core drilling. Often, though not always, their difficulty in manufacture and limited power-handling outweighs their superior laser, thermal, and doping properties compared to silica.

Other Mid-IR Glasses: Fluorides and Others Though the bulk of current interest in high-power fiber lasers is concentrated in the near IR (1 to 2 μm), some applications require high powers outside of this relatively narrow band.

The most common glass material for producing high-power fiber lasers outside of the traditional wavelength band is the fluoride glass family. The most widespread of these glasses is so called ZBLAN, named for its chemical composition containing ZnF₄, BaF₂, LaF₃, AlF₃, and NaF.¹⁴⁰ Compared to silica, ZBLAN allows more laser wavelengths in both the visible and farther into the IR.¹⁴⁰ Several reasonably high-power lasers have been reported using ZBLAN doped with Ho with outputs of 0.38 W,^{155,156} Er with outputs of approximately 10 W at 2700 nm,¹⁴⁴ and in Tm with outputs of 20-W CW and 9-W average power pulsed operation with pulse energies of 90 μJ.^{173,179,180} ZBLAN has potential in particular laser applications, where its lifetime and cross-section properties make it advantageous in terms of its having a lower threshold behavior. The characteristics of Tm:ZBLAN and Tm:silica at approximately 50-W pumping levels have been compared and it is found that ZBLAN is superior in terms of efficiency and threshold.¹⁷³ Despite the clear benefits of ZBLAN (and other fluorides) in some situations, there are also limitations stemming from fabrication difficulties, low melting point, and damage threshold making it difficult to produce fiber lasers with this material above the 50-W level.¹⁷³ This precludes it from generating the extreme powers of silica-based fiber lasers and, as a result, keeps the mid-IR wavelengths it is able to produce limited to the sub 100-W level.

There are other potential glasses for mid-IR operation including chalcogenides; however, these glasses share the difficulties of ZBLAN in terms of damage threshold and cost of manufacture. Chalcogenides are difficult and expensive to make in quantities large enough to fabricate a fiber preform. Lasers based on these materials (a Raman fiber laser) have achieved power levels of 0.64 W.¹⁸¹

25.9 SPECTRAL AND TEMPORAL MODALITIES

High output powers from fiber lasers have reached the multi-kilowatts level with diffraction limited beam qualities.²¹ However, many applications demand narrow spectral bandwidths or tunable spectral bandwidths for spectral beam combining and spectroscopy applications. Other applications call for short pulse durations (nanosecond, picoseconds, and femtosecond) with high peak powers at high repetition rate such as LIDAR and materials processing.

High-Power Spectrally Controlled Fiber Lasers

Spectral control is one of the most critical aspects of high-power fiber laser design. Fiber lasers constructed without spectral control tend to display chaotic spectral behavior with lasing occurring in multiple regions of the gain spectrum simultaneously as observed in Refs. 173, 182, and 183. This wide spectral variation and indeterminacy is unsuitable for many applications including pumping of other lasers and for spectral beam combining where significantly higher spectral brightness is desired. Some situations also call for active wavelength selectivity.

Fiber Laser Spectral Tunability Fiber laser spectral tuning is most easily accomplished by the inclusion of a dispersive element in the resonator such as a diffraction grating, prism, or volume or fiber Bragg grating. Free-space fiber laser cavities are easily configured to be wavelength tuned by simply replacing an end mirror with the tunable optical element such as a conventional diffraction grating in the Littrow configuration. Only a small amount of feedback is required to efficiently control the wavelength in fiber lasers because of its high gain, spectral narrowing of laser linewidth, due to angular dispersion of the spectrum in space is caused by the spectrally selective elements, though in many cases at the cost of efficiency.³⁶

Yb-doped fiber lasers with approximately 50-W output and more than 50 nm of tunability were reported, as were Er:Yb-doped fiber lasers with more than 100-W output and more than 40 nm of tunability and Tm-doped fiber lasers with more than 10 W output and more than 200 nm of tunability. These were in tunable oscillator configurations and showed relatively constant output powers over the tuning range.^{36,184,185}

Spectral tuning can also be achieved via volume Bragg gratings (VBGs). These are diffractive holographic grating structures that give very narrow band feedback with higher efficiency than metal or ruled gratings.¹⁸⁶ A Yb fiber laser with 4.3-W output power and a 30-nm tuning range was demonstrated.¹⁸⁷ VBGs do not work in the Littrow configuration, and thus require an integrated feedback mirror. However, they offer the benefit of higher potential efficiency due to lower losses compared to traditional gratings.¹⁸⁷

Fiber lasers can also be tuned with fiber Bragg gratings. Because FBGs are simply layers of differing photoinduced refractive index change arranged in a fiber, changing the distance between these layers by mechanical stretching or thermal tuning can change the reflectivity of the grating. This has been done in many systems. Reference 195 provides an example of 30-nm tunability and output of 43 W.¹⁸⁸ The tuning range was limited by the amount of mechanical or thermal change a grating can tolerate.

Narrow Linewidth Fiber Lasers

Narrow wavelength and wavelength control can be achieved in fiber lasers either by using narrow linewidth spectral control elements in high-power laser cavities or by seeding a high-power amplifier chain with a separate spectrally controlled light source.

Fiber Bragg gratings are a common line narrowing mechanism which can be fabricated in photo-sensitive glass or via direct femtosecond writing of FBGs into fiber laser glass. The later type of grating has the advantage of being written directly in the gain fibers with no special glass needed. This type of FBG enabled 104-W output power and 260-pm linewidth.^{74,189} As seen in Chap. 17 by choosing correct design parameters, a FBG can be tailored to the desired wavelength, bandwidth, and reflectivity with linewidths as narrow as 0.01 nm achieved in single-mode fibers. FBGs have been proven able to handle very high powers, as they were incorporated into several of the highest power lasers reported.²¹

FBGs are not easily compatible with large-mode area fibers because FBGs in large core sizes cannot easily be made to the tight tolerances demanded for laser applications. The use of volume Bragg gratings (VBGs) and guided mode resonance filters (GMRFs) do not suffer this limitation. VBGs can be designed to be nearly 100 percent reflective at normal incidence in extremely narrow wavelength ranges. Their fabrication and design is detailed in Ref. 186. The first use of VBGs in fiber lasers involved low powers in large-core PCFs.¹⁹⁰ This was extended in Yb-doped fiber reaching output power of 4.3 W and tunable linewidth of 5 GHz.¹⁸⁷ A 103-W Er:Yb laser was also demonstrated and was tuned using the VBG over approximately 30 nm at an output power of approximately 30 W.¹⁹¹ VBGs in Tm fiber lasers have also been demonstrated, exhibiting powers of up to 5 W and linewidths as small as 300 pm.¹⁸³ The highest power VBG fiber laser demonstrated was linearly polarized and reached 138-W from Yb-doped fiber. Thermal limitations to the use of VBGs is discussed in Ref. 192.

Guided mode resonant filters (GMRFs) are based on writing a layer of subwavelength gratings on top of a waveguide layer in order to use waveguide coupling effects to cause very narrow band reflectivity. The details of their operation are given in Ref. 193. These elements have not been used at high powers though they have been used to cause significant linewidth narrowing of a watt-level fiber laser.¹⁹⁴

Many applications call for linewidths less than the picometer range. Since fiber lasers usually have long cavity lengths, with closely spaced longitudinal modes, spectral control with a single dispersive element to a single mode is challenging. Short fiber laser resonators using highly doped soft glasses at the watt level have been demonstrated,^{177,195} but are not scalable to higher power.

The most effective way to achieve high powers and extremely narrow linewidths is to use a MOPA seeded by either narrow linewidth diode lasers or distributed feedback fiber lasers.¹⁴⁰ Despite their low power, DFB lasers are ideal seeds as they offer minimal temperature sensitivity with low noise, 1 to 100 kHz linewidth and single polarization. High power, single frequency lasers have been built using all the major gain media. A Yb-doped single frequency laser reached 264 W at less than 60 kHz linewidth.¹⁹⁶ The highest power Tm MOPA reported is 20 W at less than 50 kHz linewidth based on a DFB and only limited by available pump power.¹⁹⁷ High-power Er:Yb MOPAs have also been reported reaching 151 W.¹⁹⁸ The onset of SBS which causes linewidth broadening at high powers is often a limiting factor in these systems. To mitigate this, limitation fiber cores must be made larger, fiber lengths must be made shorter or special techniques must be used to eliminate SBS.^{42,199}

Nanosecond Fiber Systems

High pulse energies in fiber lasers are limited by the optical damage threshold of the fiber. Fiber lasers, so far, are limited to nanosecond pulses with energies of a few millijoules. Even in 100- μ m core fibers such nanosecond pulses approach the optical damage threshold of silica glass.¹⁷ However, fiber lasers have the advantage of being able to operate at very high repetition rates (100s of kHz) so they complement bulk lasers in the high pulsed power regime. Nanosecond laser architectures are either Q-switched or use low-power seed pulses.

Q-Switched Oscillators Conventional Q-switched fiber lasers use a light modulator (passive, electro-optical (EO) or acousto-optical (AO)), adjacent to an angle cleaved fiber facet to avoid parasitic lasing between pulses, and a feedback element in the resonator. The main challenge with Q-switched fiber lasers is maintaining hold-off between pulses as ASE can build up to the detriment of laser efficiency.^{140,200}

Higher peak powers can be reached using large core fibers, end capped with coreless caps to prevent surface damage by expanding the beam before it exits the fiber. Several high-power Q-switched

oscillator systems have been reported with millijoule-level output powers based on both conventional LMA and PCF technologies. An LMA-based laser doped with Yb produced 8.4 mJ at 500 Hz and 0.6 mJ at 200 kHz with 120 W of average power at the higher repetition rate. The beam was slightly multimode giving M^2 of approximately 4.²⁰¹ A PCF-based Yb-doped fiber laser produced 10 ns pulses with energies up to 2 mJ and an average power of approximately 100 W.^{202,203} Tm-doped fiber lasers have also been Q-switched producing 30-W output power with 270- μ J pulse energies at 125-kHz repetition rates in conventional LMA fiber.²⁰⁴

Several potential monolithic Q-switching solutions have been tested, using both active and passive switching. Passive Q-switching usually involves using some kind of saturable absorber in the cavity either as an end mirror (making the system essentially monolithic), bulk crystal, or by splicing a section of saturable absorber into the fiber.¹⁴⁰ Saturable absorber mirrors and bulk saturable absorbers are limited in output power due to damage concerns in the saturable absorber elements. Nevertheless, saturable absorber Q-switched fiber lasers have achieved watt-level powers and 100- μ J pulse energies.^{205–207} Doped fiber with absorption at the desired operation wavelength has also been used as a saturable absorber. For example, a 10-W average power, approximately 100-kHz Tm laser with microsecond pulses was Q-switched with a Ho-doped fiber section.²⁰⁸ Other alternative Q-switching methods involve using mismatched FBGs, where a resonator is formed between two FBGs and the gratings are altered in length piezoelectrically to change their reflectivity peak and modulate cavity Q.^{140,209} Many other novel methods for Q-switching fiber lasers have been proposed including passive self Q-switching using SBS and using a piezoelectrically modulated high-Q microsphere or electro-optic-based metal-filled FBG.^{210–212} None of these techniques have been tested at high powers.

There is still a general challenge to attaining very short sub-20 nanosecond pulse durations in fiber lasers due to their long length.^{140,213} Typical fiber laser pulse durations are 100s of nanoseconds, while sub-100 ns are achievable with care. In very short PCF or highly doped soft glass lasers sub-10 ns pulses are achievable. A further challenge for Q-switched oscillators is the long intracavity length in a fiber laser causing unwanted nonlinear effects and even undesirable mode locking which affect the pulse shape and energy.^{140,214,215}

Nanosecond MOPA Systems When the most consistent and shortest possible pulse durations with highest achievable peak pulse powers are desired fiber MOPA systems seeded by Q-switched fiber lasers, microchip lasers or modulated laser diodes are a common solution. Microchip lasers possess small cavity sizes and can achieve very short pulse durations with reasonable peak output powers. They operate with passive Q-switching at fixed repetition rates anywhere from 3 to 100 kHz and produce seed energies more than 5 μ J.²⁰⁰ Diode laser seeds have the advantage of being flexible in terms of pulse shape and repetition rate when driven by arbitrarily shaped current waveforms. Modifications to the pulse shape in an amplifier can be calibrated out by tailoring the input pulse shape.^{90–92,216} Direct diode laser seeds require further stages of preamplification to achieve desired seed powers to saturate a power amplifier.

The multistage nature of the MOPA system configuration allows ASE between pulses to be filtered out by narrowband spectral filters or AO gates. MOPA systems currently able to achieve the highest peak powers usually rely on either conventional single-mode excitation in LMA fibers^{90–92,216} or PCF technologies^{20,108,200} to achieve high-beam qualities and large-mode areas.

Some of the downsides to MOPA systems are their complexity, the extra components they require and their longer time for assembly and optimization. MOPAs usually require at least one preamplifier (and in the case of diode systems two or more) to boost the seed power to a point where it can efficiently extract gain from a power amplifier. In addition, MOPAs require high-power optical isolators to protect earlier stages from back-reflected pulses as well as filters to remove ASE and mode field adapters to transfer signals from small-core to large-core stages.

End caps, sections of undoped, coreless fiber (solid silica) spliced to the end of a conventional fiber or sections of PCF with the air holes thermally collapsed allow the fiber mode to expand to much larger sizes before exiting the end face where fiber damage thresholds are far lower than in the bulk.^{20,200}

A Yb-doped system was reported which produced 6.2 mJ in sub-10-ns shaped pulses at approximately 2-kHz repetition rate using conventional LMA fiber with core diameter of 80 μ m. It had M^2 of approximately 1.3.^{92,217} Another similar system based on 200- μ m LMA fiber produced 27 mJ (82 mJ)

in 50 ns pulses, but with M^2 of 6.5 (though this is significantly better beam quality than expected from such a fiber due to the use of coiling techniques). Numerous PCF-based MOPA systems have been constructed with core sizes varying from 40 to 100 μm achieving upwards of 4.4 mJ of pulse energy at 10 kHz in 100- μm core with near-diffraction-limited beam quality.^{20,218,219} At these power levels, the pulse itself began to break up due to nonlinear effects in even such a large fiber core. Additional interest in high-power eyesafe MOPA systems has lead to work on Er:Yb systems capable of more than 300- μJ pulses at 6 kHz and (or 100 μJ) at 100 kHz.^{220–222} Tm-based systems in fluoride fibers have reached 5-kW peak power in 30-ns pulses with 33-kHz repetition rates and 1-kW peak power with 125-kHz repetition rates.¹⁷⁹

High-Power Ultrashort Pulse Technologies

Ultrashort pulses (USP) and their applications are an area of increasing interest in the laser community and because of their large gain bandwidths, potential compact size and inherent stability, fiber lasers are an ideal platform for the generation and amplification of high-power USPs for use in frequency conversion, material processing, remote sensing, high harmonic generation, and production of high-power stable frequency combs. As with nanosecond pulses, fiber lasers have fundamental limitations on pulse output energy caused by nonlinearities and damage thresholds within small fiber cores at energies over a few millijoules, but fiber lasers have the ability to provide these pulses at very high average powers and repetition rates. By using pulse stretching and compressing techniques, fiber lasers are capable of producing output energies in ultrashort pulses on the same order of magnitude of nanosecond pulses. Dispersion-management techniques in fibers allow them to readily achieve pulse durations on the order of many classes of bulk ultrashort lasers. System architectures for fiber-laser-based ultrashort systems take the same two basic forms as their nanosecond counterparts: direct generation of pulses by high-power oscillator systems or amplification based on chirped pulse amplification (CPA) MOPA systems.

High-Power USP Oscillators Ultrashort pulses (USPs) have been produced in fiber lasers for many years dating back to the early interest in the generation of pulses for communications applications. Most USP systems in fibers are capable of only modest output powers of less than 100 mW and pulse durations of picoseconds.¹⁴⁰ The earliest lasers were based on temporal solitons where pulse duration was constant throughout the resonator and the high peak pulse powers of even short pulses limited potential output power. Stretched pulse additive pulse mode locking techniques allowed pulse duration to be shortened and pulse energies to increase somewhat; however, they are still relatively low compared to LMA fiber laser standards.^{140,223–225} The advent of stretched pulse, self similar, or all normal dispersion fibers where the pulse is compressed external to the resonator allowed an increase of output energies to nearly 20 nJ. This is a limitation in such lasers due to their small-core conventional fiber rather than LMA construction.^{226–229} High-power LMA fiber lasers based on all-normal dispersion techniques have been produced and have achieved record output powers.^{68,230–233}

High-power mode-locked fiber laser oscillators rely on using external cavities and very large-mode area fibers to achieve their output powers. Cavities usually take the form of ring or sigma resonators containing the gain medium, dispersion compensation, and required polarization control elements. Most of these systems use external dispersion compensation such as chirped mirrors and gratings or use none at all and rely on extra-cavity pulse compression.

Detailed descriptions of many types of mode locking used in fiber lasers are found in Refs. 140, 213, 234, and 235. The most common technique is use of saturable absorbers such as SESAMs (semiconductor saturable absorber mirrors) or carbon nanotubes; SESAM operation is described in Ref. 236, and saturable absorbers based on the use of carbon nanotubes are described in Refs. 237 to 240.

Mode-locked fiber lasers will also require dispersion compensation to compress pulses to their minimum duration. Many fiber lasers rely on traditional bulk optics making them less useful as stable fiber-based systems. Other dispersion compensation alternatives may involve the use of PCF or HOM fibers which allow for dispersion correction with very large-mode areas, and thus all-fiber LMA oscillators.^{131,241}

Despite the high-power achievements in LMA-based mode-locked oscillators, there are difficulties associated with such systems. Oscillator systems run at very high repetition rates in the megahertz regime due to the cavity length leading to low pulse energies even for 100-W-level systems. In addition, pulse durations are more difficult to control and stability can be an issue when operating at high average powers.

Ultrashort MOPA Systems There are very few differences between the construction of MOPAs for ultrashort pulse operation and nanosecond operation. The systems' architectures are quite similar; the differences lie in the way the systems are seeded. USPs have very high peak power. To amplify USPs in fiber lasers the chirped pulse amplification (CPA) technique first demonstrated in fibers in Ref. 242 must be used. Pulses from a low-power mode-locked seed laser are amplified by one or more preamplifier stages. The pulse is then stretched in duration by giving it a linear chirp using the dispersive effects of bulk grating stretchers, chirped mirrors, prisms, or even simply lengths of fiber (including potentially PCF or HOM fiber). This stretched pulse is next injected into an amplifier system in the same way as a nanosecond pulse. The bandwidth of fiber amplifiers is suitably large to handle the wide bandwidth of even sub-100 fs pulses. After amplification, the pulse is recompressed to its shortest possible duration. An added advantage of the MOPA system is that optical modulators can be incorporated after the seed laser to act as pulse pickers to reduce the megahertz repetition rates to kilohertz level rates more suitable to high pulse energy amplification.

With the use of these various types of seed lasers Yb systems have reached 100- μ J pulse energies at 90-W average powers with pulses as short as 500 fs, thus leading to 120-MW peak powers.²⁴³ Higher average power megahertz repetition rate systems have reached 131 W of average power.²⁴⁴ An even higher power system using two large core PCF amplifiers produced 1.45 mJ at 100-kHz repetition rate with more than 100-W output power in approximately 800 fs compressed pulses.²⁴⁵ Though PCFs have produced the highest output powers, LMA conventional fibers are also very capable and have produced high powers. The highest reported outputs from LMA USP lasers being 50- μ J pulses from 65- μ m core fibers which are used for x-ray generation.^{246,247} Many Er:Yb-based systems have also been constructed, the largest of which manage to produce upward of 200 μ J at 5-kHz repetition rates.²⁴⁶ New VBG-based compression and stretching techniques in Er:Yb and Yb fiber lasers have also allowed an increase in efficiency in such CPA systems.^{248,249} In addition CPA systems have been commercialized and are available for use in materials-processing applications.²⁵⁰

USP, CPAMOPAs suffer similar issues as other MOPA systems including increased parts count and complexity, the need for mode field diameter adaptation from stage to stage, and the lack of all-fiber components suitable for high-power levels, therefore necessitating free-space operation. Still, MOPAs are the most effective way to generate high peak energy and peak-power ultrashort pulses from fiber lasers.

25.10 CONCLUSIONS

Based on the discussions, data, and results presented here, fiber lasers are an effective technology for the production of high-power laser light. Despite many excellent results in many regimes, there is still a need for further development of fiber laser technologies to make the leap further in the realm currently dominated by bulk lasers. The many techniques, technologies, concepts, and systems discussed are the groundwork for enabling the advancement of fiber lasers in the near and far future.

25.11 REFERENCES

1. E. Snitzer, "Optical Maser Action of Nd^{+3} in a Barium Crown Glass," *Phys. Rev. Lett.* 7(12):444–446, 1961.
2. E. Snitzer, "Proposed Fiber Cavities for Optical Masers," *J. Appl. Phys.* 32:36, 1961.
3. C. J. Koester and E. Snitzer, "Amplification in a Fiber Laser," *Appl. Opt.* 3(10):1182–1186, 1964.

4. J. Stone and C. A. Burrus, "Neodymium-Doped Silica Lasers in End-Pumped Fiber Geometry," *Appl. Phys. Lett.* **23**:388, 1973.
5. J. Stone and C. Burrus, "Neodymium-Doped Fiber Lasers: Room Temperature CW Operation with an Injection Laser Pump," *IEEE J. Quant. Electron.* **10**(9):794–794, 1974.
6. R. J. Mears, L. Reekie, S. B. Poole, D. N. Payne, "Neodymium-Doped Silica Single-Mode Fibre Lasers," *Electron. Lett.* **21**:738, 1985.
7. R. J. Mears, L. Reekie, S. B. Poole, and D. N. Payne, "Low-Threshold Tunable CW and Q-Switched Fibre Laser Operating at 1.55 μm ," *Electron. Lett.* **22**:159, 1986.
8. L. Reekie, R. Mears, S. Poole, and D. N. Payne, "Tunable Single-Mode Fiber Lasers," *J. Lightwave Technol.* **4**(7):956–960, 1986.
9. E. Desurvire, J. R. Simpson, and P. C. Becker, "High-Gain Erbium-Doped Traveling-Wave Fiber Amplifier," *Opt. Lett.* **12**(11):888–890, 1987.
10. R. J. Mears, L. Reekie, I. M. Jauncey, and D. N. Payne, "Low-Noise Erbium-Doped Fibre Amplifier Operating at 1.54 μm ," *Electron. Lett.* **23**:1026, 1987.
11. S. Poole, D. N. Payne, R. Mears, M. Fermann, and R. Laming, "Fabrication and Characterization of Low-Loss Optical Fibers Containing Rare-Earth Ions," *J. Lightwave Technol.* **4**(7):870–876, 1986.
12. E. Snitzer, H. Po, F. Hakimi, R. Tumminelli, and B. C. McCollum, "Double-Clad, Offset Core Nd-fiber Laser," *Opt. Fiber Sensors* **1**, 1988.
13. V. P. Gapontsev, P. I. Sawvsky, and I. E. Samartsev, "1.5 μm Erbium Glass Lasers," in *Conference on Lasers and ElectroOptics*, San Jose, Calif., 1990.
14. D. Minelly, E. R. Taylor, K. P. Iedrzewski, J. Wang, and D. N. Payne, "Laser-Diode Pumped Neodymium-Doped Fibre Laser with Output Power $>1\text{ W}$," in *Conference on Lasers and ElectroOptics*, 1992.
15. H. Po, J. D. Cao, B. M. Laliberte, R. A. Minns, R. F. Robinson, B. H. Rockney, R. R. Tricca, and Y. H. Zhang, "High Power Neodymium-Doped Single Transverse Mode Fibre Laser," *Electron. Lett.* **29**(17):1500–1501, 1993.
16. V. Dominic, S. MacCormack, R. Waarts, S. Sanders, S. Bicknese, R. Dohle, E. Wolak, P. Yeh, and E. Zucker, "110 W Fibre Laser," *Electron. Lett.* **35**(14):1158–1160, 1999.
17. C. C. Ranaud, H. L. Offerhaus, J. A. Alvarez-Chavez, J. Nilsson, W. A. Clarkson, P. Turner, D. J. Richardson, and A. B. Grudinin, "Characteristics of Q-Switched Cladding-Pumped Ytterbium-Doped Fiber Lasers with Different High-Energy Fiber Designs," *IEEE J. Quant. Electron.* **37**(2):199–206, 2001.
18. J. A. Alvarez-Chavez, H. L. Offerhaus, J. Nilsson, P. Turner, W. A. Clarkson, and D. J. Richardson, "High-Energy, High-Power Ytterbium-Doped Q-Switched Fiber Laser," *Opt. Lett.* **25**(1):37–39, 2000.
19. N. G. R. Broderick, H. L. Offerhaus, D. J. Richardson, R. A. Sammut, J. Caplen, and L. Dong, "Large Mode Area Fibers for High Power Applications," *Opt. Fiber Technol.* **5**(2):185–196, 1999.
20. F. DiTeodoro and C. D. Brooks, "Multi-mJ Energy, Multi-MW Peak Power Photonic Crystal Fiber Amplifiers with Near Diffraction Limited Output," in *CLEO*, Baltimore, 2007.
21. V. Fomin, A. Mashkin, M. Abramov, A. Ferin, and V. Gapontsev, "3 kW Yb Fiber Lasers with a Single Mode Output," in *International Symposium on High Power Fiber Lasers and their Applications*, St. Petersburg, 2006.
22. H. Zellmer, A. Tünnermann, H. Welling, and V. Reichel, "Double-Clad Fiber Laser with 30 W Output Power," in *Optical Amplifiers and Their Applications*, M. Zervas, A. Willner, and S. Sasaki, eds., Vol. 16 of OSA Trends in Optics and Photonics Series (Optical Society of America, 1997), paper FAW18.
23. K. Ueda, H. Sekiguchi, and H. Kan, "1 kW CW Output from Fiber-Embedded Disk Lasers," in *Proc. CLEO*. Long Beach, CA, 2002.
24. Y. Jeong, J. Sahu, R. B. Williams, D. J. Richardson, K. Furusawa, and J. Nilsson, "Ytterbium-Doped Large-core Fibre Laser with 272 W Output Power," *Electron. Lett.* **39**:977–978, 2003.
25. J. Limpert, A. Liem, H. Zellmer, and A. Tünnermann, "500 W Continuous-Wave Fibre Laser with Excellent Beam Quality," *Electron. Lett.* **39**(8):645–647, 2003.
26. V. P. Gapontsev, N. S. Platonov, O. Shkurihin, and I. Zaitsev, "400 W Low Noise Single-Mode CW Ytterbium Fiber Laser with an Integrated Fiber Delivery," in *Proc. CLEO*, Baltimore, 2003.
27. Y. Jeong, J. Sahu, D. Payne, and J. Nilsson, "Ytterbium-Doped Large-Core Fiber Laser with 1.36 kW Continuous-Wave Output Power," *Opt. Exp.* **12**(25):6088–6092, 2004.
28. Y. Jeong, J. Sahu, S. Baek, C. Alegria, D. B. S. Soh, C. Codemard, and J. Nilsson, "Cladding-Pumped Ytterbium-Doped Large-Core Fiber Laser with 610 W of Output Power," *Opt. Commun.* **234**(1–6):315–319, 2004.

29. C. H. Liu, B. Ehlers, F. Doerfel, S. Heinemann, A. Carter, K. Tankala, J. Farroni, and A. Galvanauskas, "810W Continuous-Wave and Single-Transverse-Mode Fibre Laser Using 20 μm Core Yb-Doped Double-Clad Fibre," *Electron. Lett.* **40**(23):1471–1472, 2004.
30. Y. Jeong, J. K. Sahu, D. N. Payne, and J. Nilsson, "Ytterbium-Doped Large-Core Fibre Laser with 1 kW of Continuous-Wave Output Power," *Electron. Lett.* **40**(8):470–472, 2004.
31. A. Liem, J. Limpert, H. Zellmer, A. Tunnermann, V. Reichel, K. Morl, S. Jetschke, H. R. Unger, Muller, J. Kirchof, T. Sandrock, and T. A. Harschak, "1.3 kW Yb-Doped Fiber Laser with Excellent Beam Quality," in *Proc. CLEO*, 2004. San Francisco, CA, USA.
32. V. P. Gapontsev, D. V. Gapontsev, N. S. Platonov, O. Shkurihin, V. Fomin, A. Mashkin, M. Abramov, and A. Ferin, "2 kW CW Ytterbium Fiber Laser with Record Diffraction Limited Brightness," in *Proc. CLEO Europe*, Munich, Germany, 2005.
33. C. H. Liu, A. Galvanauskas, B. Ehlers, F. Doerfel, S. Heinemann, A. Carter, A. Tanaka, and J. Farroni, "810-W Single Transverse Mode Yb-Doped Fiber Laser," in *ASSP*, Santa Fe, N. Mex., 2004.
34. N. S. Platonov, V. Gapontsev, V. P. Gapontsev, and V. Shumilin, "135 W CW Fiber Laser with Perfect Single Mode Output," in *Proc. CLEO*, Long Beach, Calif., 2002.
35. J. W. Dawson, M. J. Messerly, R. J. Beach, M. Y. Shverdin, E. A. Stappaerts, A. K. Sridharan, P. H. Pax, J. E. Heebner, C. W. Siders, and C. P. J. Bartly, "Analysis of the Scalability of Diffraction-Limited Fiber Lasers and Amplifiers to High Average Power," *Opt. Exp.* **16**(17):13240–13266, 2008.
36. J. Nilsson, W. A. Clarkson, R. Selvas, J. K. Sahu, P. W. Turner, S. U. Alam, and A. B. Grudinin, "High-Power Wavelength-Tunable Cladding-Pumped Rare-Earth-Doped Silica Fiber Lasers," *Opt. Fiber Technol.* **10**(1): 5–30, 2004.
37. J. Nilsson, J. K. Sahu, Y. Jeong, W. A. Clarkson, R. Selvas, A. B. Grudinin, and S. U. Alam, "High Power Fiber Lasers: New Developments," *Proc. of SPIE* **4974**:51, 2003.
38. W. Yong, "Heat Dissipation in Kilowatt Fiber Power Amplifiers," *IEEE J. Quant. Electron.* **40**(6):731–740, 2004.
39. W. Yong, X. Chang-Qing, and P. Hong, "Thermal Effects in Kilowatt Fiber Lasers," *Photonics Technol. Lett. IEEE* **16**(1):63–65, 2004.
40. M. K. Davis, M. J. F. Diggonnet, and R. H. Pantell, "Thermal Effects in Doped Fibers," *J. Lightwave Technol.* **16**(6):1013–1023, 1998.
41. B. C. Stuart, M. D. Feit, S. Herman, A. M. Rubenchik, B. W. Shore, and M. D. Perry, "Nanosecond-to-Femtosecond Laser-Induced Breakdown in Dielectrics," *Phys. Rev. B* **53**(4):1749, 1996.
42. V. I. Kovalev and R. G. Harrison, "Suppression of Stimulated Brillouin Scattering in High-Power Single-Frequency Fiber Amplifiers," *Opt. Lett.* **31**(2):161–163, 2006.
43. D. N. Payne, Y. Jeong, J. Nilsson, J. K. Sahu, D. B. S. Soh, C. Alegria, P. Dupriez, et al., *Kilowatt-Class Single-Frequency Fiber Sources, (Invited Paper)*: SPIE Volume 5709: Fiber Lasers II: Technology, Systems and Applications 1, 2005.
44. P. D. Dragic, L. Chi-Hung, G. C. Papen, and A. Galvanauskas, "Optical Fiber with an Acoustic Guiding Layer for Stimulated Brillouin Scattering Suppression," in *Conference on Lasers and Electro-Optics/Quantum Electronics and Laser Science and Photonic Applications Systems Technologies*: Optical Society of America, 2005.
45. R. Paschotta, J. Nilsson, A. C. Tropper, D. C. Hanna, "Ytterbium-Doped Fiber Amplifiers," *IEEE J. Quant. Electron.* **33**(7):1049–1056, 1997.
46. W. Yong and P. Hong, "Dynamic Characteristics of Double-Clad Fiber Amplifiers for High-Power Pulse Amplification," *J. Lightwave Technol.* **21**(10):2262–2270, 2003.
47. H. M. Pask, R. J. Carman, D. C. Hanna, A. C. Tropper, C. J. Mackenchnie, P. R. Barber, and J. M. Dawes, "Ytterbium-Doped Silica Fiber Lasers: Versatile Sources for the 1–1.2 μm Region," *IEEE J. Sel. Top. Quant. Electron.* **1**(1):2–13, 1995.
48. W. W. Rigrod, "Saturation Effects in High-Gain Lasers," *J. Appl. Phys.* **36**(8):2487–2490, 1965.
49. J. A. Buck, *Fundamentals of Optical Fibers*, New York:Wiley-IEEE Press, 2004.
50. D. Marcuse, "Loss Analysis of Single Mode Fiber Splices," *Bell System Technical Journal*, **56**:703–718, 1977.
51. A. Yariv, *Optical Electronics in Modern Communications*, 5th ed., New York: Oxford University Press, 1997.
52. C. Ullmann and V. Krause, (eds.), *Diode Optics and Diode Lasers*, U.P. Office, United States, 1999.
53. W. A. Clarkson and D. C. Hanna, "Two Mirror Beam Shaping Technique for High Power Diode Bars," *Opt. Lett.* **21**:375–377, 1996.

54. L. Goldberg, J. P. Koplow, and D. A. V. Kliner, "Highly Efficient 4-W Yb-Doped Fiber Amplifier Pumped by a Broad-Stripe Laser Diode," *Opt. Lett.* **24**(10):673–675, 1999.
55. D. J. Ripin, and L. Goldberg, "High Efficiency Side-Coupling of Light into Optical Fibres Using Imbedded V-Grooves," *Electron. Lett.* **31**(25):2204–2205, 1995.
56. F. Gonthier, L. Martineau, N. Azami, M. Faucher, F. Seguin, D. Stryckman, and A. Villeneuve, *High-Power All-Fiber Components: the Missing Link for High-Power Fiber Lasers*, in *Fiber Lasers: Technology, Systems, and Applications*: SPIE, San Jose, Calif., 2004.
57. C. Headley III, M. Fishteyn, A. D. Yablon, M. J. Andrejco, K. Brar, J. Mann, M. D. Mermelstein, and D. J. DiGiovanni, "Tapered Fiber Bundles for Combining Laser Pumps (Invited Paper)," in *Fiber Lasers II: Technology, Systems, and Applications*: SPIE, San Jose, Calif., 2005.
58. A. Kosterin, V. Temyanko, M. Fallahi, and M. Mansuripur, "Tapered Fiber Bundles for Combining High-Power Diode Lasers," *Appl. Opt.* **43**(19):3893–3900, 2004.
59. B. Samson, and G. Frith, *Diode Pump Requirements for High Power Fiber Lasers*, in *ICALEO*, Orlando, Fla, 2007.
60. J. Xu, J. Lu, G. Kumar, J. Lu, and K. Ueda, "A Non-Fused Fiber Coupler for Side-Pumping of Double-Clad Fiber Lasers," *Opt. Commun.* **220**(4–6):389–395, 2003.
61. C. Codemard, K. Yla-Jarkko, J. Singleton, P. W. Turner, I. Godfrey, S. U. Alam, J. Nilsson, J. Sahu, and A. B. Grudinin, "Low-Noise Intelligent Cladding-Pumped L-Band EDFA," *Photon. Technol. Lett. IEEE* **15**(7):909–911, 2003.
62. Alam, S.-U. J. Nilsson, P. W. Turner, M. Ibsen, and A. B. Grudinin, "Low Cost Multi-Port Reconfigurable Erbium Doped Cladding Pumped Fiber Amplifier," in *Proc. ECOC'00*, vol. 2, 2000, pp. 119–120. Munich, Germany, 2000.
63. X. J. Gu, and Y. Liu, "The Efficient Light Coupling in a Twin-Core Fiber Waveguide," *Photonics Technol. Lett. IEEE* **17**(10):2125–2127, 2005.
64. P. Polynkin, V. Temyanko, M. Mansuripur, and N. Peyghambarian, "Efficient and Scalable Side Pumping Scheme for Short High-Power Optical Fiber Lasers and Amplifiers," *Photonics Technol. Lett. IEEE*, **16**(9):2024–2026, 2004.
65. "SPI Lasers," available: www.spilasers.com, accessed on: Dec. 2008.
66. P. Even, and D. Pureur. "High-Power Double-Clad Fiber Lasers: A Review" in *Optical Devices for Fiber Communication III*: SPIE, San Jose, CA, 2002.
67. F. Haxsen, A. Ruehl, M. Engelbrecht, D. Wandt, U. Morgner, and D. Kracht, "Stretched-Pulse Operation of Athulium-Doped Fiber Laser," *Opt. Exp.* **16**(25):20471–20476, 2008.
68. B. Ortaç, C. Lecaplain, A. Hideur, T. Schreiber, J. Limpert, and A. Tünnermann, "Passively Mode-Locked Single-Polarization Microstructure Fiber Laser," *Opt. Exp.* **16**(3):2122–2128, 2008.
69. J. Nilsson, and B. Jaskorzynska, "Modeling and Optimization of Low-Repetition-Rate High-Energy Pulse Amplification in CW-Pumped Erbium-Doped Fiber Amplifiers," *Opt. Lett.* **18**(24):2099, 1993.
70. P. Wang, J. K. Sahu, and W. A. Clarkson, "High-Power Broadband Ytterbium-Doped Helical-Core Fiber Superfluorescent Source," *Photonics Technol. Lett. IEEE*, **19**(5):300–302, 2007.
71. M. Kihara, and S. Tomita, "Loss Characteristics of Thermally Diffused Expanded Core Fibers," *Photonics Technol. Lett.* **4**(12):1390–1391, 1992.
72. K. O. Hill, and G. Meltz, "Fiber Bragg Grating Technology Fundamentals and Overview," *J. Lightwave Technol.* **15**(8):1263–1276, 1997.
73. L. B. Fu, G. D. Marshall, J. A. Bolger, P. Steinvurzel, E. C. Magi, M. J. Withford, and B. J. Eggleton, "Femtosecond Laser Writing Bragg Gratings in Pure Silica Photonic Crystal Fibres," *Electron. Lett.* **41**(11):638–640, 2005.
74. N. Jovanovic, A. Fuerbach, G. D. Marshall, M. J. Withford, and S. D. Jackson, "Stable High-Power Continuous-Wave Yb³⁺-Doped Silica Fiber Laser Utilizing a Point-by-Point Inscribed Fiber Bragg Grating," *Opt. Lett.* **32**(11):1486–1488, 2007.
75. Y. Lai, A. Martinez, I. Khrushchev, and I. Bennion, "Distributed Bragg Reflector Fiber Laser Fabricated by Femtosecond Laser Inscription," *Opt. Lett.* **31**(11):1672–1674, 2006.
76. A. D. Yablon, *Optical Fiber Fusion Splicing*, Springer, New York, 2005.
77. D. J. Richardson, P. Britton, and D. Taverner, "Diode-Pumped, High-Energy, Single Transverse Mode Q-Switch Fibre Laser," *Electron. Lett.* **33**(23):1955–1956, 1997.

78. H. L. Offerhaus, N. G. Broderick, D. J. Richardson, R. Sammut, J. Caplen, and L. Dong, "High-Energy Single-Transverse-Mode Q-Switched Fiber Laser Based on a Multimode Large-Mode-Area Erbium-Doped Fiber," *Opt. Lett.* **23**(21):1683–1685, 1998.
79. T. Bhutta, J. I. Mackenzie, D. P. Shepherd, and R. J. Beach, "Spatial Dopant Profiles for Transverse-Mode Selection in Multimode Waveguides," *J. Opt. Soc. Am. B* **19**(7):1539–1543, 2002.
80. J. P. Koplow, D. A. V. Kliner, and L. Goldberg, "Single-Mode Operation of a Coiled Multimode Fiber Amplifier," *Opt. Lett.* **25**(7):442–444, 2000.
81. D. Marcuse, "Curvature Loss Formula for Optical Fibers," *J. Opt. Soc. Am.* **66**(3):216, 1976.
82. D. Marcuse, "Field Deformation and Loss Caused by Curvature of Optical Fibers," *J. Opt. Soc. Am.* **66**(4):311, 1976.
83. D. Marcuse, "Influence of Curvature on the Losses of Doubly Clad Fibers," *Appl. Opt.* **21**(23):4208, 1982.
84. R. T. Schermer, "Mode Scalability in Bent Optical Fibers," *Opt. Exp.* **15**(24):15674–15701, 2007.
85. J. W. Dawson, R. J. Beach, I. Jovanovic, W. Benoit, Z. Liao, and S. Payne, "Large Flattened Mode Optical Fiber for Reduction of Nonlinear Effects," in *Proc. SPIE Volume 5335 Fiber Lasers: Technology, Systems and Applications*. 2004.
86. A. K. Ghatak, I. C. Goyal, and R. Jindal, "Design of a Waveguide Refractive Index Profile to Obtain a Flat Modal Field," in *International Conference on Fiber Optics and Photonics: Selected Papers from Photonics India '98*, New Delhi, India: SPIE, 1999.
87. P. Facq, F. de Fornel, and F. Jean, "Tunable Single-Mode Excitation in Multimode Fibres," *Electron. Lett.* **20**(15):613–614, 1984.
88. M. E. Fermann, "Single-Mode Excitation of Multimode Fibers with Ultrashort Pulses," *Opt. Lett.* **23**(1):52–54, 1998.
89. C. D. Stacey, R. M. Jenkins, J. Banerji, and A. R. Davies, "Demonstration of Fundamental Mode only Propagation in Highly Multimode Fibre for High Power EDFAs," *Opt. Commun.* **269**(2):310–314, 2007.
90. K. C. Hou, M. Y. Cheng, D. Engin, R. Changkakoti, P. Mamidipudi, and A. Galvanauskas, "Multi-MW Peak Power Scaling of Single Transverse Mode Pulses Using 80 μm Core Yb-Doped LMA Fibers," Optical Society of America *Advanced Solid State Photonics*, Tahoe, Calif., 2006.
91. A. Galvanauskas, C. Ming-Yuan, H. Kai-Chung, and A.K.-H.L. Kai-Hsiu Liao, "High Peak Power Pulse Amplification in Large-Core Yb-Doped Fiber Amplifiers," *IEEE J. Sel. Top. in Quant. Electron.* **13**(3):559–566, 2007.
92. K. -C. Hou, S. George, A. G. Mordovanakis, K. Takenoshita, J. Nees, B. Lafontaine, M. Richardson, and A. Galvanauskas, "High Power Fiber Laser Driver for Efficient EUV Lithography Source with Tin-Doped Water Droplet Targets," *Opt. Exp.* **16**(2):965–974, 2008.
93. J. A. Alvarez-Chavez, A. B. Grudinin, J. Nilsson, P. W. Turner, and W. A. Clarkson, "Mode Selection in High Power Cladding Pumped Fiber Lasers with Tapered Section," p. 247–248 in *CLEO*. Washington D.C., 1999.
94. W. A. Clarkson, "Short Course Notes SC270: High Power Fiber Lasers and Amplifiers," in *Short Courses at CLEO 2007*, Optoelectronics Research Center, University of Southampton, Baltimore, Md., 2007.
95. V. Filippov, Y. Chamorovskii, J. Kerttula, K. Golant, M. Pessa, and O. G. Okhotnikov, "Double Clad Tapered Fiber for High Power Applications," *Opt. Express* **16**(3):1929–1944, 2008.
96. D. Taverner, D. J. Richardson, L. Dong, J. E. Caplen, K. Williams, and R. V. Penty, "158- μJ Pulses from a Single-Transverse-Mode, Large-Mode-Area Erbium-Doped Fiber Amplifier," *Opt. Lett.* **22**(6):378–380, 1997.
97. P. S. J. Russell, "Photonic-Crystal Fibers," *J. Lightwave Technol.* **24**(12):4729–4749, 2006.
98. J. C. Knight, "Photonic Crystal Fibers and Fiber Lasers (Invited)," *J. Opt. Soc. Am. B* **24**(8):1661–1668, 2007.
99. T. A. Birks, P. J. Roberts, P. S. J. Russell, D. Atkin, and T. Shepherd, "Full 2-D Photonic Bandgaps in Silica/air Structures," *Electron. Lett.* **31**(22):1941–1943, 1995.
100. J. Broeng, D. Mogilevstev, S. E. Barkou, and A. Bjarklev, "Photonic Crystal Fibers: A New Class of Optical Waveguides," *Optical Fiber Technology*, **5**(3):305–330, 1999.
101. J. C. Knight, T. A. Birks, P. St. J. Russell, and D. M. Atkin, "All-Silica Single-Mode Optical Fiber with Photonic Crystal Cladding," *Opt. Lett.* **21**, 1547–1549 (1996).
102. W. J. Wadsworth, J. C. Knight, and P. S. and J. Russell, "Large Mode Area Photonic Crystal Fibre Laser," in *Conference in Lasers and Electrooptics*, Washington D.C., 2001.

103. W. J. Wadsworth, J. C. Knight, and P. St. J. Russell, "Yb³⁺-Doped Photonic Crystal Fibre Laser," *Electron. Lett.* **36**:1452–1453, 2000.
104. Crystal-Fibre. *PCF Technology Tutorial*, available: www.crystal-fibre.com, accessed on: Dec. 15, 2008.
105. T. A. Birks, J. C. Knight, and P. S. J. Russell, "Endlessly Single-Mode Photonic Crystal Fiber," *Opt. Lett.* **22**(13):961–963, 1997.
106. T. M. Monro, D. J. Richardson, N. G. R. Broderick, and P. J. Bennett, "Holey Optical Fibers: An Efficient Modal Model," *J. Lightwave Technol.* **17**(6):1093, 1999.
107. K. Furusawa, A. Malinowski, J. Price, T. Monro, J. Sahu, J. Nilsson, and D. Richardson, "Cladding Pumped Ytterbium-Doped Fiber Laser with Holey Inner and Outer Cladding," *Opt. Exp.* **9**(13):714–720, 2001.
108. K. P. Hansen, J. Broeng, P. M. W. Skovgaard, J. P. Folkenberg, M. D. Nielsen, A. Petersson, T. P. Hansen, et al., "High Power Photonic Crystal Fiber Lasers: Design, Handling and Subassemblies," in *Photonics West.*: San Jose, Calif., 2005.
109. W. Wadsworth, R. Percival, G. Bouwmans, J. Knight, and P. Russell, "High Power Air-Clad Photonic Crystal Fibre Laser," *Opt. Exp.* **11**(1):48–53, 2003.
110. B. Zintzen, T. Langer, J. Geiger, D. Hoffmann, and P. Loosen, "Heat Transport in Solid and Air-Clad Fibers for High-Power Fiber Lasers," *Opt. Exp.* **15**(25):16787–16793, 2007.
111. J. Limpert, O. Schmidt, J. Rothhardt, F. Röser, T. Schreiber, A. Tünnermann, S. Ermeneux, P. Yvernault, and F. Salin, "Extended single-mode Photonic Crystal Fiber Lasers," *Opt. Exp.* **14**(7):2715–2720, 2006.
112. G. Bonati, H. Voelkel, T. Gabler, U. Krause, A. Tuennermann, A. Liem, T. Schreiber, S. Nolte, and H. Zellmer, "1.53 kW from a Single Yb-Doped Photonic Crystal Fiber Laser," in *Late Breaking News, Photonics West*, San Jose, Calif., 2005.
113. O. Schmidt, J. Rothhardt, T. Eidam, F. Röser, J. Limpert, A. Tünnermann, K.P. Hansen, C. Jakobsen, and J. Broeng, "Single-Polarization Ultra-Large-Mode-Area Yb-Doped Photonic Crystal Fiber," *Opt. Exp.* **16**(6):3918–3923, 2008.
114. J. Broeng, G. Vienne, A. Petersson, P. M. W. Skovgaard, J. P. Folkenberg, M. D. Nielsen, C. Jakobsen, H. R. Simonsen, and N. A. Mortensen, "Air-Clad Photonic Crystal Fibers for High-Power Single-Mode Lasers," in *Photonics West.*, San Jose, Calif., 2004.
115. V. V. R. Kumar, A. George, W. Reeves, J. Knight, P. Russell, F. Omenetto, and A. Taylor, "Extruded Soft Glass Photonic Crystal Fiber for Ultrabroad Supercontinuum Generation," *Opt. Exp.* **10**(25):1520–1525, 2002.
116. H. Ebendorff-Heidepriem, and T. M. Monro, "Extrusion of Complex Preforms for Microstructured Optical Fibers," *Opt. Exp.* **15**(23):15086–15092, 2007.
117. A. Isomäki, and O. G. Okhotnikov, "Femtosecond Soliton Mode-Locked Laser Based on Ytterbium-Doped Photonic Bandgap Fiber," *Opt. Exp.* **14**(20):9238–9243, 2006.
118. A. Wang, A. K. George, and J. C. Knight, "Three-Level Neodymium Fiber Laser Incorporating Photonic Bandgap Fiber," *Opt. Lett.* **31**(10):1388–1390, 2006.
119. F. Qiang, W. Zhi, K. Guiyun, Long Jin., Yang Yue., Jiangbing Du, Qing Shi., Zhanyuan Liu., Bo Liu., Yange Liu., Shuzhong Yuan., and Xiaoyi Dong. "Proposal for All-Solid Photonic Bandgap Fiber with Improved Dispersion Characteristics," *Photonics Technol. Lett. IEEE* **19**(16):1239–1241, 2007.
120. L. Dong, J. Li, H. McKay, A. Marcinkevicius, B. Thomas, M. Moore, L. Fu, and M. E. Fermann, "Robust and Practical Optical Fibers for Single Mode Operation with Core Diameters up to 170 μm ," in *CLEO 2008*, San Jose, Calif., 2008.
121. L. Dong, X. Peng, and J. Li, "Leakage Channel Optical Fibers with Large Effective Area," *J. OSA B*, **24**(8):1689–1697, 2007.
122. Wong, William S., X. Peng, McLaughlin, M. Joseph, and L. Dong, "Breaking the Limit of Maximum Effective Area for Robust Single-Mode Propagation in Optical Fibers," *Opt. Lett.* **30**(21):2855–2857, 2005.
123. C. H. Liu, G. Chang, N. Litchinitser, D. Guertin, N. Jacobsen, K. Tankala, and A. Galvanauskas, "Chirally Coupled Core Fibers at 1550-nm and 1064-nm for Effectively Single-Mode Core Size Scaling," in *CLEO*, Baltimore, 2007.
124. M. C. Swan, C. H. Liu, D. Guertin, N. Jacobsen, A. Tanaka, and A. Galvanauskas, "33 μm Core Effectively Single-Mode Chirally-Coupled-Core Fiber Laser at 1064-nm," in *Optical Fiber Communication Conference & Exposition and the National Fiber Optic Engineers Conference*. San Diego, Calif., 2008.
125. R. J. Beach, M. D. Feit, R. H. Page, L. D. Brasure, R. Wilcox, and S. A. Payne, "Scalable Antiguided Ribbon Laser," *J. Opt. Soc. Am. B*, **19**(7):1521–1534, 2002.

126. M. Wrage, P. Glas, M. Leitner, T. Sandrock, N. N. Elkin, A. P. Napartovich, and A. G. Sukharev, "Experimental and Numerical Determination of Coupling Constant in a Multicore Fiber," *Opt. Commun.* **175**(1–3):97–102, 2000.
127. T. Erdogan, "Cladding-Mode Resonances in Short- and Long-Period Fiber Grating Filters," *J. Opt. Soc. Am. A*, **14**(8):1760–1773, 1997.
128. S. Ramachandran, J. W. Nicholson, S. Ghalmi, M. F. Yan, P. Wisk, E. Monberg, and F. V. Dimarcello, "Light Propagation with Ultralarge Modal Areas in Optical Fibers," *Opt. Lett.* **31**(12):1797–1799, 2006.
129. S. Ramachandran, and S. Ghalmi, "Diffraction Free" Self Healing Bessel Beams from Fibers," in *CLEO 2008*, San Jose, Calif., 2008.
130. S. Suzuki, A. Schülzgen, and N. Peyghambarian, "Single-Mode Fiber Laser Based on Core-Cladding Mode Conversion," *Opt. Lett.* **33**(4):351–353, 2008.
131. S. Ramachandran, S. Ghalmi, J. W. Nicholson, M. F. Yan, P. Wisk, E. Monberg, and F. V. Dimarcello, "Anomalous Dispersion in a Solid, Silica-Based Fiber," *Opt. Lett.* **31**(17):2532–2534, 2006.
132. M. Schultz, O. Prochnow, A. Ruehl, D. Wandt, D. Kracht, S. Ramachandran, and S. Ghalmi, "Sub-60-fs Ytterbium-Doped Fiber Laser with a Fiber-Based Dispersion Compensation," *Opt. Lett.* **32**(16):2372–2374, 2007.
133. M. D. Mermelstein, S. Ramachandran, J. M. Fini, and S. Ghalmi, "SBS Gain Efficiency Measurements and Modeling in a 1714 μm^2 Effective Area LP_{08} Higher-Order Mode Optical Fiber," *Opt. Exp.* **15**(24):15952–15963, 2007.
134. A. E. Siegman, "Propagating Modes in Gain-Guided Optical Fibers," *J. Opt. Soc. Am. A*, **20**(8):1617–1628, 2003.
135. A. E. Siegman, "Gain-Guided, Index-Antiguide Fiber Lasers," *J. Opt. Soc. Am. B* **24**(8):1677–1682, 2007.
136. Y. Chen, T. McComb, V. Sudesh, M. Richardson, and M. Bass, "Very Large-Core, Single-Mode, Gain-Guided, Index-Antiguide Fiber Lasers," *Opt. Lett.* **32**(17):2505–2507, 2007.
137. Y. Chen, T. McComb, V. Sudesh, M. C. Richardson, M. Bass, "Lasing in a Gain-Guided Index Antiguide Fiber," *J. Opt. Soc. Am. B* **24**(8):1683–1688, 2007.
138. V. Sudesh, T. McComb, Y. Chen, M. Bass, M. Richardson, J. Ballato, and A. E. Siegman, "Diode-Pumped 200 μm Diameter Core, Gain-Guided, Index-Antiguide Single Mode Fiber Laser," *Appl. Phys. B: Lasers and Opt.* **90**(3):369–372, 2008.
139. P. Pavel, T. Valery, M. Jerome, and N. Peyghambarian, "Dramatic Change of Guiding Properties in Heavily Yb-Doped, Soft-Glass Active Fibers Caused by Optical Pumping," *Appl. Phys. Lett.* **90**(24):241106-1–241106-3, 2007.
140. M. J. F. Digonnet, *Rare-Earth-Doped Fiber Lasers and Amplifiers*, CRC Press, New York, 2001.
141. P. Hamamatsu, "The Fiber Disk Laser Explained," *Nat Photon*, **sample**(sample):14–15, 2006.
142. Y. Jeong, S. Yoo, C. A. Codemard, J. Nilsson, J. K. Sahu, D. N. Payne, R. Horley, P. W. Turner, L. M. B. Hickey, A. Harker, M. Lovelady, and A. Piper, "Erbium:Ytterbium Co-Doped Large-Core Fiber Laser with 297 W Continuous-Wave Output Power," *IEEE J. Sel. Top. Quant. Electron.* **13**(3):573–579.
143. X. Zhu, "Mid-IR ZBLAN Fiber Laser Approaches 10 W Output," in *Laser Focus World*, **44**(3), PennWell Corp. Tulsa Okla., 2007.
144. X. Zhu, and R. Jain, "10-W-Level Diode-Pumped Compact 2.78 μm ZBLAN Fiber Laser," *Opt. Lett.* **32**(1):26–29, 2007.
145. X. Zhu, and R. Jain, "Compact 2 W Wavelength-Tunable Er:ZBLAN Mid-Infrared Fiber Laser," *Opt. Lett.* **32**(16):2381–2383, 2007.
146. S. D. Jackson, F. Bugge, and G. Erbert, "High-Power and Highly Efficient Ti^{3+} -Doped Silica Fiber Lasers Pumped with Diode Lasers Operating at 1150 nm," *Opt. Lett.* **32**(19), 2007.
147. S. D. Jackson, "Power Scaling Method for 2-mm Diode-Cladding-Pumped Ti^{3+} -Doped Silica Fiber Lasers that Uses Yb^{3+} Codoping," *Opt. Lett.* **28**(22):2192, 2003.
148. D. V. Gapontsev, N. S. Platonov, M. Meleshkevich, A. Drozhzhin, and V. Sergeev, "415 W Single Mode CW Thulium Fiber Laser in all Fiber Format," in *CLEO Europe 2007 Munich, Germany*, 2007.
149. S. D. Jackson, "Cross Relaxation and Energy Transfer Upconversion Processes Relevant to the Functioning of 2 μm Ti^{3+} -Doped Silica Fibre Lasers," *Opt. Commun.* **230**(1–3):197–203, 2004.
150. J. Wu, Z. Yao, J. Zong, and S. Jiang, "Highly Efficient High-Power Thulium-Doped Germanate Glass Fiber Laser," *Opt. Lett.* **32**(6):638–640, 2007.

151. E. Slobodtchikov, P. F. Moulton, and G. Frith, "Efficient, High Power, Tm-Doped Silica Fiber Laser," in *ASSP 2007 Postdeadline*. Vancouver, Canada, 2007.
152. S. D. Jackson, and S. Mossman, "Efficiency Dependence on the Tm and Al Concentrations for Tm-Doped Silica Double-Clad Fiber Lasers," *Appl. Opt.* **42**(15), 2003.
153. S. D. Jackson, and T. A. King, "Theoretical Modeling of Tm-Doped Silica Fiber Lasers," *J. Lightwave Technol.* **17**(5):948–956, 1999.
154. P. F. Moulton, G. A. Rines, E. V. Slobodtchikov, K. F. Wall, G. Frith, B. Samson, and A. L. G. Carter, "Tm-Doped Fiber Lasers: Fundamentals and Power Scaling," *IEEE J. Sel. Top. Quant. Electron.* **15**(1): 85–92, 2009.
155. S. D. Jackson, "Midinfrared Holmium Fiber Lasers," *J. Quant. Electron. IEEE* **42**(2):187–191, 2006.
156. S. D. Jackson, F. Bugge, and G. Erbert, "Directly Diode-Pumped Holmium Fiber Lasers," *Opt. Lett.* **32**(17):2496–2498, 2007.
157. S. D. Jackson, F. Bugge, and G. Erbert, "High-Power and Highly Efficient Diode-Cladding-Pumped Ho³⁺-Doped Silica Fiber Lasers," *Opt. Lett.* **32**(22):3349–3351, 2007.
158. S. D. Jackson, and S. Mossman, "Diode-Cladding-Pumped Yb³⁺, Ho³⁺-Doped Silica Fiber Laser Operating at 2.1- μ m," *Appl. Opt.* **42**(18):3546–3549, 2003.
159. S.D. Jackson, A. Sabella, A. Hemming, S. Bennetts, and D. G. Lancaster, "High-Power 83 W Holmium-Doped Silica Fiber Laser Operating with High Beam Quality," *Opt. Lett.* **32**(3):241–243, 2007.
160. D. Hewak, (ed.), *Properties, Processing and Applications of Glass and Rare Earth Doped Glasses for Optical Fibers*, INSPEC Publications, London, 1998.
161. D. J. Digiovanni, A. M. Vengsarkar, J. L. Wagener, and R. S. Windeler, U.S. Patent 5,802,236 *Article Comprising a Micro-Structured Optical Fiber, and Method of Making Such Fiber*: United States Patent Office Office. USA, 1998.
162. H. Bookey, K. Bindra, A. Kar, and B. A. Wherrett, "Telluride Glass Fibres for all Optical Switching: Nonlinear Optical Properties and Fibre Characterisation," in *Workshop on Fibre and Optical Passive Components, Proc. 2002 IEEE/LEOS*. Glasgow, Scotland 2002.
163. G. Boudebs, W. Berlatier, S. Cherukulappurath, F. Smektala, M. Guignard, and J. Troles, "Nonlinear Optical Properties of Chalcogenide Glasses at Telecommunication Wavelength Using Nonlinear Imaging Technique," in *Transparent Optical Networks, 2004, Proc. 2004 6th International Conf. on Transparent Optical Networks*, 2004.
164. C. C. Chen, Y. J. Wu, and L. G. Hwa, "Temperature Dependence of Elastic Properties of ZBLAN Glasses," *Materials Chemistry and Physics* **65**(3):306–309, 2000.
165. J. A. Harrington, *Infrared Fibers and Their Applications*, SPIE Press, Bellingham, Wash. 2004.
166. Y. T. Hayden, S. Payne, J. S. Hayden, J. Campbell, M. K. Aston, and M. Elder, U.S. Patent 5526369 *Phosphate Glass Useful in High Energy Lasers*, United States Patent Office, 1996.
167. "Laser Glass Properties," available: www.kigre.com, accessed on: Oct. 15 2008.
168. A. Kut'in, V. Polyakov, A. Gibin, and M. Churbanov, "Thermal Conductivity of (TeO₂)_{0.7}(WO₃)_{0.2}(La₂O₃)_{0.1} Glass," *Inorg. Mater.* **42**(12):1393–1396, 2006.
169. M. D. O'Donnell, K. Richardson, R. Stolen, A. B. Seddon, D. Furniss, V. K. Tikhomirov, C. Rivero, et al., "Tellurite and Fluorotellurite Glasses for Fiberoptic Raman Amplifiers: Glass Characterization, Optical Properties, Raman Gain, Preliminary Fiberization, and Fiber Characterization," *J. Am. Ceram. Soc.*, **90**(5):1448–1457, 2007.
170. T. Töpfer, J. Hein, J. Philipps, D. Ehrt, and R. Sauerbrey, "Tailoring the Nonlinear Refractive Index of Fluoride-Phosphate Glasses for Laser Applications," *Appl. Phys. B: Lasers Opt.* **71**(2):203–206, 2000.
171. M. Yamane, and Y. Asahara, *Glasses for Photonics*, Cambridge University Press, Cambridge, UK, 2000.
172. G. Chen, Q. Zhang, G. Yang, and Z. Jiang, "Mid-Infrared Emission Characteristic and Energy Transfer of Ho³⁺-Doped Tellurite Glass Sensitized by Tm³⁺," *Journal of Fluorescence* **17**(3): 301–307, 2007.
173. M. Eichhorn, and S. D. Jackson, "Comparative Study of Continuous Wave Tm³⁺-Doped Silica and Fluoride Fiber Lasers," *Appl. Phys. B: Lasers Opt.* **90**(1):35–41, 2008.
174. B. Richards, Y. Tsang, D. Binks, J. Lousteau, and A. Jha, "Efficient ~2 μ m Tm³⁺ Doped Tellurite Fiber Laser," *Opt. Lett.* **33**(4):402–404, 2008.
175. S. Jiang, M. J. Myers, D. L. Rhonehouse, S. J. Hamlin, J. D. Myers, U. Griebner, R. Koch, and H. Schonngel, "Ytterbium-Doped Phosphate Laser Glasses," in *Solid State Lasers VI: SPIE*, San Jose, Calif., 1997.

176. Y. W. Lee, S. Sinha, M. J. F. Digonnet, R. L. Byer, and S. Jiang, "20 W Single-Mode Yb³⁺-Doped Phosphate Fiber Laser," *Opt. Lett.* **31**(22):3255–3257, 2006.
177. S. Wei, M. Leigh, Z. Jie, A. Zhidong Yao, and A. Shibin Jiang, "Photonic Narrow Linewidth GHz Source Based on Highly Codoped Phosphate Glass Fiber Lasers in a Single MOPA Chain," *Photon. Technol. Lett. IEEE* **20**(2):69–71, 2008.
178. N. P. Barnes, B. M. Walsh, D. J. Reichle, R. J. Deyoung, and S. Jiang, "Tm:Germanate Fiber Laser: Tuning and Q-Switching," *Appl. Phys. B: Lasers Opt.* **89**(2):299–304, 2007.
179. M. Eichhorn, "High-Peak-Power Tm-Doped Double-Clad Fluoride Fiber Amplifier," *Opt. Lett.* **30**(24):3329–3331, 2005.
180. M. Eichhorn, "Development of a High-Pulse-Energy Q-Switched Tm-Doped Double-Clad Fluoride Fiber Laser and Its Application to the Pumping of Mid-IR Lasers," *Opt. Lett.* **32**(9):1056–1058, 2007.
181. S. D. Jackson, and G. Anzueto-Sanchez, "Chalcogenide Glass Raman Fiber Laser," *Appl. Phys. Lett.* **88**(22):221106-3, 2006.
182. A. F. El-Sherif, and T. A. King, "Dynamics and Self-Pulsing Effects in Tm³⁺-Doped Silica Fibre Lasers," *Opt. Commun.* **208**(4–6):381–389, 2002.
183. T. McComb, V. Sudesh, and M. Richardson, "Volume Bragg Grating Stabilized Spectrally Narrow Tm Fiber Laser," *Opt. Lett.* **33**(8):881–883, 2008.
184. W. A. Clarkson, N. P. Barnes, P. W. Turner, J. Nilsson, and D. C. Hanna, "High-Power Cladding-Pumped Tm-Doped Silica Fiber Laser with Wavelength Tuning from 1860 to 2090 nm," *Opt. Lett.* **27**(22):1989–1991, 2002.
185. D. Y. Shen, J. K. Sahu, and W. A. Clarkson, "Highly Efficient Er, Yb-Doped Fiber Laser with 188W Free-Running and > 100 W Tunable Output Power," *Opt. Exp.* **13**(13):4916–4921, 2005.
186. O. M. Efimov, L. B. Glebov, and V. I. Smirnov, "High-Frequency Bragg Gratings in a Photothermorefractive Glass," *Opt. Lett.* **25**(23):1693–1695 2000.
187. P. Jelger, and F. Laurell, "Efficient Skew-Angle Cladding-Pumped Tunable Narrow-Linewidth Yb-Doped Fiber Laser," *Opt. Lett.* **32**(24):3501–3503, 2007.
188. J. Yoonchan, C. Alegria, J. K. Sahu, L. Fu, M. Ibsen, C. Codemard, M. Mokhtar, and J. Nilsson, "A 43-W C-Band Tunable Narrow-Linewidth Erbium-Ytterbium Codoped Large-Core Fiber Laser," *Photon. Technol. Lett. IEEE* **16**(3):756–758, 2004.
189. N. Jovanovic, M. Åslund, A. Fuerbach, S. D. Jackson, G. D. Marshall, and M. J. Withford, "Narrow Linewidth, 100 W cw Yb³⁺-Doped Silica Fiber Laser with a Point-by-Point Bragg Grating Inscribed Directly into the Active Core," *Opt. Lett.* **32**(19):2804–2806, 2007.
190. P. Jelger, and F. Laurell, "Efficient Narrow-Linewidth Volume-Bragg Grating-Locked Nd:Fiber Laser," *Opt. Exp.* **15**(18):11336–11340, 2007.
191. J. W. Kim, P. Jelger, J. K. Sahu, F. Laurell, and W. A. Clarkson, "High-Power and Wavelength-Tunable Operation of an Er, Yb Fiber Laser Using a Volume Bragg Grating," *Opt. Lett.* **33**(11):1204–1206, 2008.
192. P. Jelger, P. Wang, J. K. Sahu, F. Laurell, and W. A. Clarkson, "High-Power Linearly-Polarized Operation of a Cladding-Pumped Yb Fibre Laser Using a Volume Bragg Grating for Wavelength Selection," *Opt. Exp.* **16**(13):9507–9512, 2008.
193. S. Tibuleac, and R. Magnusson, "Reflection and Transmission Guided-Mode Resonance Filters," *J. Opt. Soc. Am. A* **14**(7):1617–1626, 1997.
194. A. Mehta, R. C. Rumpf, Z. A. Roth, and E. J. Johnson, "Guided Mode Resonance Filter as a Spectrally Selective Feedback Element in a Double-Cladding Optical Fiber Laser," *Photon. Technol. Lett. IEEE* **19**(24):2030–2032, 2007.
195. J. Geng, J. Wu, S. Jiang, and J. Yu, "Efficient Operation of Diode-Pumped Single-Frequency Thulium-Doped Fiber Lasers near 2 μ m," *Opt. Lett.* **32**(4):355–357, 2007.
196. Y. Jeong, J. Nilsson, J. K. Sahu, D. B. S. Soh, C. Alegria, P. Dupriez, C. A. Codemard, et al., "Single-Frequency, Single-Mode, Plane-Polarized Ytterbium-Doped-fiber Master Oscillator Power Amplifier Source with 264 W of Output Power," *Opt. Lett.* **30**(5):459–461, 2005.
197. D. V. Gapontsev, N. S. Platonov, M. Meleshkevich, O. Mishechkin, O. Shikurikin, S. Agger, P. Varming, and J. H. Poylsen, "20 W Single-Frequency Fiber Laser Operating at 1.93 μ m," in *CLEO 2007*, Baltimore, Md., 2007.
198. Y. Jeong, J. K. Sahu, D. B. S. Soh, C. A. Codemard, and J. Nilsson, "High-Power Tunable Single-Frequency Single-Mode Erbium-Ytterbium Codoped Large-Core Fiber Master-Oscillator Power Amplifier Source," *Opt. Lett.* **30**(22):2997–2999, 2005.

199. G. P. Agrawal, *Nonlinear Fiber Optics*, Academic, San Diego, Calif., 1995.
200. M. O'Connor, and F. DiTeodoro, "Fiber Lasers in Defense: Fibers, Components and System Design Considerations," in *DEPS SSDLTR Conference Short Course*, Los Angeles, Calif., 2007.
201. Y. Jeong, J. Sahu, M. Laroche, W. A. Clarkson, K. Furusawa, D. J. Richardson, and J. Nilsson, "120 W Q-Switched Cladding Pumped Yb Doped Fiber Laser," in *CLEO Europe*, 2003. Munich, Germany.
202. J. Limpert, N. Deguil-Robin, S. Petit, I. Manek-Hönniger, F. Salin, P. Rigail, C. Hönniger, and E. Mottay, "High Power Q-Switched Yb-Doped Photonic Crystal Fiber Laser Producing Sub-10 ns Pulses," *Appl. Phys. B: Lasers Opt.* **81**(1):19–21, 2005.
203. O. Schmidt, J. Rothhardt, F. Röser, S. Linke, T. Schreiber, K. Rademaker, J. Limpert, S. Ermeneux, P. Yvernault, F. Salin, and A. Tünnermann, "Millijoule Pulse Energy Q-Switched Short-Length Fiber Laser," *Opt. Lett.* **32**(11):1551–1553, 2007.
204. M. Eichhorn, and S. D. Jackson, "High-Pulse-Energy Actively Q-Switched Tm³⁺-Doped Silica 2 μ m Fiber Laser Pumped at 792 nm," *Opt. Lett.* **32**(19):2780–2782, 2007.
205. J. Y. Huang, S. C. Huang, H. L. Chang, K. W. Su, Y. F. Chen, and K. F. Huang, "Passive Q Switching of Er-Yb Fiber Laser with Semiconductor Saturable Absorber," *Opt. Exp.* **16**(5):3002–3007, 2008.
206. R. Paschotta, R. Häring, E. Gini, H. Melchior, U. Keller, H. L. Offerhaus, and D. J. Richardson, "Passively Q-Switched 0.1-mJ Fiber Laser System at 1.53 μ m," *Opt. Lett.* **24**(6):388–390, 1999.
207. F. Z. Qamar, and T. A. King, "Passive Q-Switching of the Tm-Silica Fibre Laser near 2 μ m by a Cr²⁺:ZnSe Saturable Absorber Crystal," *Opt. Commun.* **248**(4–6):501–508, 2005.
208. S. D. Jackson, "Passively Q-Switched Tm³⁺-Doped Silica Fiber Lasers," *Appl. Opt.* **46**(16):3311–3317, 2007.
209. M. V. Andrés, "Actively Q-Switched All-Fiber Lasers," *Laser Phys. Lett.* **5**(2):93–99, 2008.
210. Z. Yu, W. Margulis, O. Tarasenko, H. Knappe, and P. Y. Fonjallaz, "Nanosecond Switching of Fiber Bragg Gratings," *Opt. Exp.* **15**(22):14948–14953, 2007.
211. H. Shuling, "Stable NS Pulses Generation from Cladding-Pumped Yb-Doped Fiber Laser," *Microwave and Opt. Technol. Lett.* **48**(12):2442–2444, 2006.
212. K. Kieu, and M. Mansuripur, "Active Q Switching of a Fiber Laser with a Microsphere Resonator," *Opt. Lett.* **31**(24):3568–3570, 2006.
213. W. Koechner, *Solid-State Laser Engineering*, Springer, New York, 1999.
214. P. Myslinski, J. Chrostowski, J. A. Koningstein, and J. R. Simpson, "Self Mode-Locking in a Q-Switched Erbium-Doped Fiber Laser," *Appl. Opt.* **32**(3):286, 1993.
215. B. N. Upadhyaya, U. Chakravarty, A. Kuruvilla, K. Thyagarajan, M. R. Shenoy, and S. M. Oak, "Mechanisms of Generation of Multi-Peak and Mode-Locked Resembling Pulses in Q-Switched Yb-Doped Fiber Lasers," *Opt. Exp.* **15**(18):11576–11588, 2007.
216. M. -Y. Cheng, Y. -C. Chang, A. Galvanauskas, P. Mamidipudi, R. Changkakoti, and P. Gatchell, "High-Energy and High-Peak-Power Nanosecond Pulse generation with Beam Quality Control in 200- μ m Core Highly Multimode Yb-Doped Fiber amplifiers," *Opt. Lett.* **30**(4):358–360, 2005.
217. S. A. George, K. -C. Hou, K. Takenoshita, A. Galvanauskas, and M. C. Richardson, "13.5 nm EUV Generation from Tin-Doped Droplets Using a Fiber Laser," *Opt. Exp.* **15**(25):16348–16356, 2007.
218. A. Liem, J. Limpert, H. Zellmer, and A. Tünnermann, "100-W Single-Frequency Master-Oscillator Fiber Power Amplifier," *Opt. Lett.* **28**(17):1537–1539, 2003.
219. J. Limpert, S. Höfer, A. Liem, H. Zellmer, A. Tünnermann, S. Knoke, and H. Voelckel, "100-W Average-Power, High-Energy Nanosecond Fiber Amplifier," *Appl. Phys. B: Lasers Opt.* **75**(4):477–479, 2002.
220. P. E. Britton, H. L. Offerhaus, D. J. Richardson, P. G. R. Smith, G. W. Ross, and D. C. Hanna, "Parametric Oscillator Directly Pumped by a 1.55- μ m Erbium-Fiber Laser," *Opt. Lett.* **24**(14):975–977, 1999.
221. S. Desmoulin, and F. Di Teodoro, "Watt-Level, High-Repetition-Rate, Mid-Infrared Pulses Generated by Wavelength Conversion of an Eye-Safe Fiber Source," *Opt. Lett.* **32**(1):56–58, 2007.
222. M. Savage-Leuchs, E. Eisenberg, A. Liu, J. Henrie, and M. Bowers, "High-Pulse Energy Extraction with High Peak Power from Short-Pulse Eye Safe All-Fiber Laser System," in *Fiber Lasers III: Technology, Systems, and Applications*: SPIE, San Jose, Calif., 2006.
223. L. E. Nelson, D. J. Jones, K. Tamura, H. A. Haus, and E. P. Ippen, "Ultrashort Pulse Fiber Ring Lasers," *Appl. Phys. B* **65**(2):277–294, 1997.

224. K. Tamura, H. A. Haus, and E. P. Ippen, "Self-Starting Additive Pulse Mode-Locked Erbium Fibre Ring Laser," *Electron. Lett.* **28**(24):2226–2228, 1992.
225. K. Tamura, E. P. Ippen, H. A. Haus, and L. E. Nelson, "77-fs Pulse Generation from a Stretched-Pulse Mode-Locked All-Fiber Ring Laser," *Opt. Lett.* **18**(13):1080, 1993.
226. A. Chong, W. H. Renninger, and F. W. Wise, "All-Normal-Dispersion Femtosecond Fiber Laser with Pulse Energy above 20 nJ," *Opt. Lett.* **32**(16):2408–2410, 2007.
227. L. M. Zhao, D. Y. Tang, and J. Wu, "Gain-Guided Soliton in a Positive Group-Dispersion Fiber Laser," *Opt. Lett.* **31**(12):1788–1790, 2006.
228. F. Ö. Ilday, J. R. Buckley, W. G. Clark, and F. W. Wise, "Self-Similar Evolution of Parabolic Pulses in a Laser," *Phys. Rev. Lett.* **92**(21):213902, 2004.
229. K. Tamura, E. P. Ippen, and H. A. Haus, "Pulse Dynamics in Stretched-Pulse Fiber Lasers," *Appl. Phys. Lett.* **67**(2):158–160, 1995.
230. C. Lecaplain, C. Chédot, A. Hideur, B. Ortaç, and J. Limpert, "High-Power All-Normal-Dispersion Femtosecond Pulse Generation from a Yb-Doped Large-Mode-Area Microstructure Fiber Laser," *Opt. Lett.* **32**(18):2738–2740, 2007.
231. B. Ortaç, J. Limpert, and A. Tünnermann, "High-Energy Femtosecond Yb-Doped Fiber Laser Operating in the Anomalous Dispersion Regime," *Opt. Lett.* **32**(15):2149–2151, 2007.
232. B. Ortaç, O. Schmidt, T. Schreiber, J. Limpert, A. Tünnermann, and A. Hideur, "High-Energy Femtosecond Yb-Doped Dispersion Compensation Free Fiber Laser," *Opt. Exp.* **15**(17):10725–10732, 2007.
233. R. Herda, S. Kivist, O. G. Okhotnikov, A. Kosolapov, A. Levchenko, S. Semjonov, and E. Dianov, "Environmentally Stable Mode-Locked Fiber Laser with Dispersion Compensation by Index-Guided Photonic Crystal Fiber," *Photon. Technol. Lett. IEEE* **20**(3):217–219, 2008.
234. J. C. Diels and W. Rudolph, *Ultrashort Laser Pulse Phenomena*, Academic Press, New York, 1996.
235. A. E. Siegman, *Lasers*, University Science Books, New York, 1986.
236. U. Keller, K. J. Weingarten, F. X. Kartner, D. Kopf, B. Braun, I. Jung, R. Fluck, C. Honninger, N. Matuschek, and J. Aus Der Au, "Semiconductor Saturable Absorber Mirrors (SESAM) for Femtosecond to Nanosecond Pulse Generation in Solid-State Lasers," *IEEE J. Sel. Top. Quant. Electron.* **2**(3):435–453, 1996.
237. H. Kataura, Y. Kumazawa, Y. Maniwa, I. Umez, S. Suzuki, Y. Ohtsuka, and Achiba, "Optical Properties of Single-Wall Carbon Nanotubes," *Synthetic Metals* **103**:2555–2558, 1999.
238. J. W. Nicholson, R. S. Windeler, and D. J. DiGiovanni, "Optically Driven Deposition of Single-Walled Carbon-Nanotube Saturable Absorbers on Optical Fiber End-Faces," *Opt. Exp.* **15**(15):9176–9183, 2007.
239. S. Y. Set, H. Yaguchi, Y. Tanaka, and M. Jablonski, "Laser Mode Locking Using a Saturable Absorber Incorporating Carbon Nanotubes," *J. Lightwave Technol.* **22**(1):51, 2004.
240. L. Vivien, P. Lancon, F. Hache, D. A. Riehl, and E. A. Anglaret, "Pulse Duration and Wavelength Effects on Optical Limiting Behaviour in Carbon Nanotube Suspensions," in *Lasers and Electro-Optics Europe, 2000. Conference Digest.* 2000.
241. L. P. Shen, W. P. Huang, G. X. Chen, and S. Jian, "Design and Optimization of Photonic Crystal Fibers for Broad-Band Dispersion Compensation," *Photon. Technol. Lett. IEEE* **15**(4):540–542, 2003.
242. D. Strickland, and G. Mourou, "Compression of Amplified Chirped Optical Pulses," *Opt. Commun.* **55**(6):447–449, 1985.
243. F. Röser, D. Schimpf, O. Schmidt, B. Ortaç, K. Rademaker, J. Limpert, and Tünnermann, "90 W Average Power 100 fJ Energy Femtosecond Fiber Chirped-Pulse Amplification System," *Opt. Lett.* **32**(15):2230–2232, 2007.
244. F. Röser, J. Rothhard, B. Ortac, A. Liem, O. Schmidt, T. Schreiber, J. Limpert, and Tünnermann, "131 W 220 fs Fiber Laser System," *Opt. Lett.* **30**(20):2754–2756, 2005.
245. F. Röser, T. Eidam, J. Rothhardt, O. Schmidt, D. N. Schimpf, J. Limpert, and A. Tünnermann, "Millijoule Pulse Energy High Repetition Rate Femtosecond Fiber Chirped-Pulse Amplification System," *Opt. Lett.* **32**(24):3495–3497, 2007.
246. A. Galvanauskas, "Mode-Scalable Fiber-Based Chirped Pulse Amplification Systems," *IEEE J. Sel. Top. Quant. Electron.* **7**(4):504–517, 2001.
247. K.-H. Liao, et al., "Generation of Hard X-Rays Using an Ultrafast Fiber Laser System," *Opt. Exp.* **15**(21):13942–13948, 2007.

- 248. G. Chang, et al., "50-W Chirped Volume Bragg Grating Based Fiber CPA at 1055 nm," in *SSDLTR*. Los Angeles, Calif., 2007.
- 249. K.-H. Liao, A. G. Mordovanakis, B. Hou, G. Chang, M. Rever, G. A. Mourou, J. Nees, and Galvanauskas, "Large-Aperture Chirped Volume Bragg Grating Based Fiber CPA System," *Opt. Exp.* **15**(8):4876–4882, 2007.
- 250. "Pioneering Ultrafast Fiber Laser Technology," available: www.imra.com, accessed on: Dec. 2008.

PART

5

X-RAY AND
NEUTRON OPTICS

This page intentionally left blank

SUBPART

5.1

INTRODUCTION AND APPLICATIONS

This page intentionally left blank

AN INTRODUCTION TO X-RAY AND NEUTRON OPTICS

Carolyn MacDonald

*University at Albany
Albany, New York*

26.1 HISTORY

X rays have a century of history of medical and technological applications. Optics have come to play an important role in many capacities. One of Roentgen's earliest observations, shortly after his discovery of x rays in 1895,¹ was that while the rays were easily absorbed by some materials, they did not strongly refract. Standard optical lenses, which require weak absorption and strong refraction, were therefore not useful for manipulating the rays. New techniques were quickly developed. In 1914, van Laue was awarded the Nobel Prize for demonstrating the diffraction of x rays by crystals. By 1929, total reflection at grazing incidence had been used to deflect x rays.² The available optics for x rays still can be classified by those three phenomena: refraction, diffraction, and total reflection. Surprisingly, given that the physics governing these optics has been well known for nearly a century, there has been a recent dramatic increase in the availability, variety, and performance of x-ray optics for a wide range of applications.

An increasing interest in x-ray astronomy was one of the major forces for the development of x-ray optics in the latter half of the last century. Mirror systems similar to those developed for astronomy also proved useful for synchrotron beam lines. Just as x-ray tubes were an accidental offshoot of cathode ray research, synchrotron x-ray sources were originally a parasite of particle physics. The subsequent development of synchrotrons with increasing brightness and numbers of beam lines have created whole new arrays of x-ray tools and a consequent demand for an increasing array of optics. The rapid development of x-ray optics has also been symbiotic with the development of detectors and of compact sources. Detectors developed for particle physics, medicine, and crystallography have found applications across fields. Similarly, the increasing capability of x-ray systems has stimulated the development of new science with evergrowing requirements for intensity, coherence, and spatial and energy resolution. X-ray diffraction and fluorescence were early tools of the rapid development of materials science after World War II, but have been greatly advanced to meet the demands of the shrinking feature sizes and allowed defect levels in semiconductors. X-ray diffraction, especially the development of dedicated synchrotron beam lines, has also been stimulated by the growing demands for rapid protein crystallography for biophysics and pharmaceutical development.

The new abundance of x-ray optics, sources, and detectors requires a fresh look at the problem of optimizing a wide range of x-ray and neutron applications. The development of x-ray technology has also advanced neutron science because a number of the optics and detectors are either applicable to neutrons or have inspired the development of neutron technology.

One question that arises immediately in a discussion of x-ray phenomena is precisely what spectral range is included in the term. Usage varies considerably by discipline, but for the purposes of this volume, the x-ray spectrum is taken to be roughly from 1 to 100 keV in photon energy (1.24 to 0.0124-nm wavelength). The range is extended down into the hard EUV to include some microscopy and astronomical optics, and upward to include nuclear medicine.

26.2 X-RAY INTERACTION WITH MATTER

X rays are applicable to such a wide variety of areas because they are penetrating but interacting, have wavelengths on the order of atomic spacings, and have energies on the order of core electronic energy levels for atoms.

X-Ray Production

X rays are produced primarily by the acceleration of charged particles, the knock out of core electrons, or black body and characteristic emission from very hot sources such as laser-generated plasmas or astronomical objects. The production of x rays by accelerated charges includes incoherent emission such as bremsstrahlung radiation in tube sources and coherent emission by synchrotron undulators or free electron lasers. Highly coherent emission can also be created by pumping transitions between levels in ionic x-ray lasers.

The creation of x rays by the knock out of core electrons is the mechanism for the production of the characteristic lines in the spectra from conventional x-ray tube sources. The incoming electron knocks out a core electron, creating a vacancy, which is quickly filled by an electron dropping down from an outer shell. The energy difference between the outer shell energy level and the core energy level is emitted in the form of an x-ray photon. This is also the origin of the characteristic lines used to identify elemental composition in x-ray fluorescence, described in Chap. 29, and for x-ray spectroscopy, described in Chap. 30. X-ray sources are described in Subpart 5.4 of this section, in Chaps. 54 to 59. Source coherence, and coherence requirements, are discussed in Chap. 27.

Refraction

In the x-ray regime, the real part of the index of refraction of a solid can be simply approximated by³

$$n_r = \text{Re}\{\sqrt{\kappa}\} = \text{Re}\left\{\sqrt{\frac{\epsilon}{\epsilon_0}}\right\} \cong \sqrt{1 - \frac{\omega_p^2}{\omega^2}} \cong 1 - \delta \quad (1)$$

where n is the index of refraction, ϵ is the dielectric constant of the solid, ϵ_0 is the vacuum dielectric constant, κ is their ratio, ω is the photon frequency, and ω_p is the plasma frequency of the material. The plasma frequency, which typically corresponds to tens of electron volts, is given by

$$\omega_p^2 = \frac{Ne^2}{m\epsilon_0} \quad (2)$$

where N is the electron density of the material, and e and m are the charge and mass of the electron. For x rays, the relevant electron density is the total density, including core electrons. Thus changes in x-ray optical properties cannot be accomplished by changes in the electronic levels, in the manner in which optical properties of materials can be manipulated for visible light. The x-ray optical properties

are determined by the density and atomic numbers of the elemental constituents. Tables of the properties of most of the elements are presented in Chap. 36. Because the plasma frequency is very much less than the photon frequency, the index of refraction is slightly less than one.

Absorption and Scattering

The imaginary part of the index of refraction for x rays arises from photoelectric absorption. This absorption is largest for low energy x rays and has peaks at energies resonant with core ionization energies of the atom. X rays can also be deflected out of the beam by incoherent or coherent scattering from electrons. Coherent scattering is responsible for the decrement to the real part of the index of refraction given in Eq. (1). Scattering from nearly free electrons is called *Thompson scattering*, and from tightly bound electrons, *Rayleigh scattering*. The constructive interference of coherent scattering from arrays of atoms constitutes *diffraction*. *Incoherent*, or Compton, scattering occurs when the incident photon imparts energy and momentum to the electron. Compton scattering becomes increasingly important for high-energy photons and thick transparent media.

26.3 OPTICS CHOICES

Slits and Pin Holes

Almost all x-ray systems, whether or not they use more complex optics, contain slits or apertures. Some aperture systems have considerable technological development. Most small sample diffraction systems employ long collimators designed to reduce the background noise from scattered direct beam reaching the detector. For θ - 2θ measurements, Soller slits, arrays of flat metal plates arranged parallel to the beam direction, are often placed after the sample to further reduce the background. Diffraction applications are described in Chap. 28.

Lead hole collimators and pinholes specifically designed for high photon energies are employed for nuclear medicine, with pinhole sizes down to tens of microns. Nuclear medicine is discussed in Chap. 32.

Refractive Optics

One consequence of Eq. (1) is that the index of refraction of all materials is very close to unity in the x-ray regime. Thus Snell's law implies that there is very little refraction at the interface,

$$(1) \sin\left(\frac{\pi}{2} - \theta_1\right) = n \sin\left(\frac{\pi}{2} - \theta_2\right) \quad (3)$$

where the first medium is vacuum and the second medium has index n , as shown in Fig. 1. In x-ray applications the angles θ are measured from the surface, not the normal to the surface. If n is very close to one, θ_1 is very close to θ_2 . Thus, refractive optics are more difficult to achieve in the x-ray regime. The lens maker's equation,

$$f = \frac{R/2}{n - 1} \quad (4)$$

gives the relationship between the focal length of a lens f and the radius of curvature of the lens R . The symmetric case is given. Because n is very slightly less than one, the focal length produced even by very small negative radii (convex) lenses is rather long. This can be overcome by using large

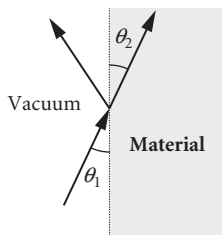


FIGURE 1 Refraction at a vacuum-to-material interface.

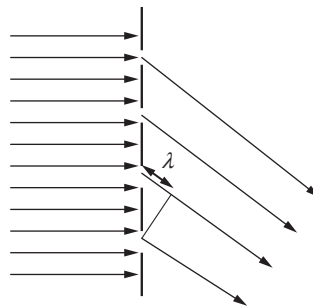


FIGURE 2 Schematic of constructive interference from a transmission grating. λ is the wavelength of the radiation. The extra path length indicated must be an integral multiple of λ . Most real gratings are used in reflection mode.

numbers of surfaces in a compound lens. Because the radii must still be small, the aperture of the lens cannot be large, and refractive optics are generally better suited to narrow synchrotron beams than isotropic point sources. Refractive optics is described in Chap. 37.

Diffractive and Interference Optics

The coherent addition of radiation from multiple surfaces or apertures can only occur for a very narrow wavelength bandwidth. Thus, diffractive and interference optics such as gratings, crystals, zone plates, multilayers, and Laue lenses, described in Chaps. 38 to 43, respectively, are all wavelength selective. Such optics are often used as monochromator to select a particular wavelength range from a white beam source.

To achieve constructive interference, the spacing of a grating must be arranged so that the radiation from successive sources is in phase, as shown in Fig. 2. The circular apertures of zone plates are similar to the openings in a transmission grating. The superposition of radiation from the circular apertures results in focusing of the radiation to points on the axis.

Diffractive optics operate on the same principal, coherent superposition of many rays.⁴ The most common diffractive optic is the crystal. Arranging the beam angle to the plane, θ , and plane spacing d as shown in Fig. 3, so that the rays reflecting from successive planes are in phase (and ignoring refraction) yields the familiar Bragg's law,

$$n\lambda = 2d \sin \theta \quad (5)$$

where n is an integer and λ is the wavelength of the x ray. Bragg's law cannot be satisfied for wavelengths greater than twice the plane spacing, so crystal optics are limited to low wavelength, high energy, x rays. Multilayers are "artificial crystals" of alternating layers of materials. The spacing, which is the thickness of a layer pair, replaces d in Eq. (5), and can be much larger than crystalline plane spacings. Multilayers are therefore effective for lower energy x rays. Diffraction can also be used in transmission, or Laue, mode, as shown in Fig. 4.

Reflective Optics

Using Snell's law, Eq. (3), the angle inside a material is smaller than the incident angle in vacuum. For incident angles less than a critical angle θ_c , no wave can be supported in the medium and the

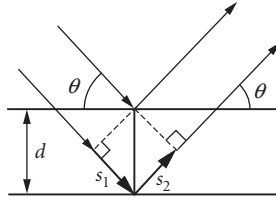


FIGURE 3 Pictorial representation of Bragg diffraction from planes with spacing d . The extra path length, $(s_1 + s_2)$, must be an integral multiple of λ .

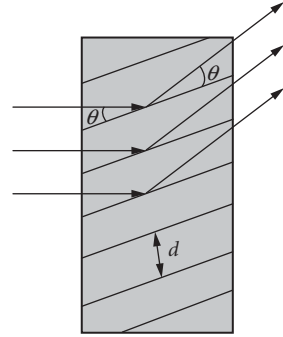


FIGURE 4 Laue transmission diffraction.

incident ray is totally externally reflected. The critical angle, the largest incident angle for total reflection, is given by

$$\sin\left(\frac{\pi}{2} - \theta_c\right) = n \sin\left(\frac{\pi}{2}\right) \quad (6)$$

Thus, using small angle approximations and Eq. (1)

$$\theta_c \approx \frac{\omega_p}{\omega} \quad (7)$$

Because the plasma frequency is very much less than the photon frequency, total external reflection occurs for very small grazing incidence angles. This phenomenon is used extensively for single reflection mirrors for synchrotrons, x-ray microscopes, and x-ray telescopes. Mirrors are described in Chaps. 44 to 47 and 51. To increase the angle of incidence, mirrors are often coated with metals, or with multilayers (although, because the multilayer depends on interference, the optic will no longer have a broadband response).

Arrays of mirrors, such as multifoil or pore optics, are described in Chaps. 48 and 49. Glass capillaries, described in Chap. 52, can be used as single bounce-shaped mirrors, or as multiple-bounce light pipes transport or focus the beam. Multiple reflections are used to transport x rays in polycapillary arrays, described in Chap. 53. Because a mirror is often the first optic in a synchrotron beam line, heat load issues are significant. For a number of synchrotron and other high-flux applications, radiation hardness and thermal stability are important considerations. A large body of experience has been developed for high-heat-load synchrotron mirrors and crystals.^{5,6} Many optics such as zone plates, microscopy objectives, and glass capillary tubes are routinely used in synchrotron beam lines and are stable over acceptable flux ranges. Adaptive optics used to mitigate the effects of thermal changes are described in Chap. 50.

26.4 FOCUSING AND COLLIMATION

Optics Comparisons

A variety of x-ray optics choices exist for collimating or focusing x-ray beams. The best choice of the optic depends to a large extent on the geometry of the sample to be measured, the information desired from the measurement, and the source geometry and power. No one optic can provide

the best resolution, highest intensity, easiest alignment, and shortest data acquisition time for all samples. A true comparison of two optics for a particular application requires careful analysis of the sample and measurement requirements, and adjustment of the source and optic design for the application. No global comparison of all optics for all applications is possible.

Four optics commonly used for collimation or focusing from x-ray tubes are bent crystals, multilayers, nested cones, and polycapillary optics. Bent crystals collect radiation from a point source and diffract it into a nearly monochromatic collimated beam. Doubly bent crystals or two singly bent crystals are required for two-dimensional collimation or focusing. Multilayer optics also work by diffraction, although in this case from the periodicity of the artificial compositional variation imposed in the multilayer. The beam is less monochromatic than for bent crystals. “Supermirrors,” or “Goebel mirrors,” are multilayers with graded or irregular spacing, and have wider energy bandwidths than multilayers.

Mirrors are broad band optics, but because curved mirrors are single-bounce optics, their maximum angular deflection is limited to the critical angle for their metallic coating, which can be about 10 mrad at 8 keV. The total capture angle is then determined by the length of the mirror. The output divergence is determined by the length of the optic and the source size. One solution to increase the capture angle is to “nest” multiple optics. Nested parabolic mirrors have smaller output divergences than nested cones. Polycapillary optics are also array optics, containing hundreds of thousands glass tubes. Because the focal spot is produced by overlap, it cannot be smaller than the channel size.

Comparison of focusing optics for diffraction also requires a detailed analysis of the effect of the convergence angle on the diffracted signal intensity and so is very sample and measurement dependent. Decreasing the angle of convergence onto the sample improves the resolution and decreases the signal to noise ratio, but also decreases the diffracted signal intensity relative to a large angle.⁷ Conventional practice is to use a convergence angle less than the mosaicity of the sample. It is necessary that the beam cross-section at the sample be larger than the sample to avoid the difficulty of correcting for intensity variations with sample angle.

Focusing for Spatial Resolution

Diffraction effects limit the resolution of visible light optical systems to within an order of magnitude of the wavelength of the light. Thus, in principle, x-ray microscopy systems could have resolution many orders of magnitude better than optical light systems. Electrons also have very small wavelength, and electron microscopes have extremely high resolution. However, electrons are charged particles and necessarily have low penetration lengths into materials. X-ray microscopy is capable of very high resolution imaging of relatively thick objects, including wet samples. In practice, x-ray microscopy covers a range of several orders of magnitude in wavelength and spot size. Very high resolution is commonly obtained with Schwarzschild objectives,⁸ or zone plates.⁹ Because Schwarzschild objectives are used in normal incidence, they are essentially limited to the EUV region. Synchrotron sources are required to provide adequate flux. The resolution of zone plates is determined by the width of the outermost zone. Zone plates are easiest to make for soft x rays, where the thickness required to absorb the beam is small. However, zone plates with high aspect ratio have been demonstrated for hard x rays. Because the diameters of imaging zone plates are small, and the efficiencies are typically 10 percent for amplitude zone plates and 40 percent for phase zone plates, synchrotron sources are required. Very small spot sizes have also been demonstrated with multilayer Laue lenses and with refractive optics.

Single capillary tubes can also have output spot sizes on the order of 100 nm or less, and thus can be used for scanning microscopy or microanalysis. Capillary tubes with outputs this small also require synchrotron sources.

For larger spot sizes, a wider variety of sources and optics are applicable. Microscopes have been developed for laser-plasma sources with both Wolter¹⁰ and bent crystal optics,¹¹ with resolutions of a few microns. These optics, and also capillary and polycapillary optics and nested mirrors, with or without multilayer coatings, can produce spot sizes of a few tens of microns with laboratory tube sources. Optics designed to collect over large solid angles, such as bent crystals, graded multilayers or polycapillary optics, will produce the highest intensities. Bent crystals will yield monochromatic radiation; polycapillary optics can be used to produce higher intensity, but broader band radiation. Polycapillary

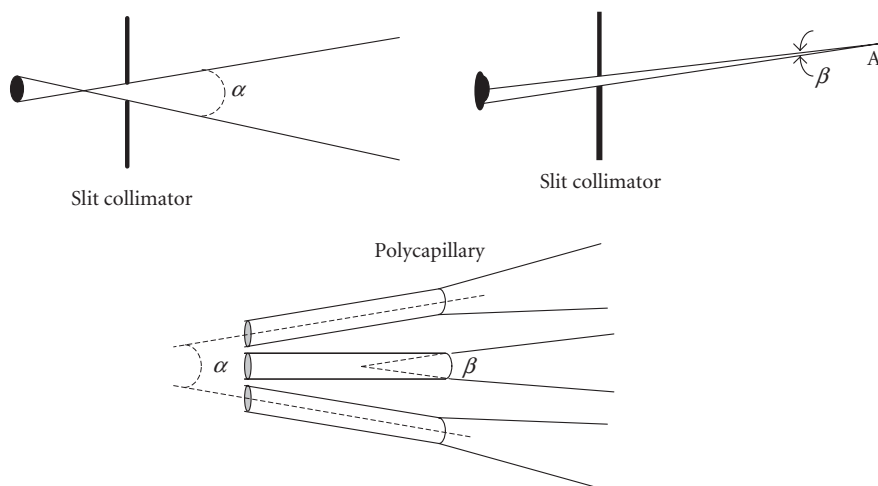


FIGURE 5 For either a slit collimator (top left and right) or an optic such as a polycapillary optic (bottom), the local divergence β seen at a point A is different from the angle subtended by the beam α .

optics have spot sizes no smaller than tens of microns, independent of source size. Mirrors and crystals will have smaller focal spot sizes for smaller sources and larger spot sizes for larger sources. Refractive optics are true imaging optics, with the potential for very small spots. Clearly, the optimal optic depends on the details of the measurement requirements and the sources available.

Collimation

As shown in Fig. 5, the output from a collimating optic has both global divergence α and local divergence β . Even if the global divergence is made very small by employing an optic, the local divergence is usually not zero. For grazing incidence reflection optics, the local divergence is generally given by the critical angle for reflection and can be increased by profile errors in the optic. The degree of collimation achieved by the optic is limited not only by technology, but by the thermodynamic constraint that the beam brightness cannot be increased by any optic.¹² Liouville's theorem states that increasing the density of states in phase space is a violation of the second law of thermodynamics. This implies that the six dimensional real space/momentum space volume occupied by the photons cannot be decreased. More simply, the angle-area product, that is, the cross-sectional area of the beam multiplied by the divergence of the beam, cannot be decreased without losing photons. The beam can never be brighter than the source. An idealized point source of x rays occupies zero area. X rays from such a source could theoretically be perfectly collimated into any chosen cross section. However, a real source has a finite size and so limits the degree of possible collimation. Small spot sources are required to produce bright, well-collimated beams.

26.5 REFERENCES

1. W. C. Röntgen, "On a New Form of Radiation," *Nature* **53**:274–276, January 23, 1896, English translation from *Sitzungsberichte der Würzburger Physik-med. Gesellschaft*, 1895.
2. Werner Ehrenberg and Felix Jentsch, "Über die Auslösung von Photoelektronen durch Röntgenstrahlen aus Metallsiegeln an der Grenze der Totalreflexion," *Zeitschrift für Physik* **54**(3, 4):227–235, March, 1929.

3. J. D. Jackson, *Classical Electrodynamics*, John Wiley & Sons, New York, pp. 227, 1962.
4. B. D. Cullity and B. D. Cullity, *Elements of X-Ray Diffraction*, Adison Wesley Longman, Reading, Mass., 1978.
5. A. Khounsary, *Advances in Mirror Technology for Synchrotron X-Ray and Laser Applications*, SPIE, Bellingham, Wash., vol. 3447, 1998.
6. A. T. Macrander and A. M. Khounsary (eds.), *High Heat Flux and Synchrotron Radiation Beamlines* (Proceedings Volume) SPIE, Bellingham, Wash., vol. 3151, 11 December, 1997, ISBN: 9780819425737.
7. U. W. Arndt, "Focusing Optics for Laboratory Sources in X-Ray Crystallography," *J. Appl. Cryst.* **23**:161–168, 1990.
8. F. Cerrina, "The Schwarzschild Objective," in *Handbook of Optics*, 3d ed., vol. V, M. Bass (ed.), McGraw-Hill, New York, 2009.
9. A. Michette, "Zone Plates," in *Handbook of Optics*, 3d ed., vol. V, M. Bass (ed.), McGraw-Hill, New York, 2009.
10. P. Trousses, P. Munsch, and J. J. Ferme, "Microfocusing between 1 and 5 keV with Wolter Type Optic," in *X-Ray Optics Design, Performance and Applications*, A. M. Khounsary, A. K. Freund, T. Ishikawa, G. Srajer, J. Lang, (eds.), SPIE, Bellingham, Wash., vol. 3773, pp. 60–69, 1999.
11. T. A. Pikuz, A. Y. Faenov, M. Fraenkel, et al., "Large-Field High Resolution X-Ray Monochromatic Microscope, Based on Spherical Crystal and High Repetition-Rate Laser-Produced Plasmas," in *EUV, X-Ray and Neutron Optics and Sources*, C. A. MacDonald, K. A. Goldberg, J. R. Maldonado, H. H. Chen-Mayer, S. P. Vernon (eds.), SPIE, Bellingham, Wash., vol. 3767, pp. 67–78, 1999.
12. D. L. Goodstein, *States of Matter*, Prentice-Hall, Englewood Cliffs, N.J., 1975.

COHERENT X-RAY OPTICS AND MICROSCOPY

Qun Shen

*National Synchrotron Light Source II
Brookhaven National Laboratory
Upton, New York*

27.1 GLOSSARY

$F(x, y)$	diffracted wave field amplitude on the detector image plane (x, y)
$q(X, Y)$	transmission function through a thin object
r	length of the position vector from point (X, Y) on the object plane to point (x, y) on the detector image plane
λ	x-ray wavelength
k	$2\pi/\lambda$ is the wave number
$\bar{q}(X, Y)$	distorted object with Fresnel zone phase factors embedded in the original object
$N_z = a^2/(4\lambda z)$	number of Fresnel zones on the object when looking back from the image plane

Traditionally the development of x-ray optics is based on ray-tracing wavefield propagation in geometric optics. This is true especially in the hard x-ray regime with energy in the multiple keV regime. With the availability of partially coherent x-ray sources such as those at modern synchrotrons, this situation has changed completely. Very often, some “artifacts” or features beyond ray-tracing can be observed in experiments due to the interference or phase effects from substantial spatial coherence of the synchrotron x-ray source. In this chapter, a brief outline is presented of a simple version of the wave propagation theory that can be used to evaluate coherent propagation of x-ray waves to take into account these effects. Based on optical reciprocity theorem, this type of coherent propagation is also required when evaluating x-ray focusing optics with the intention to achieve diffraction-limited performance.

In addition to coherent x-ray optics, there are substantial interests in the scientific community to use x-rays for imaging microscopic structures based on coherent wave propagation. For example, the success of structural science today is largely based on x-ray diffraction from crystalline materials. However, not all materials of interest are in crystalline forms; examples include the majority of membrane proteins and larger multidomain macromolecular assemblies, as well as many nanostructure specimens at their functioning levels. For these noncrystalline specimens, imaging at high spatial resolution offers an alternative, or the only alternative, to obtain any information on their internal structures. This topic is covered in the second part of this chapter.

27.2 INTRODUCTION

The spatial coherence length is defined as the transverse distance across the beam over which two parts of the beam have a fixed phase relationship. Using classical optics, the coherence length is $L_{\text{coherence}} = (\lambda/\beta)$ where λ is the wavelength of the radiation, and β is the local divergence, the angle subtended by the source. For typical laboratory sources the coherence length is only a few microns at several meters. However, for modern synchrotron sources β is much smaller and the coherence length is large enough to encompass optics such as zone plates which require coherent superposition, or entire samples for phase imaging.

27.3 FRESNEL WAVE PROPAGATION

In principle, imaging and diffraction or scattering are two optical regimes that are intrinsically interrelated based on Fresnel diffraction for wave propagation, which, under the first-order Born approximation,^{1,2} is

$$F(x, y) = \frac{i}{\lambda} \iint q(X, Y) \frac{e^{-ikr}}{r} dXdY \quad (1)$$

where $F(x, y)$ is the diffracted wave field amplitude, $q(X, Y)$ is the transmission function through a thin object, $r = [z^2 + (x - X)^2 + (y - Y)^2]^{1/2}$ is the length of the position vector from point (X, Y) on the object plane to point (x, y) on the detector image plane, λ is the x-ray wavelength, and $k = 2\pi/\lambda$ is the wave number.

Although widely used in optical and electron diffraction and microscopy,¹ the concept of Fresnel diffraction Eq. (1) has only recently been recognized in the broader x-ray diffraction community where traditionally far-field diffraction plus conventional radiography dominated the x-ray research field for the past century. This is because an essential ingredient for Fresnel-diffraction-based wave propagation is a substantial degree of transverse coherence in an x-ray beam, which had not been easily available until recent advances in partially coherent synchrotron and laboratory-based sources.

27.4 UNIFIED APPROACH FOR NEAR- AND FAR-FIELD DIFFRACTION

Coherent wave field propagation based on Fresnel diffraction Eq. (1) is usually categorized into two regimes: the near-field Fresnel or in-line holography regime, and the far-field Fraunhofer regime. A unified method for the evaluation of wave-field propagation in both regimes (Fig. 1) has been developed using the concept of *distorted object* in Fresnel Eq. (1), which can be applied both to Fraunhofer and Fresnel diffractions.

To introduce this method, we expand in Eq. (1), $r = [z^2 + (x - X)^2 + (y - Y)^2]^{1/2} \approx z + [(x - X)^2 + (y - Y)^2]/2z$ so that Eq. (1) becomes

$$F(x, y) = \frac{ie^{-ikz}}{\lambda z} \iint q(X, Y) e^{-ik \frac{(x-X)^2 + (y-Y)^2}{2z}} dXdY$$

Further expanding the terms in the exponential results in

$$F(x, y) = \frac{ie^{-ikR}}{\lambda R} \iint q(X, Y) e^{-\frac{i\pi}{\lambda z}(X^2 + Y^2)} e^{-\frac{i2\pi}{\lambda z}(xX + yY)} dXdY$$

where $R = (x^2 + y^2 + z^2)^{1/2}$. We now define a new *distorted object* $\bar{q}(X, Y)$ as follows

$$\bar{q}(X, Y) \equiv q(X, Y) e^{-\frac{i\pi}{\lambda z}(X^2 + Y^2)} \quad (2)$$

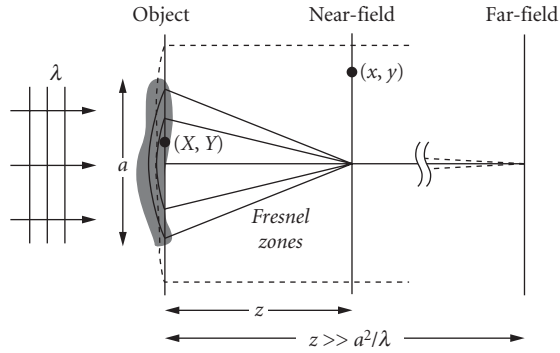


FIGURE 1 Schematic illustration of coherent x-ray wave propagation with a distorted object approach both for near-field Fresnel diffraction, where an object extends into multiple Fresnel zones (solid lines), and for far-field Fraunhofer diffraction, where an object occupies only the center of the first Fresnel zone (dashed lines). (See also color insert.)

and the scattered wave field $F(x, y)$ can then be expressed by a direct Fourier transform of this distorted object

$$F(x, y) = \frac{i e^{-ikR}}{\lambda R} \iint \bar{q}(X, Y) e^{-\frac{ik}{z}(xX+yY)} dXdY \quad (3)$$

Eq. (3) clearly shows that by embedding Fresnel zone construction into the distorted object, Eq. (2), a near-field diffraction pattern can be simply evaluated by a Fourier transform just like in the far-field approximation, with a momentum transfer $(Q_x, Q_y) = (kx/z, ky/z)$. Furthermore, it reduces to the familiar far-field result when $z \gg a^2/(4\lambda)$, where a is the transverse size of the object, since the extra Fresnel phase factor in Eq. (2) can be then approximated to unity. In general, the number of Fresnel phase zones of width π depends on distance z and is given by $N_z = a^2/(4\lambda z)$. Therefore, Eq. (3) can be used both in the near-field and in the far-field regimes, and this traditional but somewhat artificial partition of these two regimes is easily eliminated. Figure 2 shows some examples of calculated diffraction patterns at different detector to specimen distances.

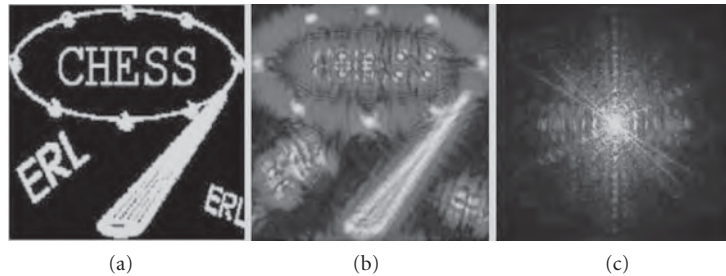


FIGURE 2 Simulated diffraction amplitudes $|F(x, y)|$, of an amplitude object (a) of $10 \mu\text{m} \times 10 \mu\text{m}$, with $\lambda = 1 \text{ \AA}$ x rays, at image-to-object distance (b) $z = 2 \text{ mm}$ and (c) $z = \infty$, using the unified distorted object approach Eq. (3) with $N_z = 500$ zones in (b) and $N_z = 0$ in (c). Notice that the diffraction pattern changes from noncentrosymmetric in the near-field (b) to centrosymmetric in the far-field (c). (See also color insert.)

27.5 COHERENT DIFFRACTION MICROSCOPY

It has been shown in recent years³ that an oversampled continuous coherent diffraction pattern from a nonperiodic object can be phased directly based on real space and reciprocal space constraints using an iterative phasing technique originally developed in optics.^{4,5} The oversampling condition requires a diffraction pattern be measured in reciprocal space at a Fourier interval finer than the Nyquist frequency used in all discrete fast Fourier transforms. Once such an oversampled diffraction pattern is obtained, as shown in Fig. 3, the iterative phasing method starts with a random set of phases for diffraction amplitudes, and Fourier transforms back and forth between diffraction amplitudes in reciprocal space and density in real space. In each iteration, the real space density is confined to within the finite specimen size and the square of diffraction amplitudes in reciprocal space is made equal to the experimentally measured intensities. This iterative procedure has proved to be a powerful phasing method for coherent diffraction imaging of nonperiodic specimens as a form of lensless x-ray microscopy.

One of the applications of the distorted object approach is that it extends the Fourier transform-based iterative phasing technique that works well in the far-field coherent diffraction imaging, into the regime of phasing near-field Fresnel diffraction or holographic images.⁶ Because the distorted object $\bar{q}(X, Y)$ differs from the original object $q(X, Y)$ by only a phase factor, which is known once the origin on the object is chosen, all real-space constraints applicable on $q(X, Y)$ can be transferred onto $\bar{q}(X, Y)$ in a straightforward fashion. In fact, most existing iterative phasing programs may be easily modified to accommodate the distorting phase factor in Eq. (2). A similar technique developed by Nugent et al.⁷ makes use of a curved wave illumination from a focusing x-ray optic in coherent diffraction imaging experiments and has demonstrated that the iterative phasing algorithm may converge much faster with a curved-beam illumination.

A significant recent development in coherent diffraction microscopy is the introduction of a scanning probe so that the coherent diffraction method can be applied to extended specimens.⁸ It has

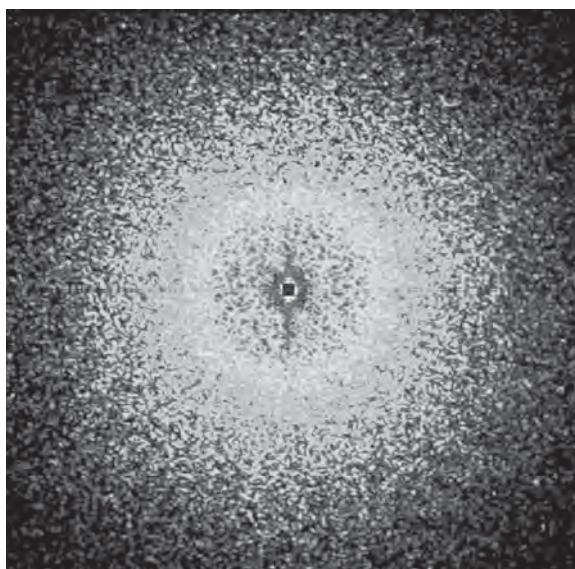


FIGURE 3 Example of a coherent x-ray diffraction pattern from a gold nanofoam specimen of $\sim 2 \mu\text{m}$ in size, using 7.35-keV coherent x rays. The corner of the image corresponds to $\sim 8 \text{ nm}$ spatial frequency. (See also color insert.)

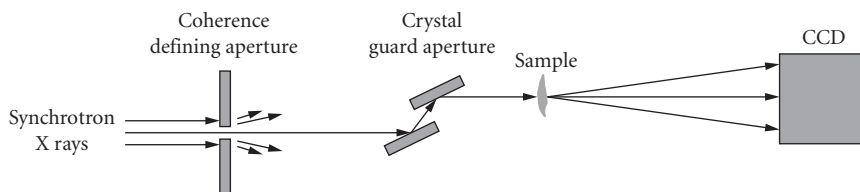


FIGURE 4 Concept of a perfect-crystal guard aperture in coherent diffraction imaging experiments for the purpose of eliminating unwanted parasitic scattering background in order to achieve high signal-to-noise in a diffraction pattern. (See also color insert.)

become apparent that the a combination of scanning x-ray microscopy with coherent diffraction may be the ultimate tool⁹ that scientists will be using in the coming years to image high resolution structures on nonperiodic specimens.

27.6 COHERENCE PRESERVATION IN X-RAY OPTICS

Coherent x-ray wavefield propagation has become an important consideration in many aspects of x-ray optics developments. There are many examples already published in the literature. For example, in order to evaluate the ultimate performance of x-ray focusing optics, it is essential to employ a coherent wave propagation theory from the x-ray optic to the focal spot which ultimately is diffraction limited.¹⁰

Another example where coherent wave propagation is needed is to preserve well-defined wavefronts through an x-ray optical system, which is often referred to as coherence preservation. For instance, one crucial issue in coherent x-ray diffraction imaging is how to increase the signal-to-noise ratio when measuring relatively weak diffraction intensities from a nonperiodic object. Based on coherent wave propagation, a crystal guard aperture concept has been developed¹¹ which makes use of a pair of multiple-bounce crystal optics to eliminate unwanted parasitic scattering background from the upstream coherence defining aperture (see Fig. 4). Recent experimental observation and theoretical analysis confirm the effectiveness of the crystal guard aperture method with coherence-preserved wave propagation through the crystal guard aperture and dramatically reduced scattering background in coherent x-ray diffraction images.¹¹

In summary, the development of coherent x-ray optics and x-ray analysis has become increasingly more important as the state-of-the-art x-ray sources and x-ray microscopic tools are becoming more readily available. It is expected that this field will continue to grow rapidly in order to satisfy the strong scientific interests in the community.

27.7 REFERENCES

1. J. M. Cowley, *Diffraction Physics*, 2nd ed., Elsevier Science Publisher, New York, 1990.
2. F. van der Veen and F. Pfeiffer, *J. Phys.: Condens. Matter* **16**:5003 (2004).
3. Original idea was proposed by D. Sayre, "Prospects for Long-Wavelength X-Ray Microscopy and Diffraction" in *Imaging Processes and Coherence in Physics*, M. Schlenker, M. Fink, J. P. Goedgebuer, C. Malgrange, J. C. Vienot, and R. H. Wade (eds.), *Springer Lect. Notes Phys.* **112**:229–35 (1980), Berlin: Springer. For a recent review, see J. Miao, et al., *Annu. Rev. Phys. Chem.* **59**:387–409 (2008).
4. J. R. Fienup, *Appl. Opt.* **21**:2758 (1982).

5. R. W. Gershberg and W. O. Saxton, *Optik* 25:237 (1972).
6. X. Xiao and Q. Shen, *Phys. Rev. B* 72:033101 (2005).
7. G. J. Williams, H. M. Quiney, B. B. Dhal, et al., *Phys. Rev. Lett.* 97:025506 (2006).
8. J. M. Rodenburg, A. C. Hurst, A. G. Cullis, et al., *Phys. Rev. Lett.* 98:034801 (2007).
9. P. Thibault et al., *Science* 321:379–382 (2008).
10. H. Yan et al., *Phys. Rev. B* 76:115438 (2007).
11. X. Xiao et al., *Opt. Lett.* 31:3194 (2006).

REQUIREMENTS FOR X-RAY DIFFRACTION

Scott T. Misture

*Kazuo Inamori School of Engineering
Alfred University
Alfred, New York*

28.1 INTRODUCTION

Many analytical tools involve the use of x-rays in both laboratory synchrotron settings. X-ray imaging is a familiar technique, with x-ray diffraction (XRD) and x-ray fluorescence (XRF) nearly ubiquitous in the materials analysis laboratory.^{1–3} A long list of additional tools incorporate x-ray optics, especially at synchrotron sources where a continuous range of x-ray wavelengths is readily accessible.

The optical components used in x-ray analysis range from simple slits and collimators to diffractive elements including crystals and multilayers to reflective elements including capillaries and mirrors (see Chaps. 39, 41, 44, 52, and 53). Regardless of the specific application, a description of x-ray optics can be divided into three components:

- Definition of the beam path
- Definition of the beam divergence
- Definition of beam conditioning, or in other words defining the energy spectrum transmitted to the sample or detector by the various optical components under given conditions

Regardless of the quantity measured—intensity, energy, or angle—the interplay of these three parameters is critical to understanding the instrument response. In order to understand the use of optical components we shall begin by describing the simplest of systems which involves slits only.

28.2 SLITS

Using simple apertures, for example, slits, pinholes, and parallel plate collimators, is often sufficient to obtain high-quality data. The most common example of such a system is the powder diffractometer in Bragg-Brentano geometry as shown in Fig. 1. Figure 1 demonstrates that the divergent beam is achieved using a system of slits to control the divergence in the direction normal to the axis of rotation of the goniometer or the *equatorial* divergence. Note that slits are generally used in a pairs (see Fig. 1), with a primary slit and an antiscatter slit designed to block any photons scattered from the edges of the primary slit.

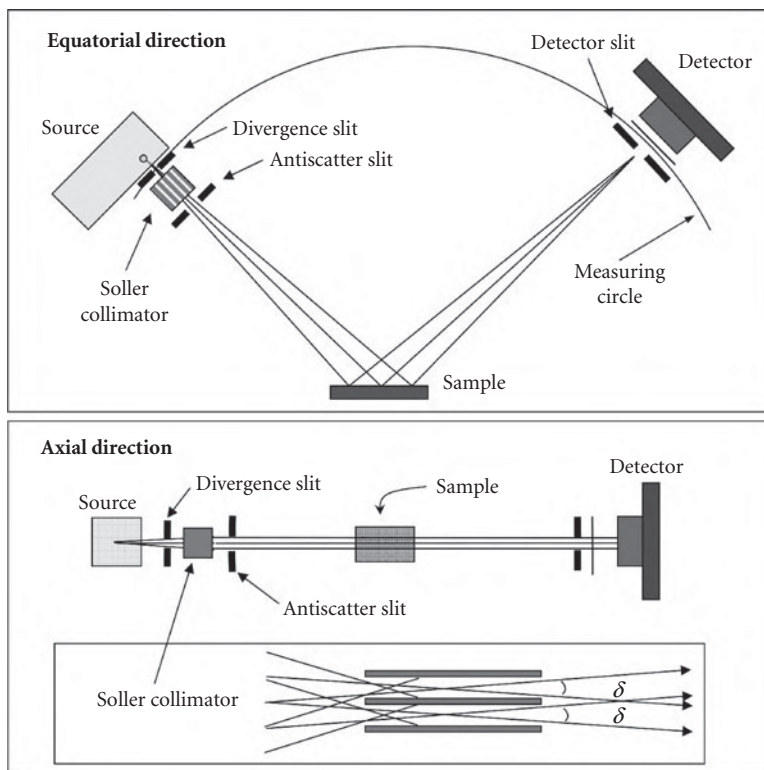


FIGURE 1 Schematic views of the Bragg-Brentano powder diffractometer.

Also shown in Fig. 1 is control of the beam divergence along the axis of the goniometer (*axial* divergence) using parallel plate collimators. The construction of the parallel plate collimator, also called a Soler collimator, includes closely spaced plates that limit the angular range of photons transmitted through the device (Fig. 1). Soler collimators are often used to limit the axial divergence to a few degrees or less, as shown in Fig. 1, but can also be used to achieve “parallel beam” conditions. By “parallel” we mean divergence ranging from the practical limit for a collimator of $\sim 0.05^\circ$ to $\sim 0.2^\circ$.

As an example, Fig. 2 shows the construction of simple parallel beam powder diffractometer incorporating a long Soler collimator on the diffracted beam side. Comparison of data collected in

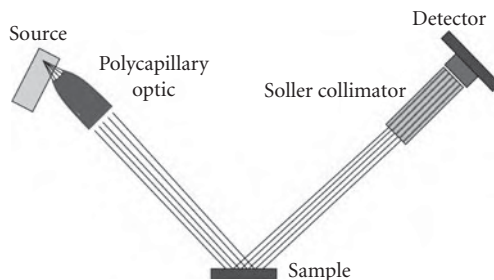


FIGURE 2 A parallel beam diffractometer employing a polycapillary optic and Soler collimator.

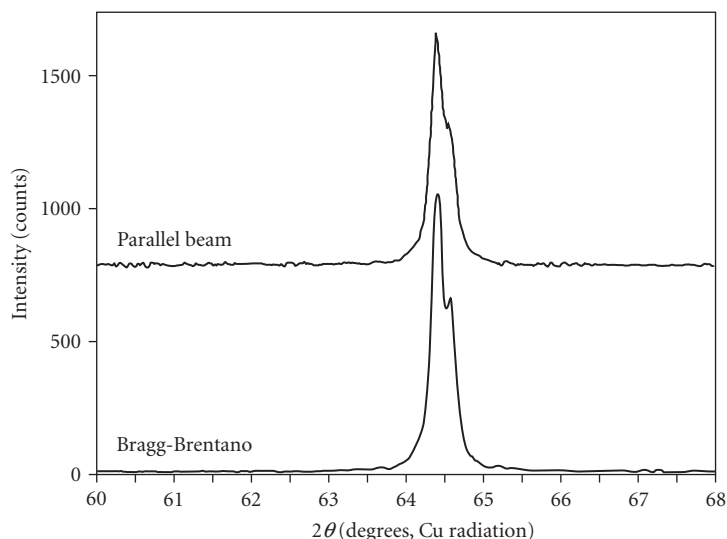


FIGURE 3 Comparison of powder diffraction data for a sample of Ag powder collected using two different instrumental geometries.

the parafofocusing Bragg-Brentano configuration to data collected in the parallel beam configuration is shown in Fig. 3 which instantly reveals that instrumental resolution is about the same for both configurations.

The use of the Soller collimator to achieve “parallel” conditions is an important concept, because the beam divergence and instrumental resolution are linked. Consider first the diffractometer in Fig. 1 that works using a divergent beam that diffracts from the sample and then focuses on the receiving slit. As a focusing system, the width of the receiving slit width plays a large role in the instrumental resolution and angular precision, with wide slits worsening both and vice versa. In sharp contrast, the parallel beam system (Fig. 2) relies on the divergence of the Soller collimator to define the measured angle. In other words, only the diffracted x rays that travel through the collimator are detected and these include only the x rays that propagate within the acceptance angle of the collimator (say 0.05° or 180 asec). In the next section, we shall reduce the divergence further by using crystal optics that can reach to $\sim 0.005^\circ$ of divergence, improving the instrumental resolution.

28.3 CRYSTAL OPTICS

Crystal optics are routinely used in two modes: as energy discriminators (monochromators) that define the range of wavelengths used in an experiment and as angular filters that define the beam divergence. The shape and cut of the crystal is critical and allows beam focusing, compression, expansion, and so on by diffracting in one or two dimensions.

Figure 4 shows four crystals, one flat, one bent and cut, one channel-cut, and the fourth doubly curved. In the case of the flat crystal, used either in diffraction or spectroscopy applications, one relies on Bragg’s law to select the wavelength of interest. Focusing crystals (in either 2-D or 3-D, Fig. 4) take on many forms but in general are bent then cut to transform a divergent beam into a focusing beam while selecting some particular wavelength. The channel-cut crystal, is so named because one channel is cut into a single crystal to facilitate diffraction from both inside faces of the channel using a single device. The channels can be cut either symmetrically so that the incident and diffracted beams make

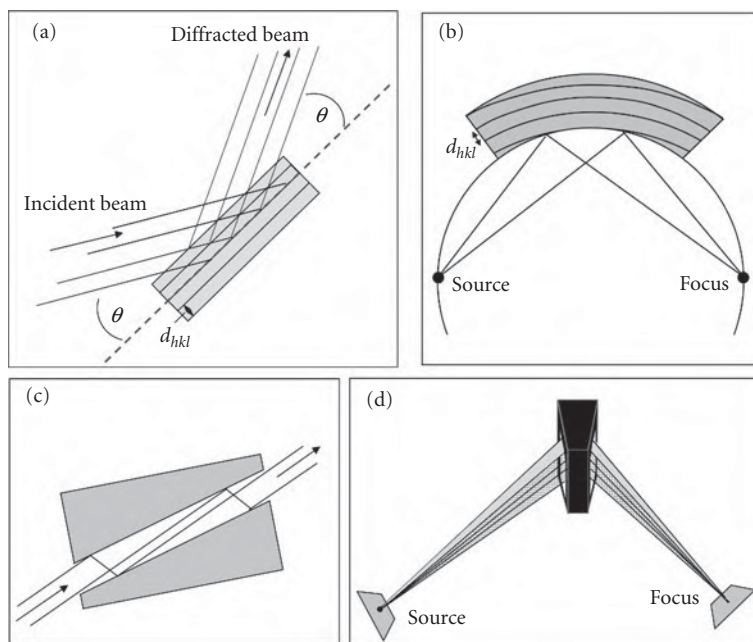


FIGURE 4 The function of several crystal optics including: (a) flat crystal; (b) a bent and cut focusing crystal; (c) an asymmetric channel-cut crystal; and (d) a 2-D focusing crystal.

the same angle with the inside of the channel or asymmetrically where the angle of incidence or diffraction is a small angle with the second angle large. The advantage of asymmetric crystals is higher throughput because of broadening of the rocking curve width.

Selection of crystals involves balancing intensity with resolution (angular or energy), with the latter defined by the rocking curve width. Measuring the intensity diffracted for a particular wavelength as a function of angle provides the rocking curve—a quantitative measure of the perfection of a crystal. The rocking curve width is the most critical aspect of any crystal optic as it defines the range of angles or energies transmitted by the crystal. In the context of energy discrimination, smaller rocking curves result in smaller ranges of energy diffracted by the crystal at some particular angle. Rocking curve widths are a function of wavelength and Miller index, but generally range from ~ 10 asec for high-perfection Si or Ge crystals to ~ 250 asec for LiF or even ~ 1000 for pyrolytic graphite. Graphite and LiF crystals are often used as diffracted beam monochromators in laboratory diffractometers, while Si and/or Ge are reserved for high-resolution epitaxial thin-film analysis or synchrotron beam lines.⁴

In the case of most laboratory diffraction experiments, one or two wavelengths are typically selected, $K\alpha_1$ and/or $K\alpha_2$, using crystal optics. Incorporating a graphite crystal that is highly defected (mosaic) can trim the energy window to include only the $K\alpha_1$ and $K\alpha_2$ components at ~ 60 percent efficiency. Using a crystal of higher perfection facilitates rejection of all but the $K\alpha_1$ radiation, but at a substantially lower efficiency. In order to improve upon the spectral purity and/or beam divergence even further, one can employ multiple crystal monochromators and/or multiple diffraction events using the channel-cut crystal described above. The number of crystals and diffraction events can become quite large for the study epitaxial films in particular, with 4-bounce monochromators on both the incident and diffracted beam sides of the specimen. The reader is referred to recent texts for a more comprehensive review.^{3,4}

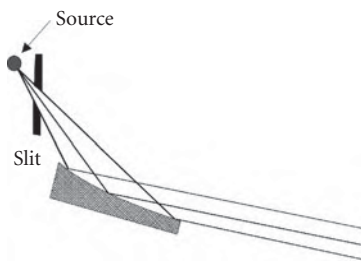


FIGURE 5 A parabolic graded multilayer that transforms a divergent beam into a parallel beam, or vice versa.

28.4 MULTILAYER OPTICS

Multilayer x-ray optics were commercialized in the late 1990s, and offer very specific advantages compared to crystal optics. As shown in Fig. 5 they are man-made crystals that are composed of alternating layers of high and low atomic number materials. They are diffractive optics, generally with large d -spacings and small diffraction angles to provide high efficiency. The rocking curve widths and efficiencies are on the order of 100 to 200 asec with 50 to 70 percent efficiency. As such, they represent a compromise between perfect crystal monochromators (Si, Ge, ~ 10 to 30 asec) and highly mosaic crystals such as graphite (~ 1000 asec).

Multilayer optics are available, like crystals, in a variety of geometrical configurations to provide focused beams and parallel beams, again in one and two dimensions. In addition, cross-coupled multilayers can be used to create point-focused or parallel beams that are today used extensively for single crystal diffraction experiments. The spectral selectivity of multilayers is a function of not only the rocking curve width but also the materials composing the multilayer that can selectively absorb, for example, beta radiation.

28.5 CAPILLARY AND POLYCAPILLARY OPTICS

Drawing hollow glass tubes to a small diameter and with smooth internal surfaces yields single capillary (monocapillary) optics that can be built into arrays to form polycapillary optics. The function of the capillary is total internal reflection of the incident photons that allows the capillary to behave as a “light pipe” to direct x rays in some particular direction. Within limits of the physics of internal x-ray reflection, a variety of beam focusing, collimating, and angular filtering can be achieved, as summarized in Fig. 6. From Fig. 6, it is clear that either mono or polycapillary optics can be used to create small x-ray spot sizes by focusing the beam. Modern capillary optics can provide beam sizes as small as 10 μm routinely, facilitating micro diffraction and micro fluorescence applications.

Another advantage of capillary optics is the ability to improve the intensity in a measurement. Harnessing a large solid angle of x rays emitted from the source or sample in an efficient manner results in 10- or even 100-fold increases in intensity. Similar principles are used in x-ray microsources described below.

28.6 DIFFRACTION AND FLUORESCENCE SYSTEMS

All of the optical components described above can be variously integrated into systems for high intensity, small spot size, large illuminated area, or high energy or angular resolution. One can enhance a diffraction or fluorescence instrument for a specific application by appropriate use of

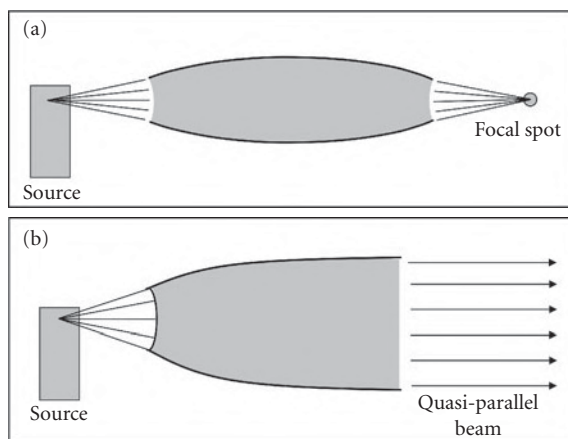


FIGURE 6 Schematics of the function of (a) focusing and (b) collimating polycapillary optics.

optics. Indeed, some modern systems employ prealigned optics that are turn-key interchangeable, affording spectacular flexibility in a single instrument. Naturally, the number of permutations of instrumental arrangements is very large, but in general one attempts to optimize the signal within the limits of the required instrumental resolution.

A clever approach to optimizing both resolution and intensity is the use of “hybrid” optics. Figure 7 shows a schematic of a high-resolution diffractometer applicable for epitaxial film characterization. The defining features of the optics in this case are very high angular resolution provided by the channel cut crystals. However, incorporating a parabolic multilayer before the first crystal monochromator notably improves the intensity. The hybrid design takes advantage of the fact that the multilayer can capture $\sim 0.5^\circ$ of divergent radiation from the x-ray source and convert it at ~ 70 percent efficiency to a beam with only ~ 100 asec divergence. Thus, a substantially larger number of photons reach the channel-cut crystal within its rocking curve width of ~ 20 asec from the multilayer than would directly from the x-ray source, improving the overall intensity.

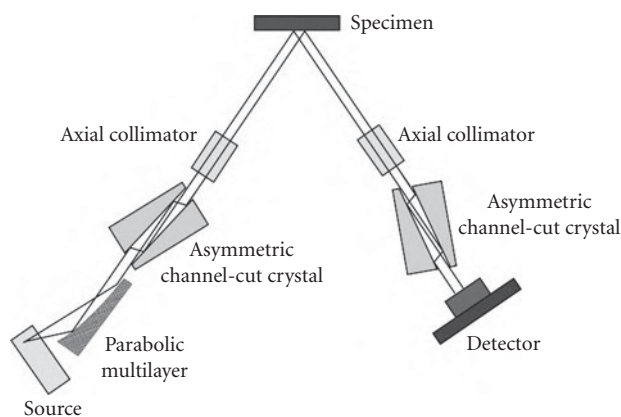


FIGURE 7 Schematic of a high-resolution diffractometer applicable for epitaxial thin film characterization.

28.7 X-RAY SOURCES AND MICROSOURCES

A notable application of x-ray optics is the x-ray “microsource.” Microsource devices in general comprise any low-power and high-flux x-ray source, technology that was enabled by clever application of optical components. The x-ray flux on a particular specimen from a standard x-ray tube is limited by the ability to cool the anode metal, limiting the input power to ~2 kW. Rotating anode sources allow for ~10-fold increases in input power, but are again limited by cooling. In either case, traditional systems use a series of slits to guide x rays from the source to the sample in a linear fashion, discarding most of the x rays produced by the source. One can use x-ray optics to harness a larger solid angle of x rays produced at the source and guide those photons to the experiment. Such approaches have been highly successful, leading to commercialization of microsources that run at power settings as low as 20 W, but provide x-ray flux comparable to traditional x-ray tubes and even rotating anode generators.

28.8 REFERENCES

1. R. Jenkins and R. L. Snyder, *Introduction to X-Ray Powder Diffractometry*, Vol. 138, J. D. Winefordner (ed.), John Wiley & Sons, New York, 1996.
2. H. P. Klug and L. E. Alexander, *X-Ray Diffraction Procedures*, 2nd ed., John Wiley & Sons, New York, 1974, p. 966.
3. B. D. Cullity and S. R. Stock, *Elements of X-Ray Diffraction*, Prentice Hall, NJ, 2001, p. 664.
4. D. K. Bowen and B. K. Tanner, *High Resolution X-Ray Diffractometry and Topography*, Taylor & Francis, London, 1998, p. 252.

This page intentionally left blank

REQUIREMENTS FOR X-RAY FLUORESCENCE

Walter Gibson*

*X-Ray Optical Systems
East Greenbush, New York*

George Havrilla

*Los Alamos National Laboratory
Los Alamos, New Mexico*

29.1 INTRODUCTION

The use of secondary x rays that are emitted from solids bombarded by x rays, electrons, or positive ions to measure the composition of the sample is widely used as a nondestructive elemental analysis tool. Such secondary x rays are called *fluorescence x rays*. The “characteristic rays” emitted from a solid irradiated by x rays or electrons¹ were shown in 1913 by Moseley to have characteristic wavelengths (energies) corresponding to the atomic number of specific elements in the target.² Measurement of the wavelength of the characteristic x rays, as well as observation of a continuous background of wavelengths, was made possible by use of the single-crystal diffraction spectrometer first demonstrated by Bragg.³ There was active development by a number of workers and by the late 1920s x-ray techniques were well developed. In 1923, Coster and von Hevesey⁴ used x-ray fluorescence to discover the unknown element hafnium by measurement of its characteristic line in the radiation from a Norwegian mineral, and in 1932 Coster and von Hevesey published the classical text *Chemical Analysis by X-Ray and Its Applications*. Surprisingly, there was then almost no further activity until after World War II. In 1947 Friedman and Birks converted an x-ray diffractometer to an x-ray spectrometer for chemical analysis,⁵ taking advantage of work on diffraction systems and detectors that had gone on in the previous decade. An x-ray fluorescence measurement in which the energy (or wavelength) spectrum is carried out by the use of x-ray diffraction spectrometry is called *wavelength-dispersive x-ray fluorescence* (WDXRF). There was then rapid progress with a number of companies developing commercial x-ray fluorescence (XRF) instruments. The early developments have been discussed in detail by Gilfrich.⁶ During the 1960s, the development of semiconductor particle detectors that could measure the energy spectrum of emitted x rays with much higher energy resolution than possible with gas proportional counters or scintillators resulted in an explosion of applications of energy-dispersive x-ray fluorescence (EDXRF). Until recently, except for the flat or curved diffraction crystals used in WDXRF, x-ray optics have not played an important role in x-ray fluorescence measurements. This situation has changed markedly during the past decade. We will now review the status of both WDXRF and EDXRF with emphasis on the role of x-ray optics without attempting to document the historical development.

*This volume is dedicated in memory of Walter Gibson.

29.2 WAVELENGTH-DISPERSIVE X-RAY FLUORESCENCE (WDXRF)

There are thousands of XRF systems in scientific laboratories, industrial laboratories, and in manufacturing and process facilities worldwide. Although most of these are EDXRF systems, many use WDXRF spectrometry to measure the intensity of selected characteristic x rays. Overwhelmingly, the excitation mechanism of choice is energetic electrons, and many are built onto scanning electron microscopes (SEMs). Electron excitation is simple and can take advantage of electrostatic and magnetic electron optics to provide good spatial resolution and, in the case of the SEM, to give elemental composition maps of the sample with high resolution. In general, the only x-ray optics connected with these systems are the flat or curved analyzing crystals. Sometimes there is a single analyzing crystal that is scanned to give the wavelength spectrum (although multiple crystals, usually two or three, are used to cover different wavelength ranges). However, some systems are multichannel with different (usually curved) crystals placed at different azimuthal angles, each designed to simultaneously measure a specific wavelength corresponding a selected element or background wavelength. Sometimes a scanning crystal is included to give a less sensitive but more inclusive spectral distribution. Such systems have the benefit of high resolution and high sensitivity in cases where the needs are well defined. In general, WDXRF systems have not been designed to take advantage of recent developments in x-ray optics, although there are a number of possibilities and it is expected that such systems will be developed. One important role that x ray optics can be used in such systems is shown in Fig. 1.

In this arrangement, a broad angular range of x-ray emission from the sample is converted into a quasi-parallel beam with a much smaller angular distribution. A variety of collimating optics could be used, for example, polycapillary (as shown), multilayer, nested cone, and so on. The benefit, represented as the gain in diffracted intensity, will depend on the optic used, the system design (e.g., the diffracting crystal or multilayer film), and the x-ray energy. With a polycapillary collimator, 8 keV x rays from the sample with divergence of up to approximately 12° , can be converted to a beam with approximately 0.2° divergence. With a diffraction width of 0.2° and a transmission efficiency for the optic of 50 percent, the gain in the diffracted beam intensity for flat crystal one-dimensional diffraction would be typically more than 30. Further discussion of the gains that can be obtained in x-ray diffraction measurements can be found in Sec. 29.5. Another potential benefit from the arrangement shown in Fig. 1 is confinement of the sampling area to a small spot defined by the collection properties

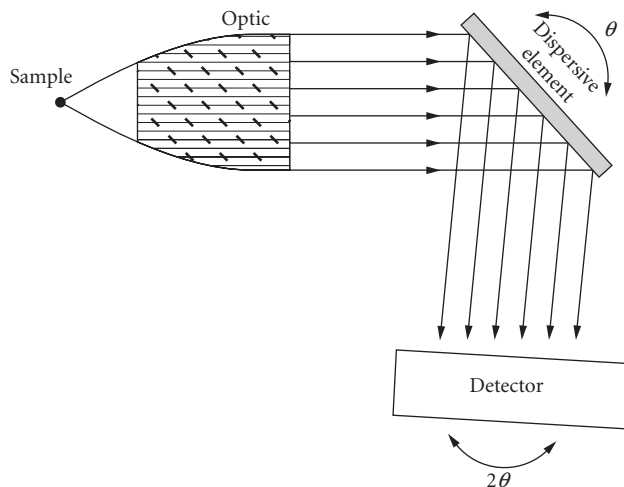


FIGURE 1 Schematic representation of collimating optic in WDXRF system.

of the optic as discussed in Chap. 53. Scanning of the sample will then give the spatial distribution of selected elements. This is also useful in so-called environmental, or high-pressure, SEMs where the position of the exciting electron beam is not so well defined.

Fine Structure in WDXRF Measurements

As noted previously, most of the WDXRF systems in use involve electron excitation either in SEM systems or in dedicated electron micro-probe systems. Photon emission from electron excitation systems contains, in addition to the characteristic lines, a continuous background due to bremsstrahlung radiation resulting from slowing down of the electrons in the solid. Although this background does not seriously interfere with many measurements of elemental composition, it can limit the measurement sensitivity, and can preclude observation of very low intensity features. If the fluorescence x rays are excited by incident x rays, or energetic charged particles, the bremsstrahlung background can be avoided. It should be noted that a continuous background still is present when a broad x-ray spectrum is used as the exciting beam due to scattering of low-energy x rays. This can be largely avoided if monoenergetic x rays are used.⁷

A dramatic illustration of the value of a low background in XRF measurements is contained in recent studies in Japan of fine structure in fluorescence spectra.⁸⁻¹¹ Accompanying each characteristic x-ray fluorescence peak is an Auger excitation peak displaced typically approximately 1 keV in energy and lower in intensity by nearly 1000 times. This peak is not usually observable in the presence of bremsstrahlung background from electron excitation. By using x-ray excitation to get a low background and WDXRF to get high-energy resolution, Kawai and coworkers⁸⁻¹¹ measured the Auger excitation peaks from Silicon in elemental Si and SiO₂ and from Al. They showed that the observed structure corresponds to x-ray absorption fine structure (EXAFS) and x-ray absorption near-edge structure (XANES) that has been observed in high-resolution synchrotron studies. Very long measurement time was necessary to obtain sufficient statistical accuracy. This type of measurement could presumably be considerably enhanced by the use of collimating optics as shown in Fig. 1.

29.3 ENERGY-DISPERSIVE X-RAY FLUORESCENCE (EDXRF)

During the 1960s, semiconductor detectors were developed with dramatic impact on energy and later position measurement of energetic charged particles, electrons, and x rays.^{12,13} Because of their high efficiency, high count rate capability, high resolution compared with gas counters, and their improved energy resolution compared with scintillation counters, these new detectors virtually revolutionized radiation detection and applications including x-ray fluorescence. Initially, the semiconductor junction detectors had a thin active area and, therefore, were not very efficient for x rays. However, by use of lithium compensation in the active area of the detector, it was possible to make very thick depletion layers¹⁴ (junctions) and, therefore, to reach a detection efficiency of 100 percent for x rays. Later, high-purity germanium was used to make thick semiconductor junctions. Such large-volume semiconductor junctions need to be cooled to obtain the highest energy resolution (typically, 130 to 160 eV).

There are now thousands of XRF systems that use cooled semiconductor detectors, most of them mounted on SEMs. Most SEM-based XRF systems do not use any x-ray optics. The detectors can be made large enough (up to 1 to 2 cm²) and can be placed close enough to the sample that they collect x-rays over a relatively large solid angle.

As pointed out previously, electron excitation produces a background of bremsstrahlung radiation that sets a limit on the signal-to-background ratio and, therefore, the minimum detection limit for impurities. This background can be avoided by using x rays or energetic charged particles as the excitation source. Consequently, it is common to see cooled lithium-drifted silicon Si(Li) or high-purity germanium (HpGe) detectors mounted on accelerator beamlines for materials analysis. The ion

beam-based (usually proton or helium ion) technique is called *particle-induced x-ray emission* (PIXE). Again, these do not require optics because the detector can be relatively close to the sample. As with the electron-based systems, optics necessary for controlling or focusing the exciting beam are electrostatic or magnetic and will not be discussed here. When the exciting beam is composed of photons from a synchrotron or free-electron laser (FEL) source, the situation is virtually the same, with no optics required between the sample and the detector. Mirrors and monochromators used to control the exciting beam are discussed in Chaps. 39 and 44.

Monocapillary Micro-XRF (MXRF) Systems

However, if the excitation is accomplished by x rays from a standard laboratory-based x-ray generator, x-ray optics have a very important role to play. In general, the need to obtain a high flux of exciting photons from a laboratory x-ray source requires that the sample be as close as possible to the source. Even then, if the sample is small or if only a small area is irradiated, practical geometrical considerations usually limit the x-ray flux. The solution has been to increase the total number of x rays from the source by increasing the source power, with water-cooled rotating anode x-ray generators becoming the laboratory-based x-ray generator of choice. (For a discussion of x-ray sources, see Chap. 54.) To reduce the geometrical $1/d^2$ reduction of x-ray intensity as the sample is displaced from the source (where d is the sample/source separation), capillaries (hollow tubes) have been used since the 1930s.¹⁵ Although metal capillaries have been used,¹⁶ glass is the overwhelming material of choice,^{17–19} because of its easy formability and smooth surface. In most of the studies reported earlier, a straight capillary was placed between the x-ray source and the sample, and aligned to give the highest intensity on the sample, the capillary length (6 to 20 mm) being chosen to accommodate the source/sample spacing in a commercial instrument. An early embodiment of commercial micro x-ray fluorescence employed metal foil apertures with a variety of dimensions which created spatially resolved x-ray beams. While these crude “optics” provided x-ray beams as small as 50 μm , the x-ray flux was quite limited due to the geometrical constraints.

In 1988, Stern et al.²⁰ described the use of a linearly tapered or conical optic that could be used to produce a smaller, more intense but more divergent beam. This has stimulated a large number of studies of shaped monocapillaries. Many of these are designed for use with synchrotron beams for which they are especially well suited, but they have also been used with laboratory sources. A detailed discussion of monocapillary optics and their applications is given in Chap. 52.

In 1989, Carpenter²¹ built a dedicated system with a very small and controlled source spot size, close-coupling between the capillary and source and variable distance to the sample chamber. The sample was scanned to obtain spatial distribution of observed elemental constituents. A straight 10- μm diameter, 119-mm-long capillary showed a gain of 180 compared to a 10- μm pinhole at the same distance and measurements were carried out with a much lower power x-ray source (12 W) than had been used before.

More recently, a commercial x-ray guide tube or formed monocapillary has been employed to produce x-ray flux gain around 50 times that of straight monocapillary. This modest flux gain enables the more rapid spectrum acquisition and elemental mapping of materials offering new spatially resolved elemental analysis capabilities at the 10 s of micrometers scale.

Polycapillary-Based MXRF

As discussed in Chap. 53, a large number of capillaries can be combined to capture x rays over a large angle from a small, divergent source and focus them onto a small spot. This is particularly useful for microfocus x-ray fluorescence (MXRF) applications.^{22–24}

Using the system developed by Carpenter, a systematic study was carried out by Gao²⁵ in which standard pinhole collimation, straight-capillary, tapered-capillary, and polycapillary focusing optics could be compared. A schematic representation of this system with a polycapillary focusing optic is shown in Fig. 2. The x rays were generated by a focused electron beam, which could be positioned electronically to provide optimum alignment with whatever optical element was being used.²¹

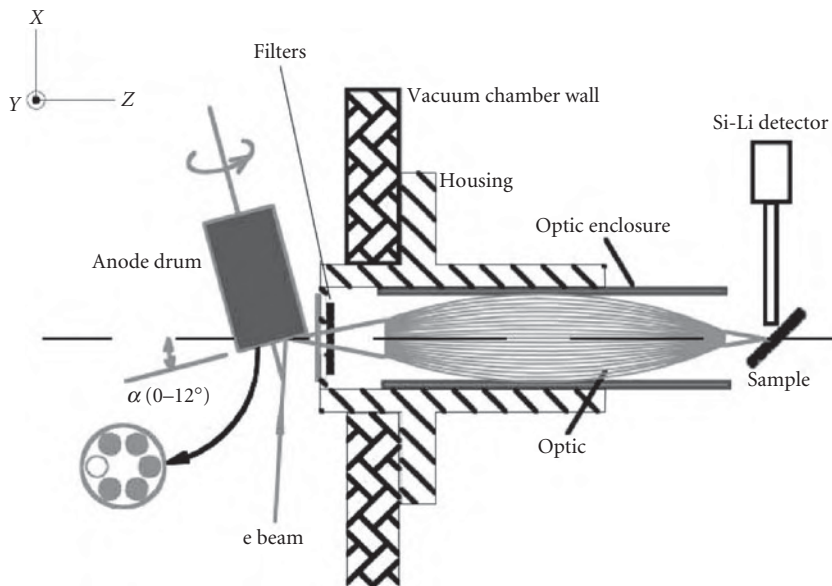


FIGURE 2 Schematic representation of microfocus x-ray fluorescence system. (From Ref. 22.)

The target material could be changed by rotating the anode as shown in Fig. 2. Measurement of the focal spot size produced by the polycapillary focusing optic was carried out by measuring the direct beam intensity while moving a knife edge across the focal spot. The result of such measurements for Cu $K\alpha$ and Mo $K\alpha$ x rays are shown in Fig. 3.

The intensity gain obtained from the polycapillary focusing optic depends on the size of the x-ray emission spot in the x-ray generator, on the x-ray energy, and on the input focal distance (distance between the source spot and the optic). This is because the effective collection angle for each of the transmitting channels is controlled by the critical angle for total external reflection (see Chap. 53). For the system shown, the flux density gain relative to the direct beam of the same size at 100 mm from the source, is shown in Fig. 4 for Cu $K\alpha$ and Mo $K\alpha$ x rays. The maximum gain is about 4400 at 8.0 keV and 2400 at 17.4 keV, respectively.

A secondary x-ray spectrum obtained by irradiating a standard NIST thin-film XRF standard sample, SRM1833, is shown in Fig. 5. Zirconium and aluminum filters were used before the optic to reduce the low-energy bremsstrahlung background from the source. The flux density of the beam at the focus was calculated to be 1.5×10^5 photons·s· μm^2 for Mo $K\alpha$ from the 12-W source operated at 40 kV. The minimum detection limits (MDLs) in picograms for 100-s measurement time were as follows: K, 4.1; Ti, 1.5; Fe, 0.57; Zn, 0.28; and Pb, 0.52. The MDL values are comparable with those obtained by Engstrom et al.²⁶ who used a 200- μm diameter straight monocapillary and an x-ray source with two orders of magnitude more power (1.7 kW) than the 12-W source used in the polycapillary measurements.

By scanning the sample across the focal spot of the polycapillary optic, the spatial distribution of elemental constituents was obtained for a rhyolitic glass inclusion in a quartz phenocryst found in a layer of Paleozoic altered volcanic ash. This information is valuable in stratigraphic correlation studies.^{27,28} The results are shown in Fig. 6. Also shown are the images obtained from Compton scattering (Comp) and Rayleigh scattering (Ray).

There are a number of commercial instruments employing the monolithic polycapillary optics to spatially form the excitation beam. Their commercial success lies in being able to generate an increase in x-ray flux 2 to 3 orders of magnitude greater than that can be obtained without the optic at a given

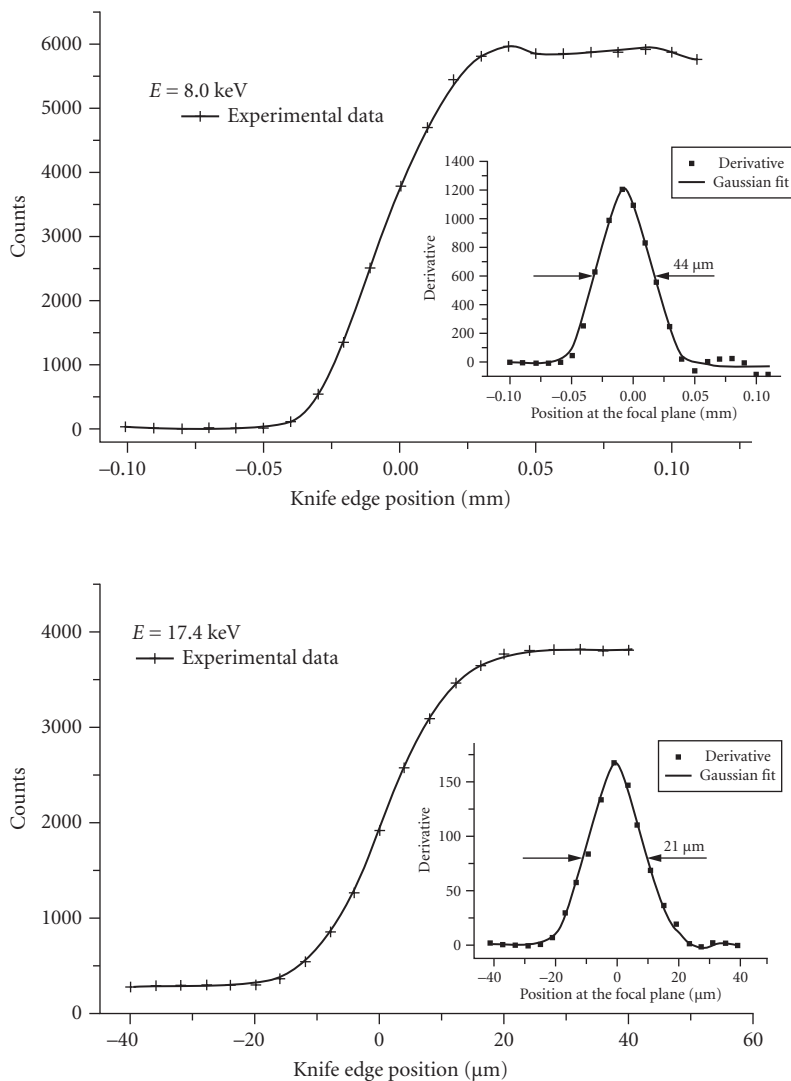


FIGURE 3 Measurement of the focal spot size for Cu and Mo x rays. (From Ref. 25.)

spot size. The future development and potential growth of MXRF rests with the continued innovation of x-ray optics in general and polycapillary optics in particular.

MXRF with Doubly Curved Crystal Diffraction

Although x-ray-induced fluorescence has a significantly lower background than electron-induced fluorescence, there is still background arising from scattering of the continuous bremsstrahlung radiation in the sample. This can be reduced by filtering of high-energy bremsstrahlung in polycapillary focusing optics (see Chap. 53) and by use of filters to reduce the low-energy bremsstrahlung as done for

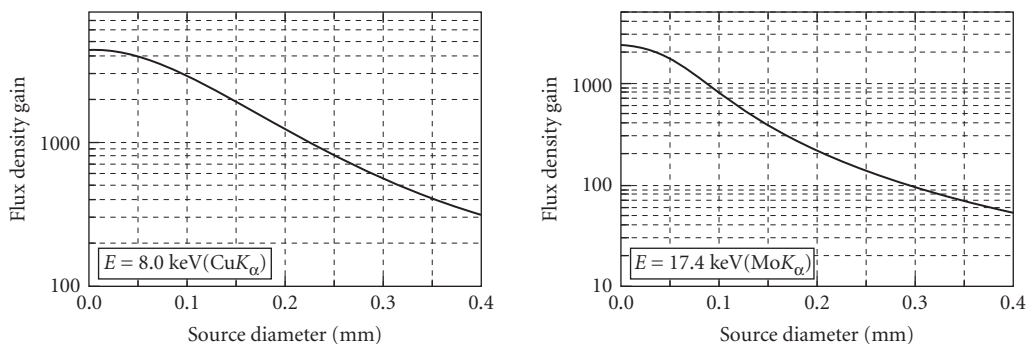


FIGURE 4 Flux density gain as a function of source size. (From Ref. 29.)

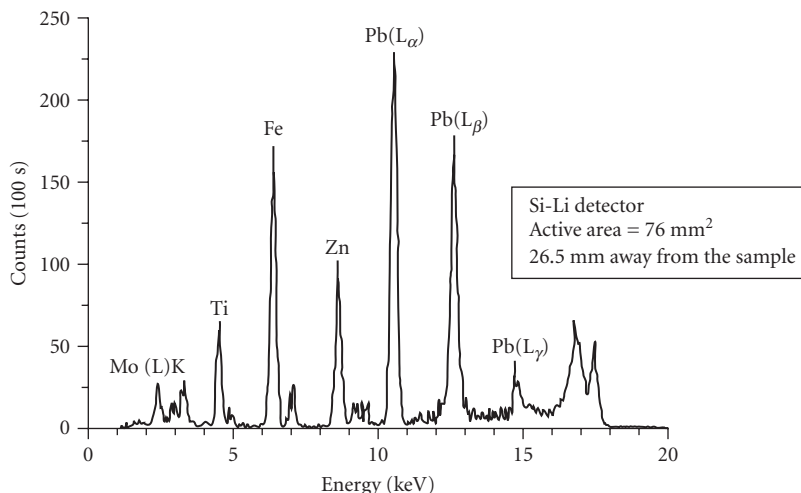


FIGURE 5 Spectrum of SRM 1833 standard XRF thin-film sample. (From Ref. 22.)

the spectrum shown in Fig. 5. Recently, efficient collection and focusing of characteristic x rays by Bragg diffraction with doubly bent single crystals has been demonstrated by Chen and Wytry.³⁰ The arrangement for this is shown in Fig. 7.³¹

Energy spectra taken with a thin hydrocarbon (acrylic) film with a polycapillary focusing optic, and with doubly curved crystal optics with a mica and with a silicon crystal are shown in Fig. 8.³¹ The background reduction for the monoenergetic excitation is evident. Various order reflections are observed with the mica crystal. The angle subtended by the mica crystal is approximately $20^\circ \times 5^\circ$, giving an x-ray intensity only about a factor of three lower than that obtained with the polycapillary lens.

The use of DCCs (doubly curved crystals) in commercial instrumentation has met with commercial success in specific elemental applications. A dual DCC instrument where a DCC is used on the excitation side to create a monochromatic beam for excitation and another DCC on the detection side to limit the region of interest of x-ray fluorescence impinging on the detector provides a highly sensitive and selective detection of sulfur in petroleum streams. Several different embodiments include benchtop, online, and handportable instruments. It is apparent that continued development of DCC-based applications will continue to increase.

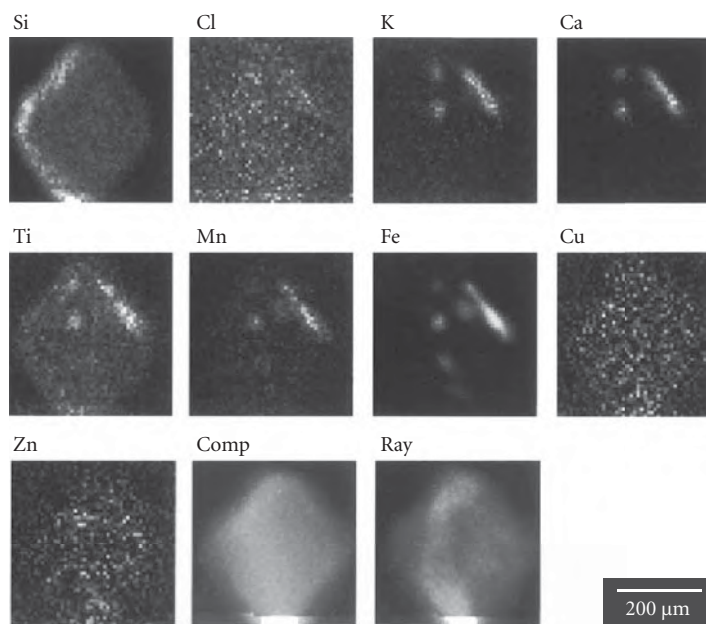


FIGURE 6 MXRF images of various elements in a geological sample that contains small volcanic glass inclusions (tens of micrometers in dimension) within a quartz phenocryst. The last two images are the Compton (energy-shifted) and Rayleigh (elastic) scattering intensity maps. (*From Ref. 25.*)

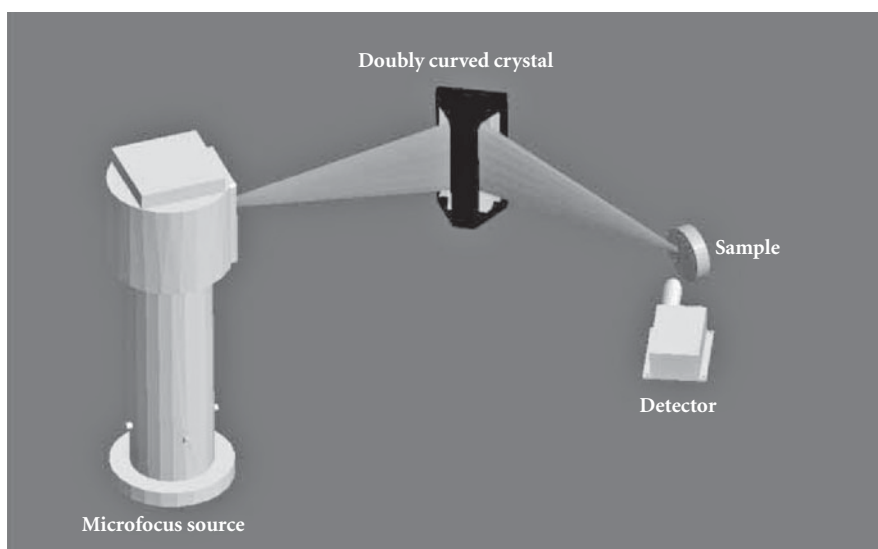


FIGURE 7 Doubly curved crystal MMEDXRF setup. (*Courtesy of XOS Inc. From Ref. 31.*)

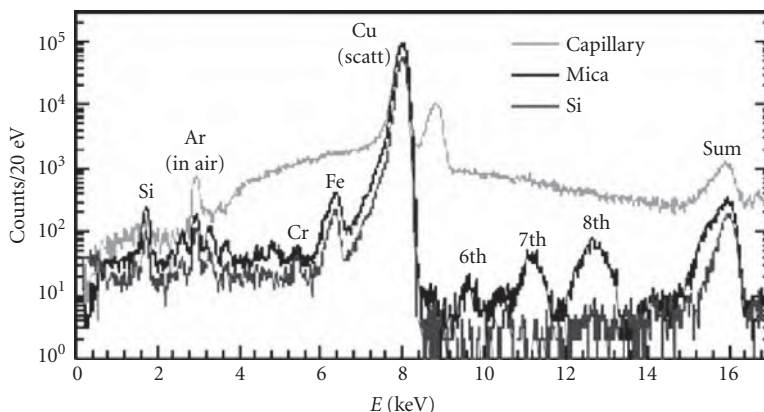


FIGURE 8 Energy spectrum obtained from a thin acrylic sample, measured with a mica doubly bent crystal, a silicon doubly bent crystal, and a polycapillary focusing optic. (From Ref. 31.)

Ultra-high Resolution EDMXRF

As discussed earlier, EDMXRF utilizing cooled semiconductor junction detectors is widely used in science and industry. The energy resolution of semiconductor detectors is typically 140 to 160 eV. During the past few years, very high resolution x-ray detectors based on superconducting transition-edge sensor (TES) microcalorimeters³² semiconductor thermistor microcalorimeters,^{33–35} and superconducting tunnel junctions³⁶ have been developed. Although these detectors are still under active development, there have been demonstrated dramatic benefits for MXRF applications.

TES microcalorimeter detectors have the best reported energy resolution (~ 2 eV at 1.5 keV).³⁷ A schematic representation of a TES microcalorimeter detector is shown in Fig. 9 and energy spectra for a titanium nitride thin film is shown in Fig. 10³⁷ and for a tungsten silicide thin film is shown in

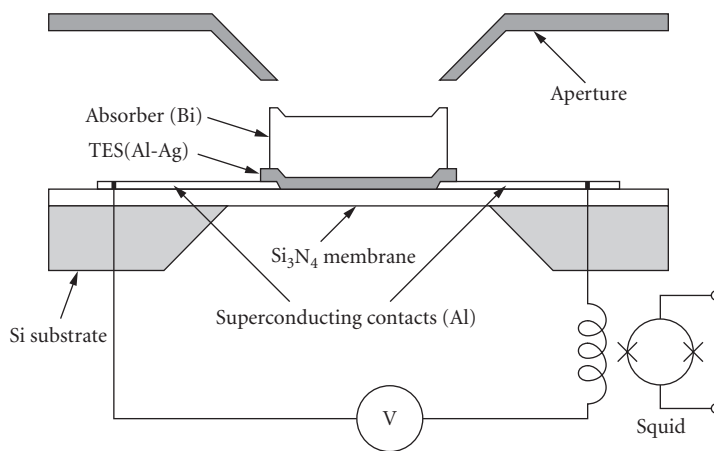


FIGURE 9 A schematic representation of a TES microcalorimeter detector. The operating temperature is ~ 50 mK. (From Ref. 32.)

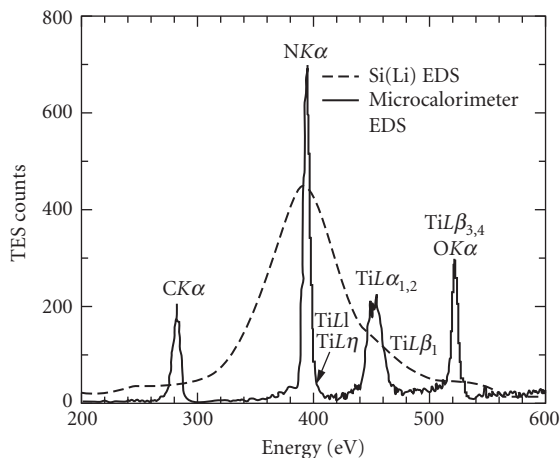


FIGURE 10 Energy spectrum for a TiN thin film on silicon. (From Ref. 37.)

Fig. 11.³² In each case, the spectrum in the same energy region from a silicon junction detector is also shown for comparison.

Because the absorbing element on microcalorimeter detectors must have a low thermal capacitance to achieve high resolution and short recovery time (for higher counting rates), and cannot operate closer than 5 mm to the sample (because of thermal and optical shielding), the sensitivity is low. However, by using a collecting and focusing optic between the sample and the detector, the effective area can be greatly increased.³⁸ A schematic of such an arrangement with a polycapillary focusing optic is shown in Fig. 12. With such a system, the effective area can be increased to approximately 7 mm², comparable with the area of high-resolution semiconductor detectors.

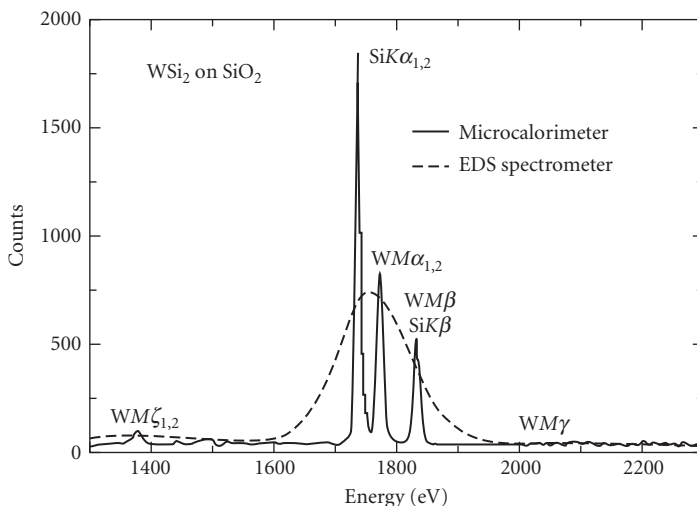


FIGURE 11 Energy spectrum for a WSi₂ film on SiO₂. (From Ref. 32.)

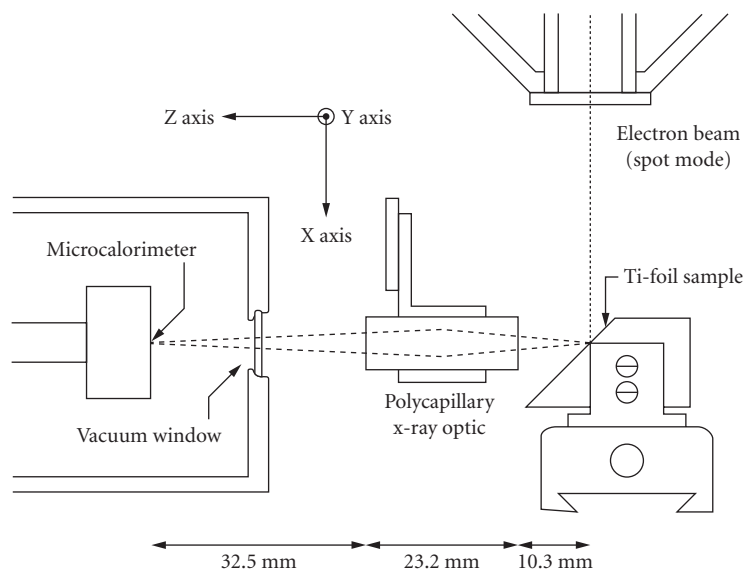


FIGURE 12 Schematic representation of focusing optic for microcalorimeter detector. (From Ref. 38.)

Figure 13 shows a logarithmic plot of the energy spectrum for an aluminum-gallium-arsenide sample.³² This shows the bremsstrahlung background that is present when electrons (in this case, 5 keV) are used as the exciting beam. It is clear that x-ray excitation (especially with monochromatic x rays) will be important in order to use such detectors to observe the low-intensity fine structure to get microstructure and microchemical information with high-resolution energy-dispersive detectors.

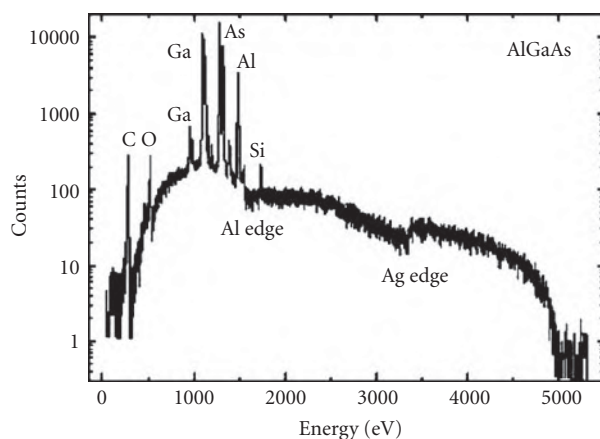


FIGURE 13 Energy spectrum from aluminum-gallium-arsenide sample measured with a TES microcalorimeter spectrometer. (From Ref. 32.)

29.4 REFERENCES

1. R. T. Beatty, "The Direct Production of Characteristic Rontgen Radiations by Cathode Particles," *Proc. Roy. Soc.* **87A**:511 (1912).
2. H. G. J. Moseley, "The High Frequency Spectra of the Elements," *Phil. Mag.* **26**:1024 (1913); **27**:703 (1914).
3. W. H. Bragg and W. L. Bragg, "The Reflection of X-Rays by Crystals," *Proc. Roy. Soc.* **88A**:428 (1913).
4. D. Coster and G. von Hevesey, "On the Missing Element of Atomic Number 72," *Nature* **111**:79, 182 (1923).
5. H. Friedman and L. S. Birks, "A Geiger Counter Spectrometer for X-Ray Fluorescence Analysis," *Rev. Sci. Instr.* **19**:323 (1948).
6. J. V. Gilfrich, "Advances in X-Ray Analysis," *Proc. of 44th Annual Denver X-Ray Analysis Conf.*, Vol. 39, Plenum Press, New York, 1997, pp. 29–39.
7. Z. W. Chen and D. B. Wittry, "Microanalysis by Monochromatic Microprobe X-Ray Fluorescence—Physical Basis, Properties, and Future Prospects," *J. Appl. Phys.* **84**:1064–1073 (1998).
8. J. Kawai, K. Hayashi, and Y. Awakura, "Extended X-Ray Absorption Fine Structure (EXAFS) in X-Ray Fluorescence Spectra," *J. Phys. Soc. Jpn.* **66**:3337–3340 (1997).
9. K. Hayashi, J. Kawai, and Y. Awakura, "Extended Fine Structure in Characteristic X-Ray Fluorescence: A Novel Structural Analysis Method of Condensed Systems," *Spectrochimica Acta* **B52**:2169–2172 (1997).
10. J. Kawai, K. Hayashi, K. Okuda, and A. Nisawa, "Si X-Ray Absorption near Edge Structure (XANES) in X-Ray Fluorescence Spectra," *Chem. Lett. (Japan)* 245–246 (1998).
11. J. Kawai, "Theory of Radiative Auger Effect—An Alternative to X-Ray Absorption Spectroscopy," *J. Electron Spectr. Rel. Phenomona* 101–103, 847–850 (1999).
12. G. L. Miller, W. M. Gibson, and P. F. Donovan, "Semiconductor Particle Detectors," *Ann. Rev. Nucl. Sci.* **33**:380 (1962).
13. W. M. Gibson, G. L. Miller, and P. F. Donovan, "Semiconductor Particle Spectrometers," in *Alpha, Beta, and Gamma Spectroscopy*, 2nd ed., K. Siegbahn (ed.), North Holland Publishing Co. Amsterdam, 1964, p. 345.
14. E. M. Pell, "Ion Drift in an *N-P* Junction," *J. Appl. Phys.* **31**:291 (1960).
15. F. Jentsch and E. Nahring, "Reflexion Von Röntgenstrahlen," *Z. The. Phys* **12**:185 (1931); *Z. Tech. Phys.* **15**:151 (1934).
16. L. Marton, "X-Ray Fiber Optics," *Appl. Phys. Lett.* **9**:194 (1966).
17. W. T. Vetterling and R. V. Pound, "Measurements on an X-Ray Light Pipe at 5.9 and 14.4 keV," *J. Opt. Soc. Am.* **66**:1048 (1976).
18. P. S. Chung and R. H. Pantell, "Properties of X-Ray Guides Transmission of X-Rays through Curved Waveguides," *Electr. Lett.* **13**:527 (1977); *IEEE J. Quant. Electr.* **QE-14**:694 (1978).
19. A. Rindby, "Applications of Fiber Technique in the X-Ray Region," *Nucl. Instr. Meth.* **A249**:536 (1986).
20. E. A. Stern, Z. Kalman, A. Lewis, and K. Lieberman, "Simple Method for Focusing X Rays Using Tapered Capillaries," *Appl. Optics* **27**:5135 (1988).
21. D. A. Carpenter, "Improved Laboratory X-Ray Source for Microfluorescence Analysis," *X-Ray Spectrometry* **18**:253–257 (1989).
22. N. Gao, I. Yu. Ponomarev, Q. F. Xiao, W. M. Gibson, and D. A. Carpenter, "Enhancement of Microbeam X-Ray Fluorescence Analysis Using Monolithic Polycapillary Focusing Optics," *Appl. Phys. Lett.* **71**:3441–3443 (1997).
23. Y. Yan and X. Ding, "An Investigation of X-Ray Fluorescence Analysis with an X-Ray Focusing System (X-Ray Lens)," *Nucl. Inst. Meth. Phys. Res.* **B82**:121 (1993).
24. M. A. Kumakhov and F. F. Komarov, "Multiple Reflection from Surface X-Ray Optics," *Phys. Reports* **191**:289–350 (1990).
25. N. Gao, "Capillary Optics and Their Applications in X-Ray Microanalysis," Ph.D. Thesis, University at Albany, SUNY, Albany, NY, 1998.
26. P. Engstrom, S. Larsson, A. Rindby, and B. Stocklassa, "A 200 mm X-Ray Microbeam Spectrometer," *Nucl. Instrum. Meth.* **B36**:222 (1989).
27. J. W. Delano, S. E. Tice, C. E. Mitchell, and D. Goldman, "Rhyolitic Glass in Ordovician K-Bentonites: A New Stratigraphic Tool," *Geology* **22**:115 (1994).

28. B. Hanson, J. W. Delano, and D. J. Lindstrom, "High-Precision Analysis of Hydrous Rhyolitic Glass Inclusions in Quartz Phenocrysts Using the Electron Microprobe and INAA," *American Mineralogist* **81**:1249 (1996).
29. J. X. Ho, E. H. Snell, C. R. Sisk, J. R. Ruble, D. C. Carter, S. M. Owens, and W. M. Gibson, "Stationary Crystal Diffraction with a Monochromatic Convergent X-Ray Beam Source and Application for Macromolecular Crystal Data Collection," *Acta Cryst.* **D54**:200–214 (1998).
30. Z. W. Chen and D. B. Wittry, "Microanalysis by Monochromatic Microprobe X-Ray Fluorescence—Physical Basis, Properties, and Future Prospects," *J. Appl. Phys.* **84**:1064–1073 (1998).
31. Ze Wu Chen, R. Youngman, T. Bievenue, Qi-Fan Xiao, I. C. E. Turcu, R. K. Grygier, and S. Mrowka, "Polycapillary Collimator for Laser-Generated Plasma Source X-Ray Lithography," in C. A. MacDonald, K. A. Goldberg, J. R. Maldonado, H. H. Chen-Mayer, and S. P. Vernon (eds.), *EUV, X-Ray and Neutron Optics and Sources*, *SPIE* **3767**:52–58 (1999).
32. D. A. Wollman, K. D. Irwin, G. C. Hilton, L. L. Dulcie, D. E. Newbury, and J. M. Martinis, "High-Resolution, Energy-Dispersive Microcalorimeter Spectrometer for X-Ray Microanalysis," *J. Microscopy* **188**:196–223 (1997).
33. D. McCammon, W. Cui, M. Juda, P. Plucinsky, J. Zhang, R. L. Kelley, S. S. Holt, G. M. Madejski, S. H. Moseley, and A. E. Szymkowiak, "Cryogenic Microcalorimeters for High Resolution Spectroscopy: Current Status and Future Prospects," *Nucl. Phys.* **A527**:821 (1991).
34. L. Lesyna, D. Di Marzio, S. Gottesman, and M. Kesselman, "Advanced X-Ray Detectors for the Analysis of Materials," *J. Low. Temp. Phys.* **93**:779 (1993).
35. E. Silver, M. LeGros, N. Madden, J. Beeman, and E. Haller, "High-Resolution, Broad-Band Microcalorimeters for X-Ray Microanalysis," *X-Ray Spectrom.* **25**:115 (1996).
36. M. Frank, L. J. Hiller, J. B. le Grand, C. A. Mears, S. E. Labov, M. A. Lindeman, H. Netel, and D. Chow, "Energy Resolution and High Count Rate Performance of Superconducting Tunnel Junction X-Ray Spectrometers," *Rev. Sci. Instrum.* **69**:25 (1998).
37. G. C. Hilton, D. A. Wollman, K. D. Irwin, L. L. Dulcie, N. F. Berggren, and J. M. Martinis, "Superconducting Transition-Edge Microcalorimeters for X-Ray Microanalysis," *IEEE Trans. Appl. Superconductivity* **9**:3177–3181 (1999).
38. D. A. Wollman, C. Jezewski, G. C. Hilton, Q. F. Xiao, K. D. Irwin, L. L. Dulcie, and J. M. Martinis, *Proc. Microscopy and Microanalysis* **3**:1075 (1997).

This page intentionally left blank

REQUIREMENTS FOR X-RAY SPECTROSCOPY

Dirk Lützenkirchen-Hecht and Ronald Frahm

*Bergische Universität Wuppertal
Wuppertal, Germany*

The basic process related to x-ray absorption spectroscopy (XAS, which includes x-ray absorption near edge spectroscopy, XANES, and extended x-ray absorption fine structure, EXAFS) is the absorption by an atom of a photon with sufficient energy to excite a core electron to unoccupied levels (bands) or to the continuum. An XAS experiment comprises the measurement of the absorption coefficient $\mu(E)$ in the vicinity of an absorption edge, where a more or less oscillatory behavior of $\mu(E)$ is observed. Monochromatic radiation is used, and the photon energy is increased to the value at which core electrons can be excited to unoccupied states close to the continuum. As an example, an absorption spectrum of a platinum metal foil is shown in Fig. 1 for the photon energy range in the vicinity of the Pt L_3 edge. The steep increase of the absorption at about 11.564 keV corresponds to the excitation of electrons from the Pt $2p_{3/2}$ level into unoccupied states above the Fermi level of the Pt material. In general, the exact energy of the absorption edge is a sensitive function of the chemical valence of the excited atom, that is, a shift of the absorption edge toward higher photon energies is observed as the chemical valence of the absorber atom is increased. In many cases, a more or less linear shift of the edge with typically 1 to 3 eV per valence unit can be found in the literature for different elements, and thus, an accurate determination of the edge position is essential for a proper valence determination.¹ Furthermore, in some cases there are also sharp features in the absorption spectrum even below the edge. These so called pre-edge peaks can be attributed to transitions from the excited photoelectron into unoccupied discrete electronic levels of the sample which can be probed by an x-ray absorption experiment. Due to the discrete nature of these transitions, again a high resolution of the spectrometer is required in order to be able to investigate such structures in the spectrum. In the case of Pt, a sharp whitenline-like feature is observed directly at the edge. As can be seen in the inset of Fig. 1, where the derivative of the Pt absorption spectrum is shown, the sharpest features in this spectrum have a full width of only a few electron volts and the energy resolution ΔE of the spectrometer has to be superior for such experiments.

In general, synchrotron radiation is used for x-ray absorption spectroscopy, because the energy range of interest can be easily selected from the broad and intense distribution provided by storage rings (see Chap. 55). X-ray monochromators select the desired photon energy by Bragg's law of diffraction

$$n\lambda = 2d \sin\Theta \quad (1)$$

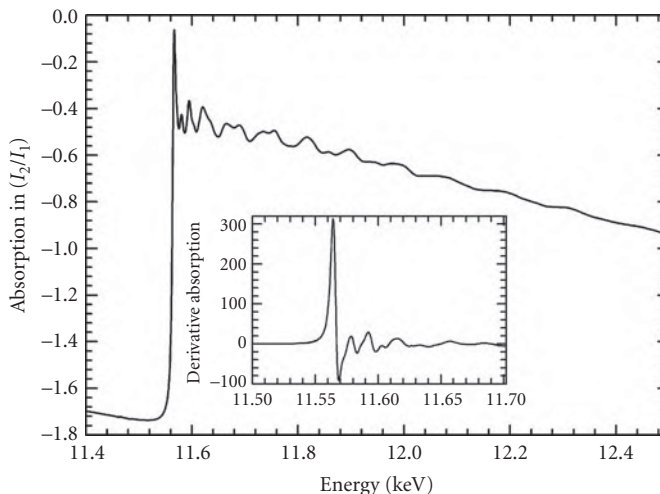


FIGURE 1 High resolution XANES spectrum of a Pt-metal foil at the L_3 -absorption edge measured in transmission mode at the DELTA-XAS beamline using a Si(111)-double crystal monochromator with adaptive optics. In the inset, the derivative of the absorption spectrum is presented in the near edge region.

where λ is the x-ray wavelength, d the lattice spacing of the monochromator crystal, and Θ the angle between the impinging radiation and the lattice planes. However, not only the fundamental wave ($n = 1$) but also higher harmonics ($n > 1$) are transmitted by the monochromator. Selection rules depending on the structure of the crystals can forbid certain harmonics, for example, in the case of the mostly used Si(111) and Si(311) monochromator crystals the second order are not allowed, whereas the third harmonics are present.

However, high-intensity radiation impinging on a monochromator crystal induces a variety of surface slope errors, such as thermal bumps and thermal bending as well as lattice constant variations. This is especially true if insertion devices (wigglers and undulators) are used at third-generation storage rings. Typical heatloads can reach the kilowatt regime, and only a small fraction of typically 10^{-4} is Bragg reflected. Therefore, different concepts have been developed to compensate these unwanted distortions of the Bragg-reflecting surfaces.²⁻⁶ It is far beyond the scope of this paper to describe these efforts in detail, here we will only refer to the literature (see, e.g., Chap. 39), however, stressing the importance of an appropriate compensation. In Fig. 2, we present rocking curves of a Si(111) double-crystal monochromator in order to demonstrate the effect of a compensation on the width (and thereby the energy resolution) of the measured curves. Having in mind that the theoretical value for the energy resolution amounts to ca. 1.3 eV at 9 keV photon energy, those results clearly demonstrate the need for a compensation mechanism. The energy resolution of the noncompensated monochromator with a rocking curve width of more than 100 arc sec, corresponding to more than 14 eV, would be useless for XANES spectroscopy.

Above an absorption edge, a series of wiggles with an oscillatory structure is visible, which modulates the absorption typically by a few percent of the overall absorption cross section as can be seen in Fig. 1. These wiggles are caused by the scattering of the ejected photoelectrons by neighboring atoms and contain quantitative information regarding the structure of the first few coordination shells around the absorbing atom such as bond distances, coordination numbers, and Debye-Waller factors.⁷ Typically, the EXAFS region extends from approximately 40 eV above the x-ray absorption edge up to about 1000 eV or even more.

Higher harmonics in the monochromatic beam may disturb the actual measurement, that is, the absorption coefficients and thus also the EXAFS will be erroneous, so that a suppression of higher

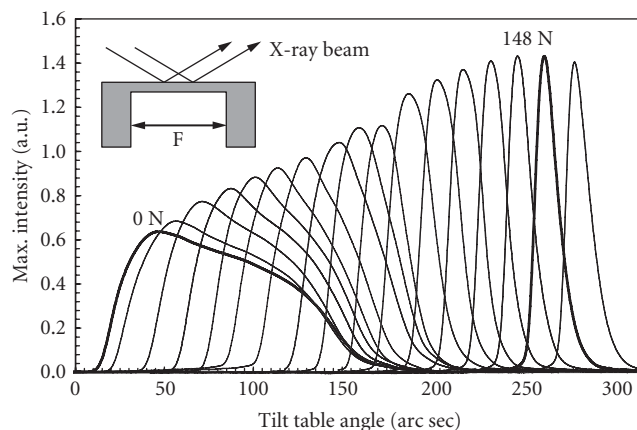


FIGURE 2 Rocking curves measured for a Si(111) crystal pair at 8.9 keV photon energy and a heat load of about 370 W for different bending forces F on the legs of a U-shaped crystal bender as depicted in the inset. The intrinsic width of the rocking curve is 8.9 arc sec (ca. 1.3 eV) compared to about 12 arc sec (ca. 1.8 eV) for the optimum compensated crystal. For comparison, the width of the noncompensated crystal amounts to more than 100 arc sec with less than half of the peak intensity.⁶

harmonics is highly desirable. This is especially important for investigations at lower edge energies, for example, at the Ti K-edge or even lower. Here, any parasitically absorbing elements in the beam path such as the x-ray windows from the beamline to ambient conditions, or even short air pathways, increasingly absorb the photons of interest while the absorption of higher harmonics is negligible. For example, if experiments at the Ti K-edge (4.966 keV) are considered, a beam path of 20 cm in air would result in an absorption of 62 percent for the fundamental wave, in contrast to only 3.5 percent absorption for the third harmonic. In addition, using a Be window of 750 μm thickness as separator between the ultrahigh vacuum system of the beamline and the experimental setup which is usually in ambient air, the transmission increases continuously from only 55 percent at 5 keV to more than 96 percent at 15 keV. In general, thus, the relative intensity of the third harmonic will increase in the course of the beam bath. This is the reason why one has to consider carefully all contributions from the third harmonic, and their influence on the measured spectra. This is illustrated in Fig. 3 for a transmission spectrum of a TiO_2 sample measured in transmission using a Si(111) double crystal monochromator and nitrogen-filled ionization chambers as detectors. Both the different absorption of the low- and the high-energy beams as well as the different ionization of the two beams have been considered. Even in the case of a beam coming from the double crystal monochromator in the vacuum section of a beamline with only 5 percent harmonic content, a significant reduction of the edge jump is visible, as well as a slight damping of the pre-edge peaks. However, this situation, in which the photons of the fundamental wave are strongly absorbed by the environment in contrast to those of the third harmonic, may be regarded as the worst case. This hardening of the beam has a major impact for x-ray studies at lower energies.

From the presented simulation it can be concluded that the x-ray optics have to ensure that the beam is really monochromatic. This may be achieved using different techniques such as detuning of the monochromator crystals⁸ or the use of a mirrored beam. Detuning makes use of the fact that the crystal rocking curve width of higher harmonics is generally much smaller compared to the fundamental wave,⁹ so that a slight tilt of the crystals with respect to each other effectively suppresses the transmitted intensity of higher harmonics. Such a detuning procedure is however not applicable in the case of a channel-cut monochromator, where the two reflecting crystal surfaces originate from a monolithic single crystal with a fixed geometric relation to each other. Here, one makes use of the fact that the x-ray reflectivity of any surface generally decreases with photon energy, that is, a mirror can also be used as

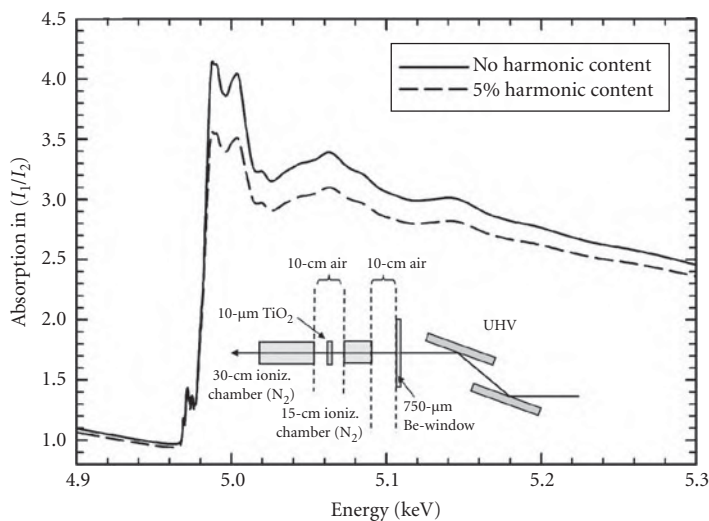


FIGURE 3 Calculation of a XANES spectrum of a nanocrystalline TiO_2 sample (10- μm thickness) for x-ray beams with varying harmonic content. In the calculations, the different absorption of the first and the third harmonic of the Si(111)-double crystal monochromator in the different materials in the optical path were included. We have used a 750-mm Be window which terminates the ultrahigh vacuum (UHV) section of the beamline as well as a 10-cm beam path through laboratory air between the Be window and the first, 15-cm long, N_2 -filled ionization chamber. A second 10-cm air path and the TiO_2 sample were present between the first and the second, (30-cm N_2) ionization chamber.

low pass filter for the harmonic rejection, if the critical energy is smaller than the energy of the corresponding harmonic wave (Refs. 10 and 11 and Chap. 44). It should be mentioned here that the heat load on the monochromator crystals is also reduced if a mirror in front of the monochromator is used, and thus the related problems mentioned above are less important. Furthermore, the mirrors can also be used for the focussing of the x-ray beam:^{11–13} Depending on the surface of the mirror (flat, spherical, cylindrical) and its bending radius, the impinging radiation is either divergent or focused in one or two directions. Using an undulator as source, the point focus of a mirror system can be as small as a few tens of microns, so that it is possible to investigate small specimen (e.g., single crystallites), and spectromicroscopy or microspectroscopic investigations are feasible, even under nonambient conditions.^{14,15}

Up to now, we have not dealt with the problem of lateral stability of the beam on the sample. Even in an idealized setup, we have to consider vertical beam movements on the sample during an EXAFS scan because the beam offset changes significantly for an extended scan as illustrated schematically in Fig. 4. More quantitatively, for a fixed distance D of the x-ray reflecting planes, the beam offset amounts to $h = 2D \cos \Theta$. Thus, h will be larger for a smaller Bragg angle Θ , as can be seen in Fig. 4. Given a typical distance of $D = 50$ mm, a Bragg-angle variation from about 16.5° to 14.5° (which would roughly correspond to an EXAFS scan at the Fe K-edge from ca. 6.9 keV to 7.9 keV) would result in a variation of the beam offset by about 1 mm. Such beam movement is not acceptable for certain experiments, for example, for the investigation of small or inhomogeneous samples or in the case of grazing incidence x-ray experiments, where the height of the sample in the beam is often limited to only a few microns.¹⁶ Furthermore, if the beam downstream the monochromator is subjected to focussing, e.g., for XANES tomography¹⁷ or spectromicroscopic investigations, such beam movements exceed the acceptance width of the focussing optics (see also Chaps. 37, 40, 42, 44, 45, 52, and 53 of this *Handbook*). Thus, a fixed exit geometry by an additional movement of the second crystal (i.e., a variation of the distance D between the Bragg-reflecting surfaces) or, alternatively, a controlled

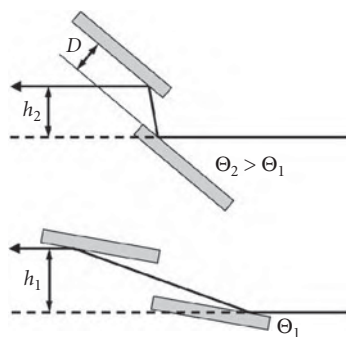


FIGURE 4 Schematic representation of vertical beam offset (h) changes during Bragg-angle variation for two different Bragg angles Θ_1 and Θ_2 , as indicated.

correction of the vertical sample position by means of a lifting table as a function of the Bragg-angle are required. In the case of a channel-cut crystal, a special form of the crystals' reflecting surfaces may also ensure that there is a constant beam height.^{18,19}

We want to conclude here by pointing out that all the topics mentioned above also apply in the case of time-resolved x-ray absorption experiments, where the Bragg angle of the monochromator crystals is moved continuously and the spectrum is collected on the fly (quick-scanning EXAFS^{20,21}). It should be mentioned here, that a high precision of all movements can be reached so that XANES data can be measured in ca. 5 ms, while EXAFS spectra spanning about 1.5 keV of real samples such as catalysts under working conditions are feasible in about 50 ms with sufficient data quality.^{21,22}

30.1 REFERENCES

1. B. Lengeler, "X-Ray Absorption and Reflection in Materials Science," *Advances in Sol. State Phys.* **29**:53–73 (1989).
2. J. Arthur, W. H. Tompkins, C. Troxel, Jr., R. J. Contolini, E. Schmitt, D. H. Bilderback, C. Henderson, J. White, and T. Settersten, "Microchannel Water Cooling of Silicon X-Ray Monochromator Crystals," *Rev. Sci. Instrum.* **63**:433–436 (1992).
3. J. P. Quintana, M. Hart, D. Bilderback, C. Henderson, D. Richter, T. Setterston, J. White, D. Hausermann, M. Krumrey, and H. Schulte-Schrepping, "Adaptive Silicon Monochromators for High-Power Insertion Devices. Tests at CHESS, ESRF and HASYLAB," *J. Synchrotron Rad.* **2**:1–5 (1995).
4. R. K. Smither, G. A. Forster, D. H. Bilderback, M. Bedzyk, K. Finkelstein, C. Henderson, J. White, L.E. Berman, P. Stefan, and T. Oversluizen, "Liquid Gallium Cooling of Silicon Crystals in High Intensity Photon Beams," *Rev. Sci. Instrum.* **60**:1486–1492 (1989).
5. C. S. Rogers, D. M. Mills, W.-K. Lee, P. B. Fernandez, and T. Graber, "Experimental Results with Cryogenically Cooled Thin Silicon Crystal X-Ray Monochromators on High Heat Flux Beamlines," *Proc. SPIE* **2855**:170–179 (1996).
6. R. Zaeper, M. Richwin, D. Lützenkirchen-Hecht, and R. Frahm, "A Novel Crystal Bender for X-Ray Synchrotron Radiation Monochromators," *Rev. Sci. Instrum.* **73**:1564–1567 (2002).
7. D. Koningsberger and R. Prins, *X-Ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS and XANES*, New York: John Wiley and Sons (1988).
8. A. Krolzig, G. Materlik, and J. Zegenhagen, "A Dynamic Control and Measuring System for X-Ray Rocking Curves," *Nucl. Instrum. Meth.* **208**:613–619 (1983).
9. G. Materlik and V. O. Kostroun, "Monolithic Crystal Monochromators for Synchrotron Radiation with Order Sorting and Polarizing Properties," *Rev. Sci. Instrum.* **51**:86–94 (1980).

10. B. W. Batterman and D. H. Bilderback, "X-Ray Monochromators and Mirrors," In: *Handbook of Synchrotron Radiation* Vol. 3: *X-Ray Scattering Techniques and Condensed Matter Research* G. Brown and D. Moncton (eds.), Amsterdam: North Holland, pp. 105–120 (1991).
11. S. M. Heald and J. B. Hastings, "Grazing Incidence Optics for Synchrotron Radiation X-Ray Beamlines," *Nucl. Instrum. Meth.* **187**:553–561 (1981).
12. P. Kirkpatrick and A. V. Baez, "Formation of Optical Images by X-Rays," *J. Opt. Soc. Am.* **38**:766–774 (1948).
13. J. A. Howell and P. Horowitz, "Ellipsoidal and Bent Cylindrical Condensing Mirrors for Synchrotron Radiation," *Nucl. Instrum. Meth.* **125**:225–230 (1975).
14. M. Newville, S. Sutton, M. Rivers, and P. Eng, "Micro-Beam X-Ray Absorption and Fluorescence Spectroscopies at GSECARS: APS Beamline 13ID," *J. Synchrotron Rad.* **6**:353–355 (1999).
15. U. Kleineberg, G. Haindl, A. Hütten, G. Reiss, E. M. Gullikson, M. S. Jones, S. Mrowka, S. B. Rekawa, and J. H. Underwood, "Microcharacterization of the Surface Oxidation of Py/Cu Multilayers by Scanning X-Ray Absorption Spectromicroscopy," *Appl. Phys. A* **73**:515–519 (2001).
16. D. Hecht, R. Frahm, and H.-H. Strehblow, "Quick-Scanning EXAFS in the Reflection Mode as a Probe for Structural Information of Electrode Surfaces with Time Resolution: An In Situ Study of Anodic Silver Oxide Formation," *J. Phys. Chem.* **100**:10831–10833 (1996).
17. C. G. Schroer, M. Kuhlmann, T. F. Günzler, et al., "Tomographic X-Ray Absorption Spectroscopy," *Proc. SPIE* **5535**:715–723 (2004).
18. P. Spieker, M. Ando, and N. Kamiya, "A Monolithic X-Ray Monochromator with Fixed Exit Beam Position," *Nucl. Instrum. Meth.* **222**:196–201 (1984).
19. S. Oestreich, B. Kaulich, R. Barrett, and J. Susini, "One-Movement Fixed-Exit Channel-Cut Monochromator," *Proc. SPIE* **3448**:176–188 (1998).
20. R. Frahm, "New Method for Time Dependent X-Ray Absorption Studies," *Rev. Sci. Instrum.* **60**:2515–2518 (1989).
21. H. Bornebusch, B. S. Clausen, G. Steffensen, D. Lützenkirchen-Hecht, and R. Frahm, "A New Approach for QEXAFS Data Acquisition," *J. Synchrotron Rad.* **6**:209–211 (1999).
22. R. Frahm, B. Griesebock, M. Richwin, and D. Lützenkirchen-Hecht, "Status and New Applications of Time-Resolved X-Ray Absorption Spectroscopy," *AIP Conf. Proc.* **705**:1411–1414 (2004).

REQUIREMENTS FOR MEDICAL IMAGING AND X-RAY INSPECTION

Douglas Pfeiffer

*Boulder Community Hospital
Boulder, Colorado*

31.1 INTRODUCTION TO RADIOGRAPHY AND TOMOGRAPHY

In the early years following Roentgen's discovery of x rays in 1895, applications and implementations of x rays in medical imaging and therapy were already being developed. The development of detectors, scatter control devices, and associated medical imaging paraphernalia has continued unabated. While more complex optics are being researched, the x-ray optics commonly used in medical imaging are apertures, filters, collimators, and scatter rejection grids. The use of x rays in nondestructive testing was actually mentioned in Roentgen's seminal paper, though this application became active only in the 1920s.

31.2 X-RAY ATTENUATION AND IMAGE FORMATION

It is generally true that medical and industrial radiography are transmission methods. X rays are passed through the object being imaged and detected on the opposite side. The image is formed via the differential attenuation of the x rays as they pass through the object. Variations in thickness or density within the object result in corresponding fluctuations in the intensity of the x-ray flux exiting the object. Most simply, this is defined by the linear attenuation coefficient of the material in the path of a narrow beam of radiation,

$$I = I_o e^{-\mu x} \quad (1)$$

where I_o and I are the incident and transmitted x-ray beam intensities, respectively, x is the thickness of the material, and μ is the linear attenuation coefficient, which has units of cm^{-1} and is dependent upon the physical characteristics of the material and the energy of the x-ray beam, as shown in Fig. 1. For heterogeneous objects, Eq. (1) is modified to account for each material, such as differing organs or a cancer, in the path of the beam, as shown in Fig. 2, so that

$$I = I_o e^{-(\mu_1 x_1 + \mu_2 x_2 \dots)} \quad (2)$$

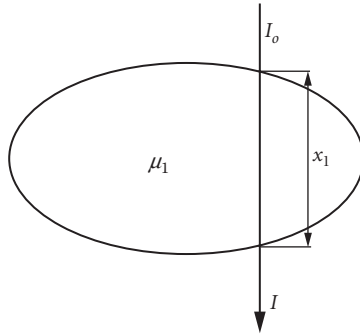


FIGURE 1 Attenuation through a uniform object.

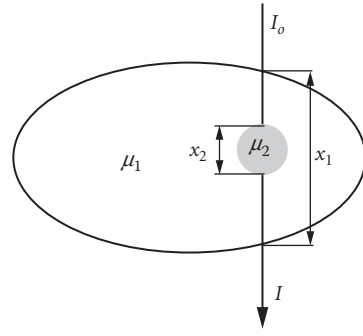


FIGURE 2 Attenuation through an object having an inclusion of differing attenuation.

For a small object, such as a void in a weld, to be visualized, there must be sufficient contrast in the attenuation between the beam passing through the object and beam passing just adjacent to it, as shown in Fig. 3. The contrast is

$$C = \frac{I_2 - I_1}{I_1} \quad (3)$$

where I_2 and I_1 are the intensities of the beams exiting from behind the small object and just adjacent to it, respectively. The amount of contrast required to confidently visualize a given object is determined through the Rose model,¹ a discussion of which is beyond the scope of this chapter. All transmission-based medical and industrial imaging is based upon these fundamental concepts.

As stated earlier, the attenuation coefficient is energy dependent and highly nonlinear. Attenuation in the medical and industrial energy range is dominated mainly by coherent and incoherent (Compton) scattering and photoelectric absorption. Mass attenuation coefficients, are linear attenuation coefficients divided by the density of the material, and thus have units of cm^2/g . Representative mass attenuation curves are demonstrated in Fig. 4, where (a) demonstrates the attenuation curve of water and (b) demonstrates the attenuation curve of iodine.² The discontinuities in the curve for iodine, a commonly used contrast medium in medical imaging, are due to the K and L electron shell absorption edges.

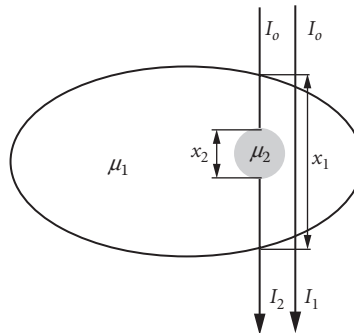


FIGURE 3 The physical basis of radiographic contrast, the difference in attenuation between two adjacent regions.

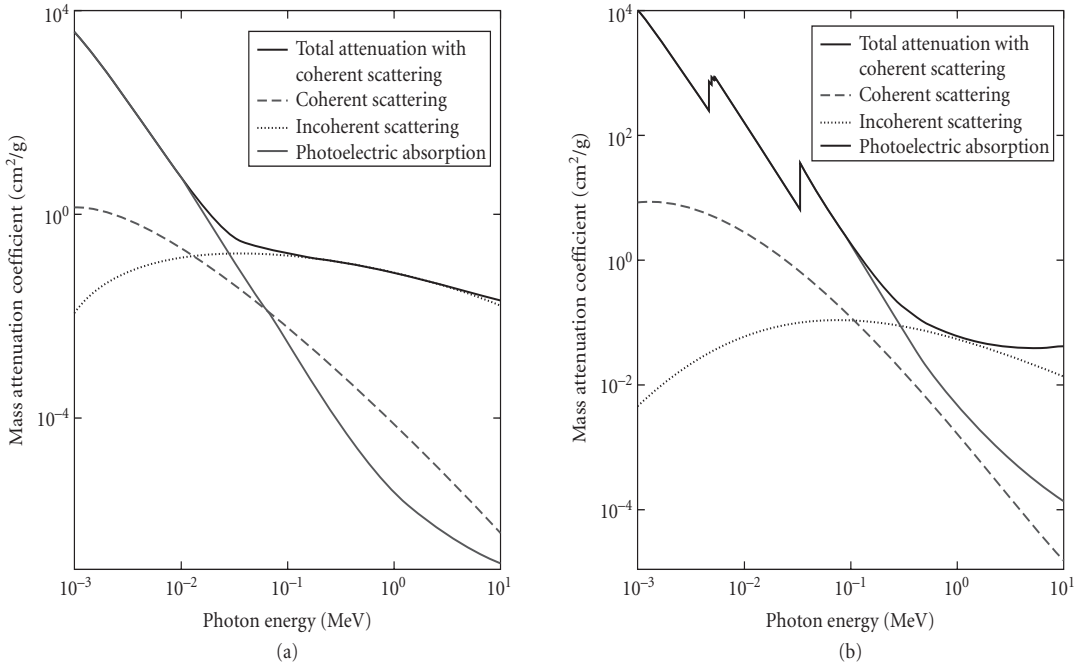


FIGURE 4 X-ray attenuation in (a) water and (b) iodine.

Because of these attenuation characteristics, the energy of the radiation used for imaging must be matched to the task. Roentgen made mention of this in his early work. Diagnostic x rays are typically generated via an x-ray tube having a tungsten anode, operated at an accelerating potential of 60 to 120 kV. A typical unfiltered spectrum is shown in Fig. 5. The spectrum displays the roughly 99 percent component that is bremsstrahlung radiation and the more intense but narrow characteristic peaks of

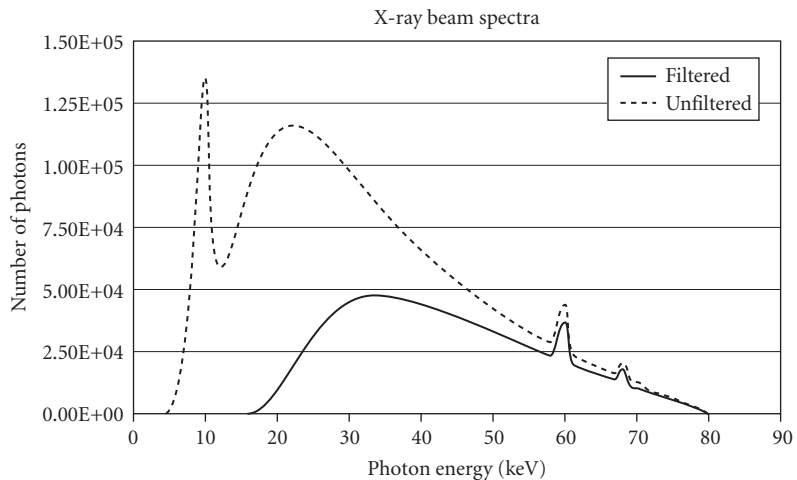


FIGURE 5 Unfiltered and filtered x-ray spectra from a tungsten anode x-ray tube operated at 80 kVp.

tungsten. The accelerating potential for this spectrum is 80 kV. The lower energy radiation is easily attenuated in tissue, therefore, contributing only to patient dose and not to the image. For this reason, medical-use x-ray beams are filtered at the x-ray tube, typically with several millimeters of aluminum. Such a filtered beam is shown in the lower curve of Fig. 5.

31.3 X-RAY DETECTORS AND IMAGE RECEPTORS

Image receptors used for medical and industrial imaging have undergone great development over the last century also. Starting from photographic glass plates, dedicated x-ray film emerged early in the twentieth century. Intensifying screens, which convert x-ray photons to light photons, were introduced for medical imaging soon thereafter, greatly reducing both patient and operator radiation dose. While film-based image receptors are still widely used in industrial and medical imaging, digital detectors came to the scene in the late twentieth century and are gaining broad acceptance. In 2008, digital mammography (breast x-ray imaging) replaced film-screen mammography at a rate of about 6 percent per month.³

The use of film as a radiographic receptor has the advantage of very high spatial resolution. Film alone has resolution greater than 20 lp/mm (20 line pairs per millimeter, equivalent to 25 μm); use of an intensifying screen reduces this to between 10 and 20 lp/mm for most systems. Industrial imagers, for which radiation dose is not a concern, continue to use film directly as the image receptor due to the high resolution.

Particularly for medical imaging, digital image receptors offer a number of advantages. While the dynamic range of film is on the order of 10^2 , for digital receptors the dynamic range is 10^4 or more. Further, the characteristic curve of film is highly nonlinear, leading to variability in the contrast of the image depending on the specific exposure reaching each point of the film. Contrary to this, the response of digital detectors is linear throughout the dynamic range, as shown in Fig. 6. Film serves as x-ray detector, display, and storage medium, meaning that compromises must be made to each role. Since data acquisition and display are separated for digital receptors, each stage can be optimized independently. With physical storage space always at a premium, digital images also have the advantage of requiring little physical space for each image. Multiterabyte storage systems have a small footprint, are relatively inexpensive, and hold large numbers of images.

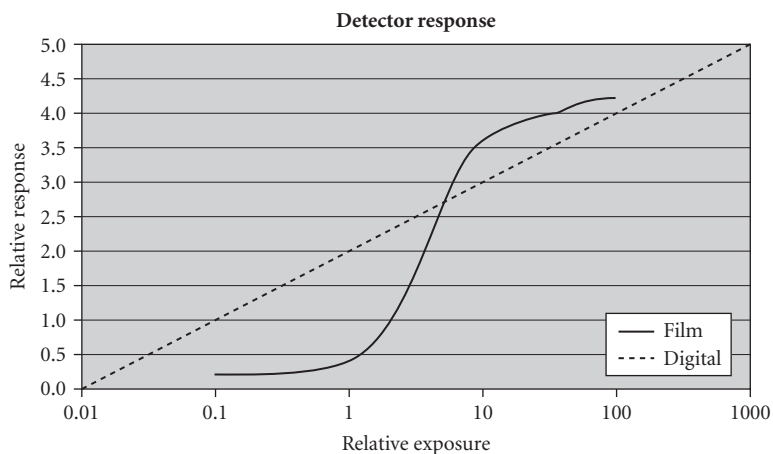


FIGURE 6 Comparison of the relative response of film compared to common digital detectors. Note the linear response and the much wider dynamic range of digital receptors compared to film.

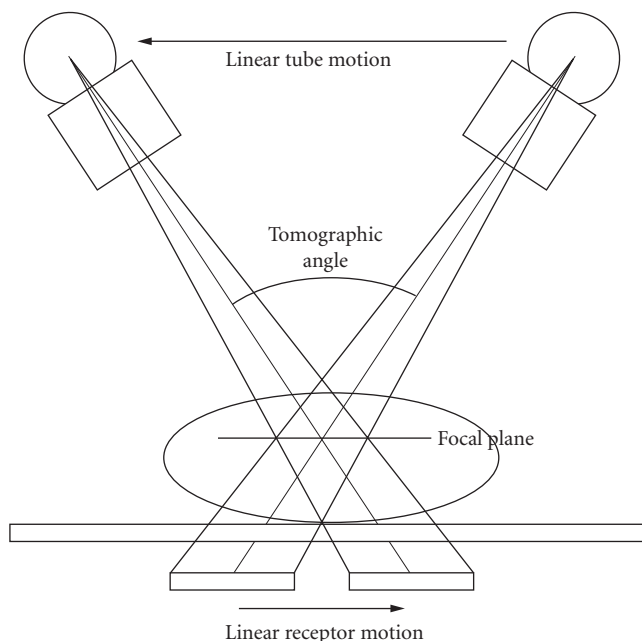


FIGURE 7 Linear tomography. As the tube moves one direction, the image receptor moves the other, with the fulcrum at the plane of interest.

31.4 TOMOGRAPHY

Tomography grew out of conventional x-ray imaging. In its simplest form, linear tomography, the x-ray tube and image receptor move in opposite directions about a fulcrum, which is placed at the height of the object of interest, as shown in Fig. 7. The paired motion of the tube and detector creates an image in which the structures above and below the fulcrum are blurred while leaving a sharp image at the level of the fulcrum. Due to the two-dimensional nature of the x-ray beam, the tomographic plane has a discrete thickness. The wider the tomographic angle, the thinner is the tomographic plane. In the mid to late twentieth century, more complicated paths were used to provide more complete blurring of the off-plane anatomy, but these devices became obsolete by the end of the century due to their bulk and dedicated use.

31.5 COMPUTED TOMOGRAPHY

In 1967, Sir Godfrey Newbold Hounsfield developed the first viable computed tomography scanner,⁴ while Allan McLeod Cormack developed a similar device in parallel. Both were awarded the Nobel Prize in Medicine in 1979 for their work.

Computed tomography (CT) is fundamentally different from conventional tomography. The concept is perhaps best understood by considering the original scanner, which was produced by EMI. This early unit used a pencil beam of radiation and a photomultiplier tube (PMT) detector. The beam and PMT were translated across the object being scanned and a single line of transmission data was collected. The beam and PMT were then rotated 1° and then translated back across the

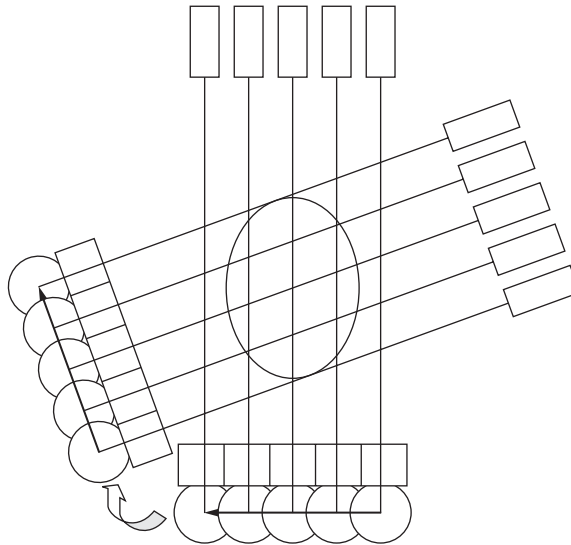


FIGURE 8 Simplified computed tomography data collection, as used in the first CT scanner. A pencil-shaped x-ray beam and detector translated across the object being imaged. This was repeated after the gantry was rotated by a small amount until the full data set had been collected.

patient, generating another line of transmission data as shown in Fig. 8. This process was repeated for 180° until the entire data set had been collected. This is known as “translate-rotate” geometry. Algebraic reconstruction techniques on a computer were then used to reconstruct an image depicting the attenuation of each matrix element. In the first scanner, the brain was divided into slices, each represented by a matrix of just 80×80 elements. Taking several hours to create, this coarse image, however, literally changed how physicians viewed the body.

Developments of this original scanner quickly ensued. More robust reconstruction algorithms, such as filtered back-projection, were created. This algorithm is not as sensitive to large discontinuities in the data, allowing for the technology to be applied not just in the head, but in the body as well. A fan beam of x rays and a linear array of detectors, each of which rotate around the object being scanned, known as “rotate-rotate geometry,” allowed for much more rapid data collection as shown in Fig. 9. Due to the fan beam, a full data set required $180^\circ + \text{fan beam angle} \approx 240^\circ$.

With the development of slip ring technology, the rotation speed of the x-ray tube and detector increased, allowing for improved imaging due to more rapid coverage of the anatomic volume. Increases in computer power and better reconstruction algorithms enabled the development of helical scanning, wherein the table supporting the patient translates through the scanner as the data collection takes place, as seen in Fig. 10.⁵ This has become the standard mode of operation for most medical CT imaging. Further advances in detector arrays, computing power, and reconstruction algorithms have led to multislice scanning. With multislice scanning, multiple data channels, representing multiple axial slices, are collected simultaneously. A continuous increase in the number of slices has been realized since the introduction of this technology, such that modern scanners have up to 320 slices, with 512 slices on the horizon.

One of the most important aspects of current multislice scanners is that the z-axis dimension of the image pixel is now reduced from approximately 12 to 0.625 mm or less. The implication of this is that scanners now provide a cubic voxel, or equivalent resolution in all three directions. With this key development, images may now be constructed in any plane desired. From a single volume of data, the physician may see axial, coronal, sagittal, or arbitrary angle images.

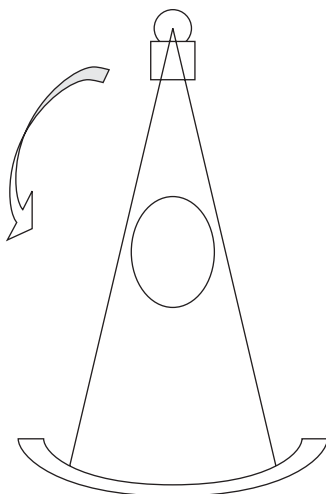


FIGURE 9 Modern computed tomography. A thin, fan beam of x rays is projected through the object and recorded by an array of detectors. The x-ray tube and detectors rotate continuously around the object.

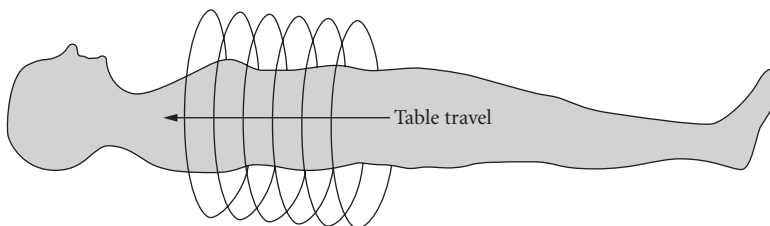


FIGURE 10 Helical computed tomography. The tube and detectors rotate continuously around the patient as the table is translated through the gantry.

Much work is going into the development of “cone-beam CT.” In these devices, the large number of detectors is replaced by a single two-dimensional panel detector with dimensions of tens of centimeters on a side. This is another promising approach to the acquisition of volumetric rather than slice-based data sets.

While most of the focus of computed tomography is on medical imaging, it has also found application in industrial imaging. The imaging principles remain unchanged, the geometry, however, may be adjusted. For example, systems are frequently configured such that the object being imaged is rotated while the imaging system remains fixed. Additionally, energies may range from the approximately 100 kV to over 10 MV, depending on the material to be inspected.

31.6 DIGITAL TOMOSYNTHESIS

The advent of digital projection radiography has led to the development of a technology known as *digital tomosynthesis*. Like conventional tomography, the x-ray tube is translated across or rotated around the object for a specified angle, although the image receptor may remain fixed. Typically,

several tens of images are acquired during the exposure. With this volumetric data set, an image can be reconstructed of any plane within the object. The observer may then scroll through these planes with the overlying anatomy removed. This allows for low contrast objects to be more readily visualized.

31.7 DIGITAL DISPLAYS

Of importance to all digital imaging is the display of the images. Most systems provide 12-bit images having 4096 shades of gray. Most computer monitors, even high-quality medical displays, display at 8 bits, or 256 shades of gray. The human eye is capable of discerning between 6 and 9 bits, or approximately 60 shades of gray. Mapping of the 4096 shades of gray in the image to the 60 shades usable by the observer is achieved through adjusting the window width and level. Figure 11a demonstrates

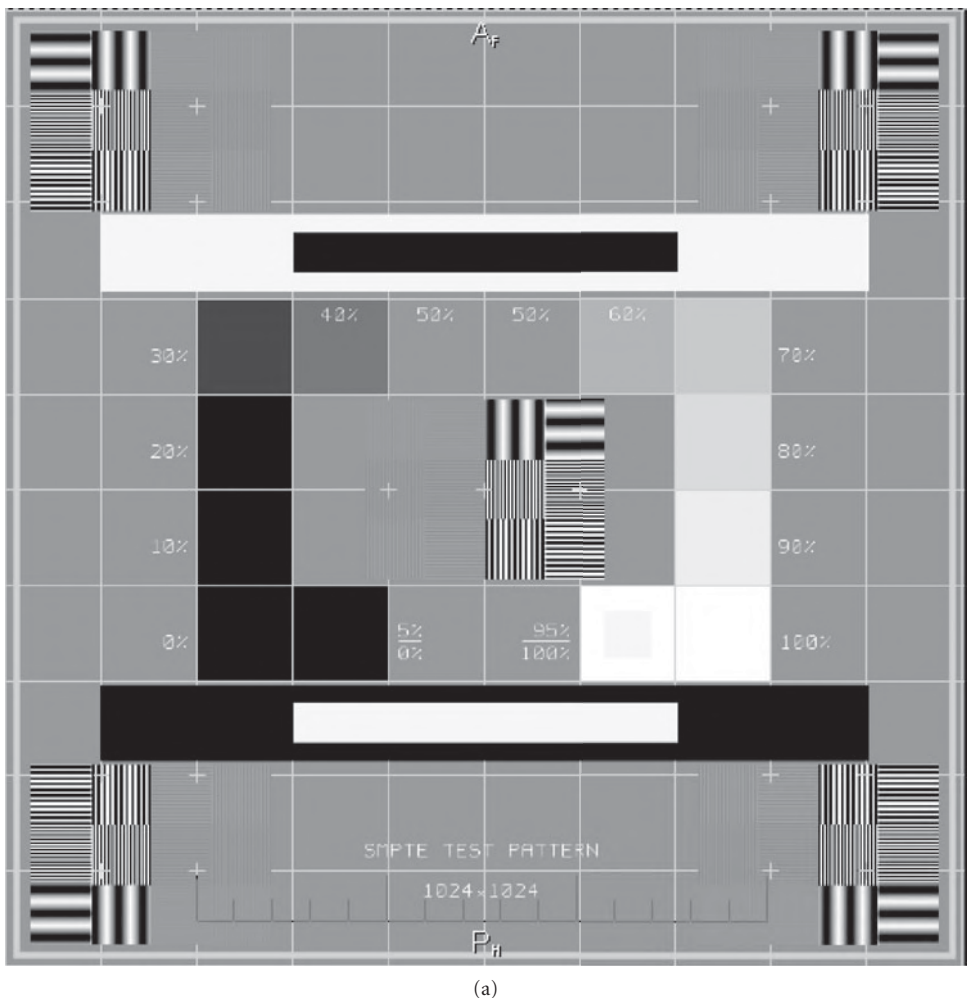


FIGURE 11 The SMPTE test pattern demonstrating (a) good contrast and good display calibration, (b) very high contrast and poor display calibration, and (c) very low contrast and poor display calibration.

good contrast. A narrow window width means that relatively few shades of gray are displayed, yielding a very high contrast image as seen in Fig. 11*b*. Increasing the width compresses the shades of gray in the original data, yielding a low contrast image, shown in Fig. 11*c*. The test pattern created by the Society of Motion Picture and Television Engineers (SMPTE), as shown in these images, is commonly used for the calibration of video displays.⁶

31.8 CONCLUSION

Progress in imaging technology has driven increases in the demands of the applications of that technology. Simple projection radiography will continue to be important to both medicine and industry. Computerized techniques, however, are giving additional tools for diagnosis to both. For example, the speed and image quality of multidetector helical computed tomography are allowing for screening of heart conditions that had been solely the realm of interventional procedures with their associated risks

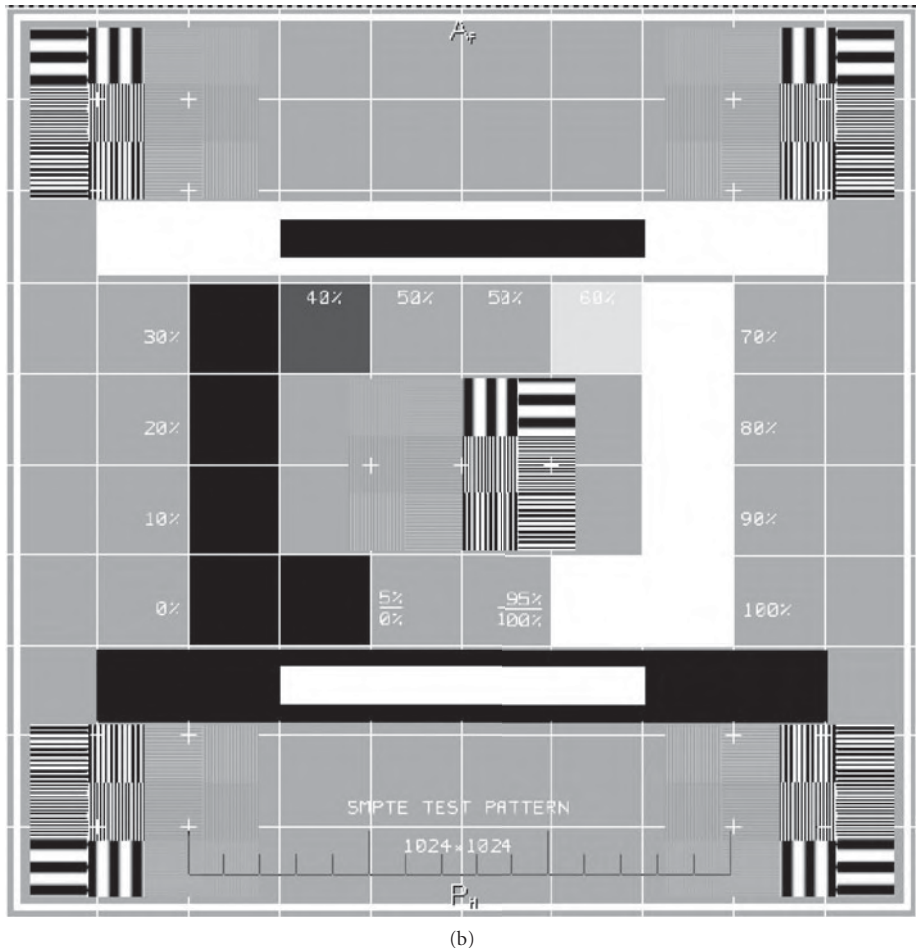


FIGURE 11 (Continued)



FIGURE 11 (Continued)

of complications. Similarly, traditional colonoscopic screening may be giving way to virtual colonoscopy through CT imaging.

Developments such as these show no sign of diminishment. However, for their continued progress, parallel development of x-ray sources, optics, and detectors must also continue.

31.9 REFERENCES

1. O. Glasser, *Wilhelm Conrad Roentgen and the Early History of the Roentgen Rays*, Springfield, IL, Charles C Thomas, 1934.
2. M. J. Berger, J. H. Hubbell, S. M. Seltzer, J. Chang, J. S. Coursey, R. Sukumar, and D. S. Zucker, "XCOM: Photon Cross Sections Database," NIST Standard Reference Database 8 (XGAM), <http://physics.nist.gov/PhysRefData/Xcom/Text/XCOM.html> (retrieved 5/8/2007).

3. P. Butler, April 2007, private communication.
4. G. N. Hounsfield, "Computerized Transverse Axial Scanning (Tomography): Part I. Description of System," *Br. J. Radiol.* **46**(552):1016–1022 (Dec. 1973).
5. W. A. Kalender, P. Vock, A. Polacin, and M. Soucek, "Spiral-CT: A New Technique for Volumetric Scans. I. Basic Principles and Methodology," *Röntgenpraxis* **43**(9):323–330 (Sep. 1990).
6. SMPTE RP133, "Specifications for Medical Diagnostic Imaging Test Pattern for Television Monitors and Hardcopy Recording Cameras," Society of Motion Picture & Television Engineers (SMPTE), White Plains, New York.

This page intentionally left blank

REQUIREMENTS FOR NUCLEAR MEDICINE

Lars R. Furenlid

University of Arizona
Tucson, Arizona

32.1 INTRODUCTION

Single Photon Emission Computed Tomography (SPECT) is a cross-sectional imaging modality that makes use of gamma rays emitted by radiopharmaceuticals (or equivalently, radiotracers) introduced into the imaging subject. As a tomographic imaging technique, the resulting images represent the concentration of radiotracer as a function of three-dimensional location in the subject. A related two-dimensional imaging technique is known as *scintigraphy*. The designation *Single Photon* in the SPECT acronym is used to distinguish the technique from a related emission tomography known as *Positron Emission Tomography* (PET) in which a correlated pair of annihilation photons together comprise the detected signal.

SPECT is one of the *molecular imaging* techniques and its strength as a diagnostic and scientific research tool derives from two key attributes: (1) properly designed radiotracers can be very specific and thus bind preferentially or even exclusively at sites where particular *molecular targets* are present and (2) gamma-ray signals in detectors originate from radioisotope tags on individual molecules, leading to potentially very high sensitivity in units of counts per tracer concentration. Virtually all SPECT systems are photon counting, i.e., respond to and record signals from individual photons, in order to extract the maximum possible information about the location and energy of each detected gamma ray.

Gamma rays are photons that are emitted as a result of a nuclear decay and are thus distinguished from x rays, which result from inner-shell electronic transitions, by the physical process they originate from. Gamma rays and x rays used for imaging overlap in energy ranges, with x rays as a rule of thumb covering the range between approximately 1 to 100 keV, and gamma rays covering the range of 10 keV up to several MeV. Most clinical SPECT imaging is performed with gamma rays (or secondary x rays) with energies of 80 keV (^{201}Tl), 140 keV ($^{99\text{m}}\text{Tc}$), 159 keV (^{123}I), or 171 keV and 245 keV (^{111}In). Preclinical imaging of murine species can be carried out with substantially lower energies, such as 30 keV (^{125}I), due to the smaller amount of tissue that needs to be traversed with a low probability of scatter or absorption.

An ensemble of SPECT radioisotopes emit their photons isotropically, i.e., with equal probability in all directions. In order to form a useful projection image of an extended source on a two-dimensional detector, an image forming principle, generally based on a physical optic, must be employed. The purpose

of the optic is to establish a relationship between locations in the object volume and pixels on the detector. In equation form, this can be expressed as

$$\bar{g}_m = \int h_m(\mathbf{r}) f(\mathbf{r}) d^3\mathbf{r}$$

where $f(\mathbf{r})$ is the gamma-ray photon emission rate (which is proportional to tracer concentration) as a function of 3D position $\mathbf{r}$, $h_m(\mathbf{r})$ is the sensitivity of pixel m to activity in different regions of the object volume, and $\bar{g}_m$ is the mean signal rate ultimately registered in pixel m . A variety of physical factors contribute to the three-dimensional *shape* (and *magnitude*) of the sensitivity functions, but the most important are the parameters of the image-forming optic, and the intrinsic resolution and detection efficiency of the detector.

All modern gamma-ray detectors convert the energy of individual gamma rays into electrical signals that are conditioned and digitized. Most clinical gamma cameras utilize an intermediate step in which the gamma-ray energy excites a burst of secondary lower-energy scintillation photons and are closely related to the scintillation camera designed by Hal Anger in the mid to late 1950s. The design comprises a relatively large slab of inorganic scintillation crystal which is viewed by an array of photomultiplier tubes whose signals are processed to estimate gamma-ray interaction location and energy. Achieved detector resolutions are on the order of 2- to 3-mm FWHM. Research and preclinical SPECT imagers are making increasing use of semiconductor detectors which directly convert gamma-ray energy into electron-hole pairs that migrate under the influence of a bias potential and induce signals in pixel or strip electrodes that can have dimensions down to approximately 50 μm . This is roughly the diameter of the region of space in which the energy of a gamma ray is deposited when it interacts with a solid in a complicated cascade of secondary photons and energetic electrons. SPECT imagers generally incorporate multiple cameras to increase the efficiency of tomographic acquisition, which can involve the measurement of typically 60 to 180 planar projections depending on the system and application.

Tomographic imaging almost always involves a reconstruction operation in which a collection of observations, projection images in the case of SPECT, are processed to recover an estimate of the underlying object. This can be expressed as

$$\hat{f}(\mathbf{r}) = O(\mathbf{g})$$

where O represents the reconstruction operation that acts on data vector $\mathbf{g}$, and $\hat{f}(\mathbf{r})$ the estimate of the object, generally in the form of activity in voxels. In SPECT, the reconstruction operation is currently most often either a filtered backprojection (FBP)¹ or an iterative statistical algorithm such as maximum-likelihood expectation maximization (MLEM)^{2,3} or ordered subsets expectation maximization (OSEM).⁴ The latter reconstruction methods incorporate a forward model of the imaging process that makes it possible to compensate for imperfections in the imaging system, especially if careful calibration measurements are made, and also make it possible to approximately account for absorption and scatter occurring within the object.

32.2 PROJECTION IMAGE ACQUISITION

Although it is possible and sometimes advantageous to perform SPECT imaging without ever forming intermediate projection images, using so-called listmode reconstruction methods that utilize unprocessed lists of gamma-ray event attributes,⁵ most SPECT systems do acquire projection images at regular angular intervals about the imaging subject. Virtually any of the physical processes that redirect or constrain light, such as absorption, refraction, reflection, and diffraction, can in principle be used to form an image in the conventional sense—mapping light from an object source point or plane to an image plane. The need to map an object volume to an image plane, and the requirement to function at the short wavelengths characteristic of gamma rays, place severe restrictions on the types and geometries of optics that can be considered and used successfully for SPECT.

Conventional SPECT systems use pinholes and parallel-hole collimators, or closely related variations such as converging or diverging collimators, slits, and slats, to form images. By permitting only the fraction of light traveling through the open area of the aperture along a restricted range of angles to reach a detector pixel, pinholes and parallel hole collimators define sensitivity functions that are nonzero in conical regions of the object volume. Parallel-hole collimators are the most commonly employed image-forming elements in clinical applications, in part because the projections they produce are the easiest to understand and process. Key features of parallel-hole collimators are that resolution degrades as a function of distance away from the collimator face while sensitivity is nearly unchanged. The parallel bores ensure that most gamma rays enter into the detector with nearly normal incidence angle, minimizing the parallax errors associated with uncertainty in depth of interaction in the detector. Parallel-hole collimators have no magnifying properties (the converging or diverging versions do) and require a tradeoff between distance dependent loss of resolution and overall sensitivity as governed by the collimator aspect ratio. There is a further design tradeoff between resolution loss from leakage between collimator bores and sensitivity loss from reduced fill factor of open collimator area.⁶

SPECT imager design with pinhole apertures also involves a set of design tradeoffs involving resolution, sensitivity, and field of view. Pinhole cameras have magnifications that depend on the ratio of pinhole to detector and pinhole to source distances. Since a three-dimensional object necessarily has a range of pinhole to source distances, resulting in different magnifications for different parts of the object, and the sensitivity cones intersect the object volume with different angles that depend on individual detector pixel locations, pinhole SPECT projections can appear complicated to a human observer. There are also further design tradeoffs between blur from leakage through the pinhole boundaries versus loss of sensitivity from vignetting at oblique angles. Nonetheless, pinhole and multipinhole apertures are currently providing the highest-resolution preclinical SPECT images, as reconstruction algorithms have no difficulty unraveling the geometric factors in the projections.

New optics, and systems, developed for SPECT should be evaluated in comparison to the conventional absorptive apertures discussed above, ideally with objective measures of system performance such as the Fourier crosstalk matrix.

32.3 INFORMATION CONTENT IN SPECT

Barrett et al.⁷ have suggested the use of the Fourier crosstalk matrix as a means of characterizing SPECT system performance. This analysis derives a measure of resolution, similar to a modulation transfer function (MTF), from the diagonal elements of the crosstalk matrix, and a second measure of system performance from the magnitude of the off-diagonal elements. An ideal system has significant amplitudes on the diagonal elements out to high spatial frequencies, representing high spatial resolution, and very small off-diagonal elements representing minimal aliasing between three-dimensional spatial frequencies (denoted by $\mathbf{p}_k$). The crosstalk matrix analysis is best carried out with measured calibration data, but it can also be carried out with accurate system models. The diagonal elements,

$$\beta_{kk} = \sum_{m=1}^M \left| \int_S h_m(\mathbf{r}) \exp(2\pi i \mathbf{p}_k \cdot \mathbf{r}) d\mathbf{r} \right|^2$$

are easily understood to represent the magnitude of the k th Fourier component of the overall system sensitivity function.

One of the primary motivations for optics and aperture development for SPECT is to increase the optical efficiency. Conventional pinholes and parallel hole collimators image by excluding most of the photon distribution function, the phase-space description of photon trajectories that encodes both locations and directions. They therefore are very inefficient at collecting light. Since SPECT radio-pharmaceuticals need to adhere to the “tracer” principle, i.e., be administered in limited amounts,

SPECT images almost always have relatively low total counts per voxel compared to many imaging techniques, and therefore significant Poisson noise. New system designs with large numbers of pinholes are helping to address these concerns.⁸

32.4 REQUIREMENTS FOR OPTICS FOR SPECT

Image-forming elements for SPECT systems need to meet criteria to be useful for imaging that can be formulated as a set of requirements. One of the primary requirements is that the fraction of gamma rays that are able to reach the detector via undesired paths is small relative to the fraction that traverses the desired optical path. For example, for pinholes, the open area of the pinhole must be large relative to the product of the probability of gamma rays penetrating the shield portion of the aperture and the relevant area of that shield. Interestingly, as the number of pinholes increases, the tolerance to leakage also increases, which is an aid in system design. Given the relatively high energies of the gamma rays emitted by the most useful SPECT radioisotopes, this is not a trivial requirement to meet. A comparable condition for parallel hole collimators is that only a small fraction of the photons reach the detector surface after having traversed (or penetrated) at least one bore's septal wall.

The overall system sensitivity function, which represents a concatenation of all of the sensitivity functions in all of the projections in the tomographic acquisition, needs to have significant values for high spatial frequencies in the diagonal elements of the Fourier crosstalk matrix, but have relatively small off-diagonal terms. A variety of conditions can break this requirement. For example, certain geometric features in the sensitivity functions, such as might arise from a symmetric arrangement of multiple pinholes with respect to the imager axis, can introduce correlations that result in enhanced off-diagonal elements. Significant holes in the sensitivity in regions of the object space can also be problematic, as can issues whenever there is an uncertainty as to which of several alternative paths a photon took to arrive at the detector.

Since SPECT imaging generally involves three-dimensional objects, the image-forming element needs to work with photons emitted from a volume of space consistent with the dimensions of the objects or subjects being imaged. In some very specialized applications, for example, in preclinical scanning applications involving known locations of uptake in limbs, the volume required can be small, on the order of 1 cm³ or less. But for most general preclinical imaging, the object volume is measured in units of tens of cubic centimeters. In clinical applications, the object volume is often measured in units of hundreds of cubic centimeters. For practical application with living imaging subjects, SPECT acquisitions should not exceed roughly 30 to 60 minutes in total time, which makes it impractical, or at least undesirable, to employ optics that require rastering to build up single planar projections.

The materials used for gamma-ray optics can sometimes cause problems with secondary fluorescence, depending on the energy spectra of interest. For example, collimators made of lead give off fluorescence k_α and k_β photons at around 75 and 85 keV, respectively, that can be hard to distinguish from the 80 keV emissions from ²⁰¹Tl labeled tracers. Care in geometric design can often minimize the potential for collimator fluorescence reaching the detector.

32.5 REFERENCES

1. P. E. Kinahan, M. Defrise, and R. Clackdoyle, "Analytic Image Reconstruction Methods," in *Emission Tomography*, M. N. Wernick and J. N. Aarsvold, eds., chap. 20 Elsevier, San Diego, 2004.
2. L. A. Shepp and Y. Vardi, "Maximum Likelihood Estimation for Emission Tomography," *IEEE Trans. Med. Imaging* 1:113–121 (1982).
3. K. Lange and R. Carson, "EM Reconstruction Algorithms for Emission and Transmission Tomography," *J. Comput. Assist. Tomogr.* 8:306–316 (1984).

4. H. M. Hudson and R. S. Larkin, "Accelerated Image Reconstruction Using Ordered Subsets of Projection Data," *IEEE Trans. Med. Imaging* **13**:601–609 (1994).
5. H. H. Barrett, T. White, and L. C. Parra, "List-Mode Likelihood," *J. Opt. Soc. Amer. A* **14**:2914–2923 (1997).
6. D. L. Gunter, "Collimator Characteristics and Design," in *Nuclear Medicine*, R. E. Henkin, M. A. Boles, G. L. Dillehay, J. R. Halama, S. M. Karesh, R. H. Wagner, and A. M. Zimmer, eds., Mosby, St. Louis, MO, 96–124 (1996).
7. H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective Assessment of Image Quality. II. Fisher Information, Fourier Cross-Talk, and Figures of Merit for Task Performance," *J. Opt. Soc. Amer.* **12**:834–852 (1995).
8. N. U. Schramm, G. Ebel, U. Engeland, T. Schurrat, M. Behe, and T. M. Behr, "High-Resolution SPECT Using Multipinhole Collimation," *IEEE Trans. Nucl. Sci.* **50**(3): 315–320 (2003).

This page intentionally left blank

REQUIREMENTS FOR X-RAY ASTRONOMY

Scott O. Rohrbach

Optics Branch

Goddard Space Flight Center, NASA

Greenbelt, Maryland

33.1 INTRODUCTION

X-ray astronomy can be generally defined as the observation of x-rays from extraterrestrial sources. A variety of methods are used to determine, with more or less accuracy, where extraterrestrial photons come from, but the most common are simple collimation, coded aperture masks and focusing mirror systems. Collimator systems are used to achieve imaging resolution down to approximately 1° , and are independent of energy. Coded aperture masks, such as the system on the Swift Burst Alert Telescope (BAT), can localize the direction of strong signals to within a few minutes of arc, and are similarly energy independent. Finally, true imaging systems can achieve sub-arc-second angular resolution, the best demonstration of which is the Chandra X-Ray Observatory, with 0.5 arc-second half-power-diameter (HPD). While the technical details of space-based x-Ray observatories are covered in Chap. 47, the various requirements and trade-offs are outlined here. For x-ray imaging systems, a large number of factors come into play, including

1. Imaging resolution
2. Effective collecting area
3. Cost
4. Focal length
5. Field of view
6. Mass

For missions based on targeted objects, the first three are usually the primary factors and trade offs must be made between them. For survey missions, where the telescope is constantly scanning the sky, field of view also comes into play, but then imaging resolution is not as critical. For either type of observatory, the telescope mass is defined more by the mission mass budget, which is set by the choice of launch vehicle, than any other factor. On top of these mirror requirements are considerations of what detector system is being used, since a higher quality (and thus more expensive) imaging system coupled to a low-resolution detector would be a waste of money.

Current technological limitations in optics and detectors mean that the x-ray regime is split into two observational regions, referred to as the soft and hard x-ray bands. The soft band spans the 0.1 to ~ 10 keV range, while the hard band spans the ~ 10 to ~ 300 keV range. In the soft band, detector efficiencies and filters required to absorb visible light effectively define the low energy cutoff, while the throughput of focusing systems and the average critical angle of total external reflection for simple thin films (see Chap. 26) define the high energy cutoff. In the hard band, the upper limit is generally determined by the quantum efficiency of the detector system being used. Due to the focal length limitations for spacecraft (~ 10 m without an extendable optical bench) and the fact that multilayer coatings that can extend the bandpass of focusing optics have only recently become available, hard x-ray observation on space-based observatories has been limited to simple collimation and coded-aperture mask instruments.

33.2 TRADE-OFFS

As noted above, the most obvious competing factors are imaging resolution, effective area, and cost. Both imaging quality and effective area can be limited by the available financial budget the mission has. Other trade-offs exist. For example, one consideration is mirror substrate thickness. Thicker mirror substrates allow for conventional grinding and polishing, yielding excellent imaging quality. But all of the projected area occupied by the cross section of each substrate means less effective area for the given optical aperture of the telescope. Also, grinding and polishing of each surface can be very costly. On the other hand, the desire for large effective collecting area drives the thickness of each substrate down, in order to minimize the optical aperture obscured by the substrate. This leads to substrates that are effectively impossible to individually polish, both due to time and cost constraints as well as the problems of polishing very thin substrates without introducing print-through errors or making the flexible substrates weaker. And, thinner, more flexible substrates are harder to integrate into a flight housing without introducing significant figure distortions.

For each mission concept that gains significant community and financial support, a team of astronomers studies the various performance trade-offs with respect to the science the mission will pursue. Will the primary targets be bright sources or dim, or a mixture? Will they be compact or span many minutes of arc? Will they change quickly like the rapid oscillations of a pulsar, or be more static like the gas between stars and galaxies? The answers to these questions depend to some degree on what the most compelling scientific questions of the day are. Once the scientific scope of the mission has been defined, the functional requirements begin to take shape, but there are still many other questions to be answered. High-resolution spectroscopy requires more collecting area than pure imaging, but a telescope with good imaging quality also reduces the background inherent in any measurement, improving the spectroscopic results. And since even the most modest observatory will cost more than \$100 million to design, launch, and operate, any new observatory must represent a significant improvement over previous missions in at least one area, be it imaging quality, collecting area, spectral resolution, bandpass, timing accuracy, and so on.

In a short review such as this, it would be impossible to list all of the various targets, science areas, and trade-offs that need to be made for any particular observatory. Instead, a short case study, using three current missions and one upcoming mission study, is presented. Each observatory has different goals and budgets, and as such, approaches the question of optimization from a different point of view. The missions are summarized in Table 1.

At one corner of the imaging/effective area/cost trade space is the Chandra X-Ray Observatory, which was primarily designed to yield the best x-ray imaging ever achieved, and is also the most expensive x-ray observatory ever made. It uses four nested shells with the largest being 1.4 m in diameter, and each of the primary and secondary pairs is figured out of thick monolithic shells of Zerodur, nearly 1 m long. The emphasis on imaging quality translated into a small number of mirror elements that could be figured and polished precisely, yielding a point spread function of 0.5 arc-second HPD, even after all four shells were aligned to one another. During the conceptual development of the mission, the observatory included a microcalorimeter detector that would not only yield high-quality

TABLE 1 Comparison of Mirror Assemblies of Various X-Ray Observatories

Observatory, Approximate Cost	Effective Area @ 1, 6 keV, cm ²	Angular Resolution (arc-second, half-power-diameter)	No. of Shells per Telescope, Maximum Diameter (m), f-number	Mirror Substrate Material	Total No. of Mirror Elements in Observatory
Chandra \$2.5B	800, 300	<0.5	4, 1.4, f/7	16–24 mm Zerodur	4
XMM-Newton \$640M	4,500, 2,700	12–15	58, 0.7, f/11	0.5–1-mm electroless nickel	174
Suzaku \$140M	1,750, 1,200	110	180, 0.4, f/11	0.17-mm segmented aluminum	5,760
Constellation-X ~\$2.5B	15,000, 6,000	5–15	~200, 1.4, f/7	0.4-mm segmented glass	12,000–15,000

imaging, but unparalleled energy resolution in a nondispersive system. But budget restrictions led to the use of a less expensive grating array in order to achieve the spectroscopic energy resolution required. Budget restrictions also meant that only four out of the originally envisioned six mirror shells could be manufactured and flown. The focal length of the system was chosen to be 10 m—to give an f-number of approximately $f/7$. This is a relatively fast compared to most x-ray observatories, but being an imaging mission, there was also a need for a wide field of view in order to fully capture extended sources in a single exposure.

During the development of the XMM-Newton observatory, the science team stressed effective collecting area over imaging. To quote the XMM-Newton web page,* “The design of the optics was driven by the requirement of obtaining the highest possible effective area over a wide range of energies, with particular emphasis in the region around 7 keV.” In contrast to Chandra, XMM-Newton was designed with 58 nested shells per telescope, and has three identical telescopes on board the spacecraft. That, in conjunction with a much slower $f/11$ optical system, means a higher effective collecting area—6 times more at 1 keV and nearly 10 times more at 6 keV—due to the shallower grazing angles improving the reflectivity at the higher energy range. But, the drawback is that each mirror shell needed to be between 0.5 and 1 mm thick in order to get them all to fit in the spacecraft volume, compared to the 16- to 24-mm-thick shells on Chandra. Figuring and polishing so many thin substrates would have taken far too long and cost too much, so a replication process was used to manufacture the shells. Fifty-eight thick negative masters, termed “mandrels,” were figured and polished with their outer surface matching the desired inner surface of each shell. The masters were then coated with a gold release layer, and electroless nickel was deposited on top of the gold, forming the mirror shells. The mirrors and mandrels were then separated through a cryogenic process that did not distort the mirror shells. But, since each mirror shell is so thin, and the time and budget available for figuring each mandrel was not unlimited, the imaging quality of final telescopes range from 12 to 15 arc-second HPD.

The Suzaku observatory is the third case to be examined. It is a small mission, with a relatively limited budget. The primary goal of this mission was to perform high-resolution spectroscopy by flying five telescopes with upward of 200 nested shells. Four of the mirror assemblies use CCD detectors as their imaging focal plane, similar to both Chandra and XMM-Newton, but the fifth uses a microcalorimeter to achieve unprecedented 12-eV resolution in a nondispersive system. Due to the relatively small volume of the spacecraft, each of the five mirror assemblies is contained within a 40-cm diameter. Each shell is made up of four 90° segments of revolution, and the reflecting substrates are 0.17-mm-thick aluminum. Obviously, it would be impractical to figure each of thousands of very thin mirrors, and with the budget available it would even be too costly to figure the forming

*http://xmm.esac.esa.int/external/xmm_user_support/documentation/technical/Mirrors/index.shtml

mandrels used to thermally shape the aluminum substrates, so a conical approximation to the Wolter-I geometry is used. This fact and the deformations associated with mounting and integration of the very thin aluminum, result in poorer imaging quality (arc-minute scale HPD), but the system yields more collecting area as Chandra in a significantly smaller package and at $\sim 1/20$ th of the cost.

Finally, the next large x-ray observatory under development at NASA is Constellation-X. This mission will concentrate on high-resolution spectroscopy in the soft x-ray band. Proposed “science enhancement packages” include more sophisticated multiple layer coatings to increase the low-energy effective area, dedicated hard x-ray telescopes, various grating arrays to improve low-energy spectral resolution, and so on. The core mission, however, includes four large Wolter-I telescopes, each 1.4 m in diameter coupled to imaging microcalorimeter arrays. The mission is designed to be a revolutionary step in capability, as opposed to an evolutionary step, as it will be 3 to 4 times larger than any previous mission and use an imaging detector with >30 times better energy resolution than any previously successfully flown. The primary trade-off is more effective collecting area, at the expense of poorer angular resolution than missions with individually polished mirror shells. In light of the science results achieved by Chandra, however, at the time of this writing, the science team is strongly considering changing the imaging requirement from 15 to 5 arc-second HPD. Whether any of the other requirements are adjusted to compensate for the tighter imaging requirement is still under debate.

33.3 SUMMARY

As can be seen in Table 1, there are a variety of competing restrictions in building a large x-ray observatory. Among the science questions that drive the design of any telescope are the breadth, signal strength, energy spectrum, and variability of the observation targets in question. Within the currently envisioned upcoming observatories^{*,†,‡,§} the angular resolution requirement varies from 2 arc-seconds to 1 arc-minute. The energy spectra of interest are as low as 0.25 to 10 keV and as high as 6 to 80 keV. The resulting telescopes have focal distances ranging from 10 to 30 m, and have 600 to 15,000 cm² of effective collecting area. From these contrasts, it is clear that there is no one-size-fits-all solution to the questions of telescope imaging quality, field of view, energy resolution, and so on, especially when the various factors of cost, launch vehicle mass budget, energy budget, volume, and schedule are taken into account. The number of variables than can be traded-off against one another mean that for each mission concept, a significant effort needs to be made to balance all of these factors in a way that results in a telescope tailored to the science goals at hand and within the given budgets. A very short list is given above that demonstrates the wide range of mission concepts, budgets and capabilities. But, while any space flight mission is costly, as a result of careful configuration trade-off studies, all are valuable to the astronomical community.

^{*}Simbol-X—<http://www.asdc.asi.it/simbol-x/>

[†]XEUS—<http://sci.esa.int/science-e/www/area/index.cfm?fareaid=103>

[‡]Constellation-X—<http://constellation.gsfc.nasa.gov/>

[§]<http://www.nustar.caltech.edu/>

EXTREME ULTRAVIOLET LITHOGRAPHY

Franco Cerrina and Fan Jiang

*Electrical and Computer Engineering & Center for NanoTechnology
University of Wisconsin
Madison, Wisconsin*

34.1 INTRODUCTION

The evolution of the integrated circuit is based on the continuous shrinking of the dimensions of the devices, generation after generation of products. As the size of the transistors shrink, the density increases and so does the functionality of the chip. This trend is illustrated in Table 1, derived from the Semiconductor Industry Association (SIA) forecasts of the Semiconductor Technology Roadmap.¹ During the fabrication process, optical imaging (optical lithography [OL]) is used to project the pattern of the circuit on the silicon wafer. These are exceedingly complex images, often having more than 10^{12} pixels. OL provides very high throughput thanks to the parallel nature of the imaging process with exposure times of less than 1 sec. In semiconductor manufacturing the ArF laser line (at 193 nm) is the main source of radiation used in OL. The resolution of the imaging systems has been pushed to very high levels, and it is very remarkable how it is possible to image patterns on large areas with dimensions smaller than 1/5th of the wavelength: Rayleigh's criterion ($\Delta \approx \lambda/2\text{NA}$) can be defeated by using super-resolution techniques, commonly referred to as resolution enhancement techniques (RET).² This is possible because in patterning super-resolution techniques can be applied more effectively than in microscopy or astronomy.³ All of this—throughput and resolution—makes OL today's dominant technology. However, even with the best of wavefront engineering it is very unlikely that features as small as 19 nm and less can be imaged with 193 nm—ultimately, the wavelength is the limiting factor in our ability to image and pattern. Hence, a shorter wavelength is needed to continue to use optical patterning techniques, or some variation thereof based on the paradigm of projecting the pattern created on a mask. For various reasons the 157-nm line of the F_2 laser is not suitable, and no other effective sources nor optical solutions exist in wavelength region from 150 to 15 nm—the vacuum ultra violet region, today often called extreme ultraviolet (EUV). Notably, in this part of the spectrum the index of refraction of all materials has a very large imaginary part due to the photoexcitation of valence band states. Thus all materials are very absorptive, and as a consequence transmission optics cannot be implemented. This has far-reaching implications for lithography.

Extreme ultraviolet lithography (EUV-L) is one of the latest additions to the list of patterning techniques available to the semiconductor industry;^{4,5} it aims to maintain the basic paradigm of pattern projection in the context of EUV-based technology. Table 2 lists some of the main goals of these imaging systems.

TABLE 1 Critical Dimensions of Devices in Integrated Circuits

Year	2008	2010	2012	2014
CD (nm)	38	30	24	19
Wavelength (nm)	193	193	13.4	13.4

TABLE 2 EUV Lithography Exposure Tool Parameters

Generation	25 – 12-nm CD
Wavelength	13.4 nm
NA	0.3 – 0.4
Field size	10–25 mm
Power density	10 mW/cm ²
Wafer size	300 mm

Because it is not possible to use any form of transmission optics (i.e., lenses) the imaging systems have to rely on mirrors. Very-high-resolution optical systems based on near-normal incidence designs are possible that use at least 4 to 6 reflecting surfaces,⁶ with aspheres further improving imaging. However, the reflectivity of a surface in the VUV is of the order of 10^{-3} to 10^{-4} , too small by order of magnitudes. As pointed out by Spiller,⁷ it is possible to use multilayer coatings to enhance the reflectivity even if the materials are highly absorbing. Indeed, the choice of the wavelength of 13.4 nm for EUV-L is dictated by the availability of efficient multilayer reflectors. These are essentially $\lambda/4$ stacks of molybdenum and silicon, producing a reflectivity higher than 70% in the region of $\lambda = 10 - 15$ nm. As mentioned above, in the VUV all materials are strongly absorbing, so that the propagation in the stack is limited to 40 to 60 bilayers,⁷ and this in turn limits the maximum reflectivity of the stack. Contrary to $\lambda/4$ dielectric mirrors, the bandpass is fairly large ($\sim 10\%$) and the maximum reflectivity does not go beyond 70 to 75%. Thus, the optical system is very lossy, with a transmission of $T \sim 0.75^6 = 0.18$, and high-power sources are required to achieve the final power density. Additionally, the multilayers must have extremely smooth interfaces to reduce incoherent scattering that worsens the quality of imaging by introducing flare.⁸ We note that while glancing angle mirrors can be used very effectively, the aberrations of these strongly off-axis systems forbid their use in large area imaging.

34.2 TECHNOLOGY

Like all imaging systems, EUV lithography requires a source of radiation, a condenser to uniformly illuminate the mask, a mask with the layout to be exposed, imaging optics and mechanical stage for step and repeat operation.^{4,5,9}

- While synchrotrons are by far the most efficient sources of EUV radiation, for manufacturing a more compact source based on the emission from plasma is more appealing. There are two main types of sources, one based on electrical discharge in a low-pressure gas and the other based on a laser-initiated plasma.¹⁰
- The condenser is a fairly complex optical system designed to match the source phase space to the optics. It includes both near-normal incidence mirrors (Mo-Si) and glancing mirrors. Since the sources are still relatively weak, it is important that the condenser collects as large a fraction as possible of the source phase space.¹¹

- The mask is formed by a patterned absorber (equivalent to 20- to 40-nm Cr) deposited on a multilayer stack (60 to 80 bilayers) formed on a thick quartz or zerodur substrate. The pattern is $4\times$ the final image, thus relaxing somewhat the constraints on the resolution and accuracy of the layout.
- The imaging optics include a set of near-normal incidence mirrors. In the first designs, the number of mirrors was 4, becoming 6 in more recent designs to increase the final NA of the system. Designs with up to 8 mirrors are also being considered. The first generation of tools operated at 0.25 NA, and the second generation works at 0.3 NA, with a goal of reaching $NA \sim 0.4$.^{5,6,12}
- The step-and-scan stage is an essential part of the exposure system, since the relatively small field of view of the optics must be repeated over the whole wafer surface. The wafer is typically held in place by an electrostatic chuck, and the motion is controlled by laser interferometers.⁹
- The whole assembly is enclosed in a high-vacuum system, as mandated by the high-absorption coefficient of EUV radiation in any media, including air.
- The photoresist is the material used to record the image projected on the wafer. These materials are very similar to those already in use in OL, and are based on the concept of chemical amplification to increase their sensitivity to the exposing radiation.¹³

Figure 1 shows such a tool, developed at Sandia National Labs¹⁴ and installed at the Advanced Light Source, Berkeley. Figure 2 shows a photograph of one of the ASML Alpha Tools, installed at SUNY-Albany and at IMEC.⁸ Notice the vacuum enclosure and the overall large size of the lithography system. The ASML tool uses a plasma source. These tools are used to study advanced devices, and to develop the processes required for device manufacturing.

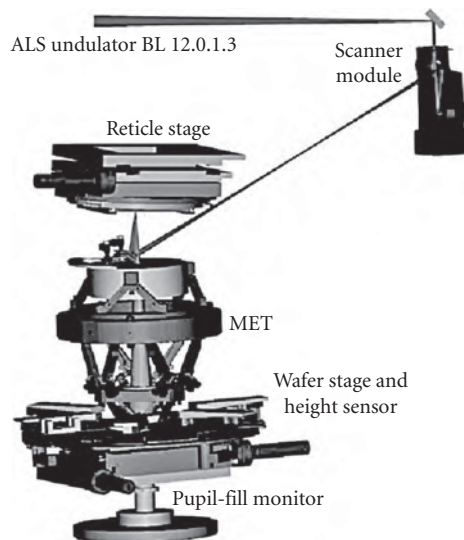


FIGURE 1 EUV exposure tool. The design includes 4 mirrors, and the mask (reticle) and wafer location. The whole system is under vacuum, and the condenser optical system is not shown.¹⁴ The whole system is enclosed in a vacuum chamber, not shown. The overall size is of several meters. This tool is installed at the Advanced Light Source, Berkeley. (See also color insert.)



FIGURE 2 ASML alpha demo tool during installation. The whole system is under vacuum, and a plasma source is used to generate the EUV radiation (not shown).⁸ (See also color insert.)

EUV-Interferometric Lithography (EUV-IL)

In addition to the imaging steppers described above, there are other ways of creating very-high-resolution patterns, in particular for materials-oriented studies where simple structures are sufficient. Interference can be used to produce well-defined fringe patterns such as linear gratings, or even zone plates—in one word, to pattern.^{15–18} Interferometric lithography (IL) is the process of using the interference of two or more beams to form periodic patterns of fringes to be recorded in an imaging material—originally a photographic film, today often a photoresist. IL is useful for the study of the properties of the recording material because it forms well-defined, simple, and high-resolution periodic fringe patterns over relatively large areas. The recording properties of a photoresist material can be studied by measuring its response to a simple sinusoidal intensity distribution; a series of exposures of different period allows us to simply and quickly determine its ability to record increasingly finer images.¹⁹ From this, we can project its ability to later resolve more complex and nonperiodic patterns. IL allows the study and optimization of resist materials without the need of high-resolution optics and complex masks—the diffraction provides the high-resolution fringe patterns needed. These arguments led us to the development of high-resolution EUV-IL as a platform for the development of the materials needed for nanolithography, well beyond the reach of the imaging ability of the current generation of imaging tools (steppers).^{16,17}

In Fig. 3 we present some of the high-resolution exposures from our system.^{16,17} One of the advantages of EUV-IL is that the interference pattern period is half that of the original grating, thus considerably facilitating the fabrication process—a 25-nm period (12.5-nm lines and spaces) is generated from

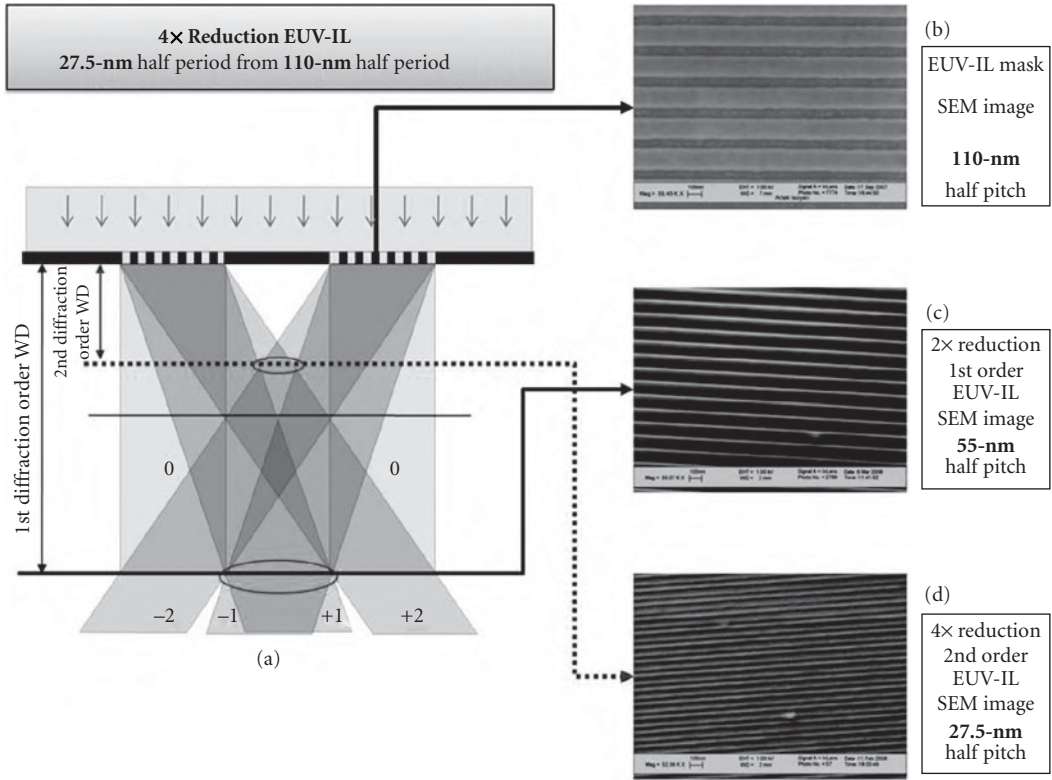


FIGURE 3 EUV interferometric lithography. The diffraction gratings are illuminated by a synchrotron, and the diffracted beams interfere as shown. The beams overlap creating 1st and 2nd order interference patterns of excellent visibility. Right, SEM images of the grating, and of the first- and second-order exposures. Notice the relative period of the images—the 1 $\times$ period is half of that of the diffracting grating, and the 2 $\times$ is 1/4.¹⁷ (See also color insert.)

a 50-nm period grating. The use of second-order interference increase this leverage by another factor of 2, so that a 50-nm period grating could in principle generate 12.5-nm periodic fringes, that is, 6.25-nm lines and spaces without the need of complex optical systems.

34.3 OUTLOOK

The continuing evolution of optical lithography has pushed EUV farther in the future. At the time of writing (2009), industry sources do not expect that EUV will be needed until the 17-nm node.^{4,5} Among the most active developer of systems are Nikon (Japan) and ASML (Belgium). There are several issues that are still unsolved, or only partially solved. Specifically:

- **Sources**—The power delivered by current plasma sources is too low by at least one order of magnitude. The photoresist should have a sensitivity of 5 mJ/cm², thus requiring at least 1 to 5 mW/cm² of power delivered to the resist surface.¹⁰
- **Source debris**—Since the EUV radiation is generated from hot plasmas (with the exception of synchrotrons) the elimination of reduction of debris ejected by the plasma is a major concern.

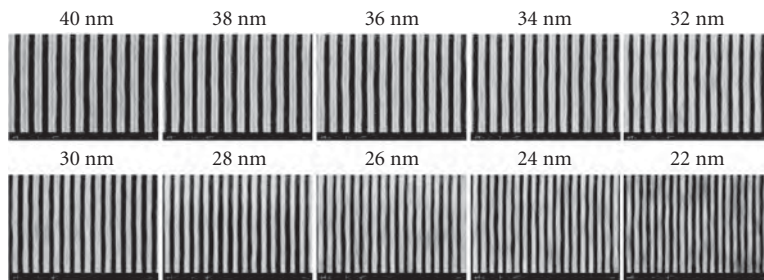


FIGURE 4 Twenty-two-nm dense lines printing of chemically amplified resists achieved at the Advanced Light Source facility. High-quality imaging is demonstrated, but line-edge roughness appears on the higher-resolution patterns.¹⁴

- **Masks**—Much progress has been made in this area, but the issue of defects has not yet been fully solved. Specifically, defects submerged in the reflecting stack underlying the absorber pattern create phase errors that affect severely the image.²⁰ These “buried defects,” only a few nanometer in size, are very difficult to detect short of a full-mask at-wavelength inspection, an expensive and time-consuming task.
- **Photoresist materials**—There is as yet no resist material that can balance the contrasting requirements of sensitivity, resolution, and low-edge roughness.²¹ As shown in Figs. 3 and 4 at the highest resolution the photoresist lines become less smooth, showing a degree of graininess (line edge roughness [LER]) that makes the pattern less clearly defined. The LER is a major problem at these small dimensions, since current materials have LER of ~ 5 to 7 nm, unacceptably large for patterns of 20 nm. Much work is going on in trying to reduce the LER in resist.
- **Cost and infrastructure**—The costs of the exposure tools and masks have increased dramatically from the original projections of the mid-90s. Today, the cost of a stepper is projected to be well above US\$ 100 million, with the cost of a single EUV mask projected to more than US\$ 200 thousand. These costs are substantially higher than those of OL. Another challenge is the development of the required infrastructure network of suppliers for mask blanks, sources, tools, and so forth. The very specialized nature of EUV technology is making the development of this network slow and difficult.

There is a considerable amount of activity in all these areas, particularly in the development of more powerful sources, better resist materials and more effective mask defect inspection and repair techniques. The optics, multilayer coatings, and mechanical stages are well developed, and do not appear to be show stoppers. Industrial, academic, and national laboratory research remains strong.

In summary, at the time of writing EUV lithography has not yet reached its full potential. The appeal of a parallel imaging system capable of delivering a patterning resolution able to support sub-20-nm imaging is very strong, and explains the continuing interest of the industry in supporting the development of the EUV-L technology. Other techniques that rely mostly on electron beams in various forms of parallelization have not yet been demonstrated as a convincing alternative, while earlier precursors like proximity x-ray lithography have been discontinued. It is reasonable to expect that EUV will mature in the time frame 2008 to 2012, reaching a fully developed system in 2014 to 2015.

34.4 ACKNOWLEDGMENTS

We thank P. Naulleau (U.C. Berkeley) and G. Lorusso (IMEC, Belgium) for sharing the pictures presented in the chapter. This work would not have been possible without the support of the staff and students of the Center for Nano Technology at the University of Wisconsin.

34.5 REFERENCES

The most updated sources of information on EUV-IL can be found in the proceedings of the main lithography conferences: SPIE (www.spie.org), EIPBN (www.eipbn.org), MNE (mne08.org) and MNC (imnc.jp), as well as in the Sematech (www.sematech.org) and SIA (www.itrs.net) Web pages.

1. [http://www.itrs.net/Links/2007ITRS/2007 Chapters/2007 Lithography.pdf](http://www.itrs.net/Links/2007ITRS/2007%20Chapters/2007%20Lithography.pdf).
2. C. Mack, *Fundamental Principles of Optical Lithography* Wiley, New York, 2008.
3. G. T. di Francia, "Resolving Power and Information," *J. Opt. Soc. Amer.* **45**:497, 1955.
4. S. Wurm, C. U. Jeon, and M. Lercel, "SEMATECH's EUV Program: A Key Enable for EUVL Introduction," *SPIE Proceeding* **6517**(4), 2007.
5. I. Mori, O. Suga, H. Tanaka, I. Nishiyama, T. Terasawa, H. Shigemura, T. Taguchi, T. Tanaka, and T. Itani, "Selete's EUV Program: Progress and Challenges," *SPIE Proceeding* **6921**(1), 2008.
6. T. E. Jewell, *OSA Proceeding on EUV Lithography* in F. Zernike and D. Attwood (eds.) **23**:98, 1994 and reference therein.
7. E. Spiller, *Soft X-Ray Optics*, SPIE Optical Engr. Press, Bellingham, Wash., 1994.
8. G. F. Lorusso, J. Hermans, A. M. Goethals, et al., "Imaging Performance of the EUV lpha Demo Tool at IMEC," *SPIE Proceeding* **6921**(24), 2008.
9. N. Harned, M. Goethals, R. Groeneveld, et al., "EUV Lithography with the Alpha Demo Tools: Status and Challenges," *SPIE Proceeding* **6517**(5), 2007.
10. U. Stamm, "Extreme Ultraviolet Light Sources for Use in Semiconductor Lithography—State of the Art and Future Development," *Journal of Physics D: Applied Physics* **37**(23):3244–3253, 2004.
11. K. Murakami, T. Oshino, H. Kondo, H. Chiba, H. Komatsuda, K. Nomura, and H. Iwata, "Development Status of Projection Optics and Illumination Optics for EUV1," *SPIE Proceeding* **6921**(26), 2008.
12. T. Miura, K. Murakami, K. Suzuki, Y. Kohama, K. Morita, K. Hada, Y. Ohkubo, and H. Kawai, "Nikon EUVL Development Progress Update," *SPIE Proceeding* **6921**(22), 2008.
13. T. Wallow, C. Higgins, R. Brainard, K. Petrillo, W. Montgomery, C. Koay, G. Denbeaux, O. Wood, and Y. Wei, "Evaluation of EUV Resist Materials for Use at the 32 nm Half-Pitch Node," *SPIE Proceeding* **6921**(56), 2008.
14. P. Naulleau, UC Berkeley, private communication.
15. H. H. Solak, D. He, W. Li, S. Singh-Gasson, F. Cerrina, B. H. Sohn, X. M. Yang, and P. Nealey, "Exposure of 38 nm Period Grating Patterns with Extreme Ultraviolet Interferometric Lithography," *Applied Physics Letters* **75**(15):2328–2330, 1999.
16. F. Cerrina, A. Isoyan, F. Jiang, Y. C. Cheng, Q. Leonard, J. Wallace, K. Heinrich, A. Ho, M. Efremov, and P. Nealey, "Extreme Ultraviolet Interferometric Lithography: A Path to Nanopatterning," *Synchrotron Radiation News* **21**(4):12–24, 2008.
17. A. Isoyan, A. Wüest, J. Wallace, F. Jiang, and F. Cerrina, "4× Reduction Extreme Ultraviolet Interferometric Lithography," *Optics Express* **16**(12):9106–9111, 2008.
18. H. H. Solak, "Nanolithography with Coherent Extreme Ultraviolet Light," *Journal of Physics D: Applied Physics* **39**(10):R171–R188, 2006.
19. W. M. Moreau, *Semiconductor Lithography: Principles and Materials*, New York: Plenum, 1988.
20. H. Han, K. Goldberg, A. Barty, E. Gullikson, T. Ikuta, Y. Uno, O. Wood II, and S. Wurm, "EUV MET Printing and Actinic Imaging Analysis on the Effects of Phase Defects on Wafer CDs," *SPIE Proceeding* **6517**(10), 2007.
21. G. M. Gallatin, P. Naulleau, D. Niakoula, R. Brainard, E. Hassanein, R. Matyi, J. Thackeray, K. Spear, and K. Dean, "Resolution, LER, and Sensitivity Limitations of Photoresist," *SPIE Proceeding* **6921**(55), 2008.

This page intentionally left blank

RAY TRACING OF X-RAY OPTICAL SYSTEMS

Franco Cerrina

*Electrical and Computer Engineering & Center for NanoTechnology
University of Wisconsin
Madison, Wisconsin*

Manuel Sanchez del Rio

*European Synchrotron Radiation Facility
Grenoble, France*

35.1 INTRODUCTION

The first step before the construction of any x-ray system, such as a synchrotron beamline, is an accurate conceptual design of the optics. The beam should be transported to a given image plane (usually the sample position) and its characteristics should be adapted to the experimental requirements, in terms of flux monochromatization, focus, time structure, and so forth. The designer's goal is not only to verify compliance to a minimum set of requirements, but also to optimize matching between the source and the optics étendue (angle-area product) to obtain the highest possible flux.

The small difference between the index of refraction (see Chap. 36) of most materials and vacuum leads to critical angles of a few milliradians, and thus to the need for reflection glancing optics or diffraction systems (crystals and glancing gratings and multilayers). This complicates the job of the optical designer, because the aberrations become asymmetric and the power unequal. Yet, complex optical systems must be designed, with evermore stringent requirements. This is well exemplified in the quest for high-resolution x-ray microprobes and microscopes, or high-resolution phase contrast imaging systems. The traditional approach used in optical design often fails in x rays because of the differences in the approach to designing a visible versus an x-ray optical system. With few exceptions, most optical systems are either dioptric (e.g., lens-based) or catoptric (mirrors only) with a few catadioptric (mixed) systems for special applications. But most importantly, in the visible region the angle of incidence of the principal ray is always near the normal of the lens (or mirror), in what is often a paraxial optical system. By contrast, x-ray optical systems are strong off-axis systems, with angles of incidence close to 90 degrees.

Today, optical design relies more and more on computer simulation and optimization, and indeed very powerful programs such as CodeV (www.opticalres.com) and Zemax (www.zemax.com), to name only two, are widely used. These programs are however unwieldy when applied to the x-ray domain, particularly because of the off-axis geometry that makes cumbersome to define the geometry. A modeling code dedicated to the x-ray domain is thus a powerful tool for the optical designer. In addition, x-ray sources are peculiar, ranging from synchrotron bending magnets, to insertion devices and free electron lasers.

Often, when dealing with synchrotron-based optical systems, the optical elements themselves are just a subsystem of a complex beamline. Thus, the ability of modeling the progress of the radiation through the beamline in a realistic way is essential. The main question that an optical designer must answer is "Will the optical system deliver the performance needed by the experimental station?" The answer often

generates a second question, that is, “What is the effect of optics imperfections, thermal loading, material changes on the performance?” Clearly, the second question is “practical,” and yet essential in determining the performance of the beamline in a realistic world. This is where *SHADOW* was born: as a Monte-Carlo simulation code capable of ray tracing the progress of a beam of x rays (or any other photon beam) through a complex, sequential optical system. *SHADOW* models and predicts the properties of a beam of radiation from the source, through multiple surfaces, taking into account the physics in all reflection, refraction, and diffraction processes. While it is of general use, *SHADOW* was designed from the ground-up for the study of the glancing and diffractive optics typical of x-ray systems.

For synchrotron radiation applications, the code *SHADOW* has become the *de facto* standard because (1) it is modular and flexible, capable of adapting to any optical configurations, (2) it has demonstrated its reliability during more of 20 years of use, as shown in hundred of publications, (3) it is simple to use, and (4) it is in the public domain. Indeed, almost all of the synchrotron beamlines today in existence have in some way benefited from the help of *SHADOW*.^{*,†} From the list of selected references at the end of this chapter, the interested reader can form a good idea of the many uses of *SHADOW*.¹⁻¹¹

35.2 THE CONCEPTUAL BASIS OF *SHADOW*

The computational model used in *SHADOW* follows the evolution of a “beam of radiation.” This beam is a collection of independent “rays.” A “ray” is a geometrical entity defined by two vectors: a starting position $\mathbf{x} = (x_0, y_0, z_0)$ and the direction vector $\mathbf{v} = (v_x, v_y, v_z)$. In addition, a scalar, $k = 2\pi/\lambda$, defines the wavenumber and the electric field is described by two vectors $\mathbf{A}_\sigma$ and $\mathbf{A}_\pi$ for the two polarizations, with two scalars for the phases ϕ_σ and ϕ_π .

The first step in simulating an optical device with a ray-tracing code is the generation of the source, that is, the beam at the source position. The beam is a collection of rays, usually many thousands, which are created by a Monte Carlo sampling of the spatial (for starting positions) and divergences (for the direction) source distributions. *SHADOW* includes models for synchrotron (bending magnets, wigglers, and undulators) and geometrical (box, Gaussian, and so forth) sources. A detailed discussion of *SHADOW*’s source models can be found in Cerrina.¹² Although a single ray is monochromatic, with a well-defined wavelength, white and polychromatic beams are formed by a collection of rays with nonequal wave numbers, giving an overall spectral distribution. Essentially, *SHADOW* samples the wave-front of an ensemble of point sources. Rays are propagated (in vacuum or air) in straight lines, until they interact with the optical elements (mirrors or gratings). The set of optical elements constitute the optical system. The tracing is sequential: the beam goes to the first optical element, which modifies it, then it goes to a second element, and so on, until arriving at the final detection plane. The optical elements are defined by the equation of the mathematical surface (plane, sphere, ellipsoid, polynomials, and so forth), and the intercept point between each ray and the surface is calculated by solving the equation system of the straight line for the ray and for the surface (in general, a quadric equation, a toroidal or a polynomial surface). At the intercept point, the normal to the surface $\mathbf{n}_D$ is computed from the gradient. The intercept point is the new starting point for the ray after the interaction with the optical element. The direction is changed following either the specular reflection law (for mirrors) or the boundary conditions at the surface (for gratings and crystals). These equations are written in vector form to allow calculations in 3D and improve efficiency (vector operations are faster in a computer than trigonometric calculations). A (compact) vector notation for the specular reflectivity can be written as $\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{in}} - 2(\mathbf{v}_{\text{in}} \cdot \mathbf{n}_D)\mathbf{n}_D$, where all the vectors are unitary. The change in the direction of a monochromatic beam *diffracted* by an optical surface can be calculated (i) using the boundary condition at the surface $\mathbf{k}_{\text{out},\parallel} = \mathbf{k}_{\text{in},\parallel} + \mathbf{G}_\parallel$, where $\mathbf{G}_\parallel$ is the projection of the reciprocal lattice vector $\mathbf{G}$ onto the optical surface, and (ii) selecting $\mathbf{k}_{\text{out},\perp}$ parallel to $\mathbf{k}_{\text{in},\perp}$, which guarantees the elastic scattering in the diffraction process, that is, the conservation of momentum $|\mathbf{k}_{\text{out}}| = |\mathbf{k}_{\text{in}}|$.

^{*}The executables for *SHADOW* and related files can be downloaded from http://www.nanotech.wisc.edu/CNT_LABS/shadow.html.

[†]The graphical interface to *SHADOW* developed at ESRF can be downloaded from <http://www.esrf.eu/UsersAndScience/Experiments/TBS/SciSoft/xop2.3/shadowvui/>.

In a diffraction grating the scattering vector is originated by the ruling, with $\mathbf{G}_{\parallel} = (m2\pi/p)\mathbf{u}_g$, p being the local period (`SHADOW` takes into account that p may not be constant along the grating surface), m the spectral order, and $\mathbf{u}_g$ a unitary vector tangent to the optical surface and pointing perpendicular to the grating grooves. It is straightforward to verify that these the scattering equations give the grating equation in its usual form: $m\lambda = p(\sin\alpha + \sin\beta)$ where $\alpha(\beta)$ is the angle of incidence (reflection) defined from the normal to the surface (see Chap. 38).

For crystals, $\mathbf{G}_{\parallel}$ is zero for the Bragg-symmetric crystals, when the crystal surface is parallel to the atomic planes. There is no additional scattering vector here, thus the Bragg-symmetric crystals are nondispersive systems. Any other crystal (asymmetric Bragg or any Laue) gives nonzero scattering: $\mathbf{G}_{\parallel} = (2\pi \sin \gamma / d_{hkl})\mathbf{u}_g$ being d_{hkl} the d-spacing of the selected crystal reflection, γ the angle between Bragg planes and surface, and now $\mathbf{u}_g$ a unitary vector tangent to the optical surface and pointing perpendicular to the lines created by the termination of the crystal planes at the surface. In other words, in crystal with a given asymmetry ($\gamma \neq 0$), the truncation of the Bragg planes with the crystal surface mimics a grating that is used to compute the changes in the direction of the beam. As $\mathbf{G}_{\parallel}$ is a function of λ , Bragg-asymmetric and Laue crystals are dispersive systems.

In addition to the change of direction of the ray at the optical surface, there is a change in amplitude and phase that can be computed using an adequate physical model. The ray electric-field vectors are then changed using Fresnel equations (for mirrors and lenses) or the dynamical theory of diffraction for crystals. In addition, nonidealities are easily prescribed. Roughness is described by the addition of a random scattering vector generated by a stochastic process; given a power spectrum of the roughness, the spectrum is sampled to generate a local “grating” corresponding to the selected spatial frequency. Finally, deterministic surface errors are included by adding a small correction term to the ideal surface, and finding the “real” intercept with an iterative method.

Slits are easily implemented: if the ray goes through the slit, it survives, otherwise it is discarded or rather labeled as “lost”. The beam (i.e., collection of rays) is scored at the detector plane, and several statistical tools (integration, histogramming, scatter and contour plots) may be used for calculating the required parameters (intensity, spatial and angular distributions, resolution, and so forth).

35.3 INTERFACES AND EXTENSIONS OF SHADOW

It is desirable to add a user-friendly graphical user interface (GUI) to `SHADOW` to help the user in analyzing the considerable output from the modeling. The `SHADOW` package includes a complete menu for defining the source and system parameters and basic plotting and conversion utilities, for displaying the results. A GUI written with free software (TCL-TK) is shipped with `SHADOW`,^{*} but much more sophisticated and powerful GUIs can be written using very powerful graphical commercial packages. Some users developed displaying utilities in Matlab, Mathematica, or IDL. A complete user interface (`SHADOWVUI`) written in IDL is freely available under the XOP package,^{13†} and it allows the user to run `SHADOW` using a multiwindow environment. This helps the user to modify the optical system, rerun `SHADOW` with modified inputs, and quickly refresh all the screens showing interesting information for the user (XY plots, histograms, etc.). Another powerful feature in `SHADOWVUI` is the availability of macros. These macros permit the user to run `SHADOW` in a loop, to perform powerful postprocessing, to make parametric calculations, and to compute a posteriori some basic operations (tracing in vacuum, vignetting, etc.). A beamline viewer application (BLViewer) helps in creating three-dimensional schematic views of the optical system. A tutorial with many examples is available.

`SHADOW` is an open code, and user contributions are not only possible but also encouraged. Whereas the `SHADOW` developers control the “official” code revisions and releases, anyone can contribute routines or interfaces. User contributions that are especially useful and of general applicability may be incorporated (with the obvious author’s agreement and collaboration) to the official version. This is effectively possible due to the file-oriented structure of `SHADOW`.

^{*}The executables for `SHADOW` and related files can be downloaded from http://www.nanotech.wisc.edu/CNT_LABS/shadow.html.

[†]The graphical interface to `SHADOW` developed at ESRF can be downloaded from <http://www.esrf.eu/UsersAndScience/Experiments/TBS/SciSoft/xop2.3/shadowvui/>.

35.4 EXAMPLES

More than 250 papers have been published describing the use of `SHADOW` in a broad range of applications. Here, we limit the discussion to a few examples to illustrate the use of program.

Monochromator Optics

Almost all synchrotron beamlines in the world have a monochromator to select and vary the energy of the photons delivered to the sample. Soft x-ray beamlines use grating monochromators, and hard x-ray beamlines use crystal monochromators. The performance of a monochromator is linked not only to the optical element itself, but also to other parameters of the beamline such as the source dimensions and divergences, slits and focusing that, in turn, can be done externally (upstream mirrors) or by using curved optical elements at the monochromator.

`SHADOW` calculates very accurately the aberrations and resolving power for all possible grating monochromators (PGM, SGM, TGM, SX700, DRAGON, etc.). For hard x-ray beamlines, a Si (alternatively Ge or diamond) double-crystal monochromator is commonly used. Bending magnet beamlines commonly use sagittal focusing with the second crystal. It is well known that the efficiency of the sagittal focusing depends on the magnification factor, and this can be efficiently computed by `SHADOW` (Fig. 1).

`SHADOW` also models more complex crystal monochromators, using crystals in transmission geometry for splitting the beam or for increasing efficiency at high energies. Polychromatic focusing (as in XAFS dispersive beamlines, as described in Chap. 30) or monochromatic focusing (including highly asymmetric back-scattering) can be analyzed with `SHADOW`. A detailed discussion of the crystal optics with `SHADOW` can be found in M. Sanchez del Rio.¹³

Mirror Optics

Mirrors are used essentially to focus or collimate the x-ray beam. They can also be used as low pass filters to reject the higher harmonics from diffractive elements or as x-ray filters to avoid high heat load on other devices such as the monochromators. Together with the monochromator, they constitute the most commonly used element in a beamline.

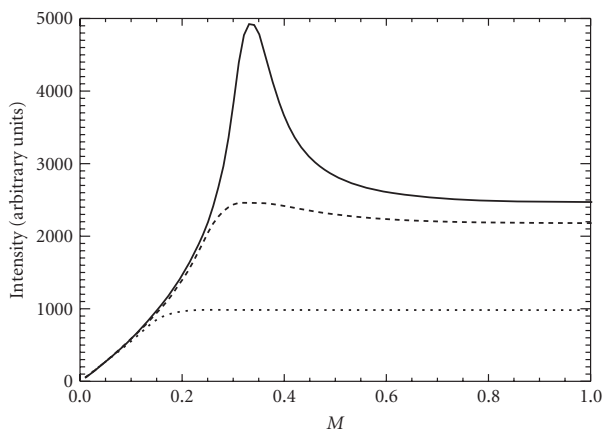


FIGURE 1 Intensity (in arbitrary units) versus magnification factor M for a point and monochromatic ($E = 20$ keV) source placed at 30 m from the sagittally Si bent crystals. Three beam divergences are considered: 1 mrad (lower), 2.5 mrad (middle), and 5 mrad (upper). We clearly observe the maximum of the transmission at $M = 0.33$ when focusing the 5-mrad beam, as predicted by the theory. (C. J. Sparks, B. S. Borie, and J. B. Hastings.¹⁴)

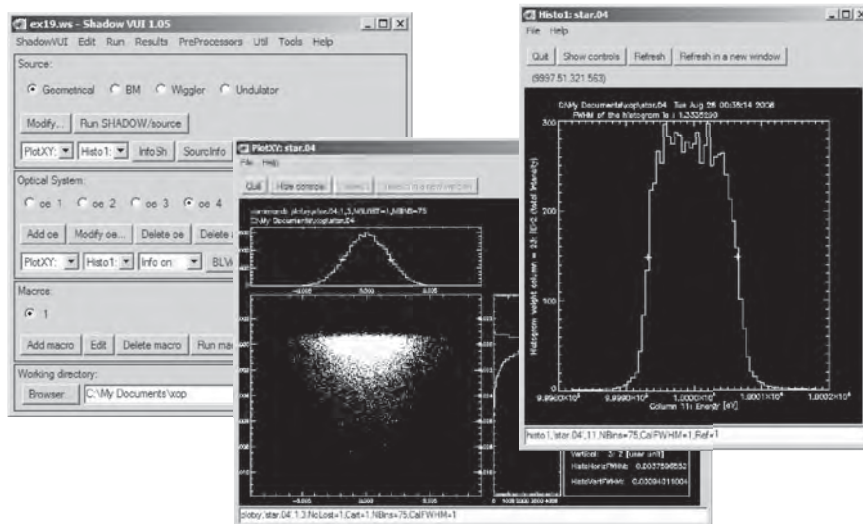


FIGURE 2 SHADOWVUI windows with the outputs of the ray tracing for the hard x-ray beamline (see text).

Reflection of x rays requires glancing incidence, which implies the use of very long mirrors (typically 30 to 100 cm). Glancing angles also magnify the effect of geometrical aberrations and surface irregularities (figure errors, slope errors, and roughness, as described in Chaps. 44 and 45). These effects are difficult to study with fully analytical methods, and it is even more difficult to analyze their combination with the geometrical and spectral characteristics of the synchrotron sources. The mirrors may be bent spherically (cylinders, spheres, and toroids) but elliptical curvature would be ideal in most cases. Spherical mirrors are usually preferred because of their lower cost and higher finishing quality. Dynamically bent mirrors are becoming very popular, and can also be used to compensate figure errors. SHADOW is well suited to perform reliable and accurate calculations of all these mirror systems under synchrotron radiation.

An important part of a ray-tracing calculation is the study of the tolerances to source movements, alignments, sample displacements, etc. SHADOW is well suited for these calculations because it allows the user to freely displace the source and the optical elements whilst conserving the same initial reference frame.

A Hard X-Ray Beamline

As an example, a full hard x-ray beamline has been raytraced with SHADOW. The undulator source can be simplified using Gaussian spatial ($\sigma_x = 0.57 \cdot 10^{-2}$ cm, $\sigma_y = 0.104 \cdot 10^{-2}$ cm) and angular distributions (horizontal $\sigma_x = 88.5$ μ rad, vertical $\sigma_y = 7.2$ μ rad), at $10,000 \pm 2$ eV (box distribution), with theoretical flux at this energy 5×10^{13} ph/sec/0.1%bw. The beamline has two mirrors, M1, a Rh-coated cylindrical collimating mirror in the vertical plane at 25 m from the source (glancing angle 0.12 mrad), and M2, a refocusing mirror (same coating and angle as M1) at 35 m from the source, focusing at the sample position. A Si (111) double crystal monochromator with second crystal sagittally bent is placed between the mirrors, at 30 m from source. The SHADOWVUI windows are shown in Fig. 2 and the resulting parameters are (i) Beam size at the sample position ($37.6 \times 9.4 \mu\text{m}^2$), (ii) energy resolution (1.33 eV), (iii) transmissivity of the whole beamline ($T = 0.85 \text{ eV}^{-1}$) and number of photons at the sample position (5.65×10^{12} photons/s).

35.5 CONCLUSIONS AND FUTURE

The SHADOW code is now about 20 years old, a statement to forward looking software design and continuing evolution. It has helped generations of postdocs and optical designers in developing

incredibly complex x-ray beamlines. While `SHADOW` performs very well, it begins to suffer from limitations, inherent in its kernel. For instance, the development of free electron laser sources requires a code capable of dealing with the sophisticated time structure, and derives coherence properties of these novel sources. The manifold increase in power of computers from the mid-80s, both in term of speed and memory space, makes this kind of calculation possible. Thus, `SHADOW`'s developers are planning to rewrite completely the code of the kernel (2008). The new structure will overcome technical and physical limitations of the current version. From the physical point of view, it is important to include in the simulation coherent beams, and the effect of the optical elements on them. Specifically, we need to have efficient tools for computing imaging and propagation of diffracted fields. This is important for many techniques already in use (e.g., phase contrast imaging) and will be essential for the new generation sources, which will be almost completely coherent. In the meantime `SHADOW` remains available to the x-ray optical community.*†

35.6 REFERENCES

1. L. Alianelli, M. Sanchez del Rio, M. Khan, and F. Cerrina, "A Comment on A New Ray-Tracing Program RIGTRACE for X-Ray Optical Systems," [*J. Synchrotron Rad.* 8:1047–1050 (2001)]. *Journal of Synchrotron Rad.* **10**:191–192, 2003.
2. B. Lai, K. Chapman, and F. Cerrina, "`SHADOW`: New Developments," *Nuclear Instruments & Methods in Physics Research Section A: Accelerators Spectrometers Detectors and Associated Equipment* **266**:544–549 (1988).
3. M. Sanchez del Rio, "Experience with Ray-Tracing Simulations at the European Synchrotron Radiation Facility," *Review of Scientific Instruments* **67**(9) (1996) [+CD-ROM].
4. M. Sanchez del Rio, "Ray Tracing Simulations for Crystal Optics," *Proceedings of the SPIE—The International Society for Optical Engineering* **3448**:230–245 (1998).
5. M. Sanchez del Rio, S. Bernstorff, A. Savoia, and F. Cerrina, "A Conceptual Model for Ray Tracing Calculations with Mosaic Crystals," *Review of Scientific Instruments* pt.11B **63**(1):932–935 (1992).
6. M. Sanchez del Rio and F. Cerrina, "Asymmetrically Cut Crystals for Synchrotron Radiation Monochromators," *Review of Scientific Instruments* pt. 11B **63**(1):936–940 (1992).
7. M. Sanchez del Rio and F. Cerrina, "Comment on 'Comments on the Use of Asymmetric Monochromators for X-Ray Diffraction on a Synchrotron Source,'" [*Rev. Sci. Instrum.* 66:2174 (1995)]. *Review of Scientific Instruments* **67**:3766–3767 (1996).
8. M. Sanchez del Rio and R. J. Dejus, "XOP 2.1—A New Version of the X-Ray Optics Software Toolkit," *American Institute of Physics Conference Proceedings* 705:784 (2004).
9. M. Sanchez del Rio, C. Ferrero, G. J. Chen, and F. Cerrina, "Modeling Perfect Crystals in Transmission Geometry for Synchrotron Radiation Monochromator Design," *Nuclear Instruments & Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment* **347**:338–43 (1994).
10. C. Welnak, P. Anderson, M. Khan, S. Singh, and F. Cerrina, "Recent Developments In `SHADOW`," *Review of Scientific Instruments* **63**:865–868 (1992).
11. C. Welnak, G. J. Chen, and F. Cerrina, "`SHADOW`—A Synchrotron-Radiation and X-Ray Optics Simulation Tool," *Nuclear Instruments & Methods in Physics Research Section A: Accelerators Spectrometers Detectors and Associated Equipment* **347**:344–347 (1994).
12. F. Cerrina, "Ray Tracing of X-Ray Optical Systems: Source Models," *SPIE Proceedings* **1140**:330–336 (1989).
13. M. Sanchez del Rio and R. J. Dejus, "XOP: Recent Developments," *SPIE Proceedings* **3448**:230–245, 340–345 (1998).
14. C. J. Sparks, B. S. Borie, and J. B. Hastings, "X-Ray Monochromator Geometry For Focusing Synchrotron Radiation above 10 keV," *Nuclear Instruments & Methods* **172**:237–224 (1980).

*The executables for `SHADOW` and related files can be downloaded from http://www.nanotech.wisc.edu/CNT_LABS/shadow.html.

†The graphical interface to `SHADOW` developed at ESRF can be downloaded from <http://www.esrf.eu/UsersAndScience/Experiments/TBS/SciSoft/xop2.3/shadowvui/>.

X-RAY PROPERTIES OF MATERIALS

Eric M. Gullikson

*Center for X-Ray Optics
Lawrence Berkeley National Laboratory
Berkeley, California*

The primary interaction of low-energy x rays within matter, namely, photoabsorption and coherent scattering, have been described for photon energies outside the absorption threshold regions by using atomic scattering factors, $f = f_1 + if_2$. The atomic photoabsorption cross section, μ , may be readily obtained from the values of f_2 using the relation

$$\mu = 2r_0\lambda f_2 \quad (1)$$

where r_0 is the classical electron radius, and λ is the wavelength. The transmission of x rays through a slab of thickness, d , is then given by

$$T = \exp(-N\mu d) \quad (2)$$

where N is the number of atoms per unit volume in the slab. The index of refraction, n , for a material is calculated by

$$n = 1 - Nr_0\lambda^2(f_1 + if_2)/(2\pi) \quad (3)$$

The (semiempirical) atomic scattering factors are based upon photoabsorption measurements of elements in their elemental state. The basic assumption is that condensed matter may be modeled as a collection of noninteracting atoms. This assumption is, in general, a good one for energies sufficiently far from absorption thresholds. In the threshold regions, the specific chemical state is important and direct experimental measurements must be made. Note also that the Compton scattering cross section is not included. The Compton cross section may be significant for the light elements ($Z < 10$) at the higher energies considered here (10 to 30 keV).

The atomic scattering factors are plotted in Figs. 1 and 2 for every 10th element. Tables 1 through 3 are based on a compilation of the available experimental measurements and theoretical calculations. For many elements there is little or no published data, and, in such cases, it was necessary to rely on theoretical calculations and interpolations across Z . To improve the accuracy in the future, considerably more experimental measurements are needed.¹ More data and useful calculation engines are available at www.cxro.lbl.gov/optical_constants.

36.1 X-RAY AND NEUTRON OPTICS

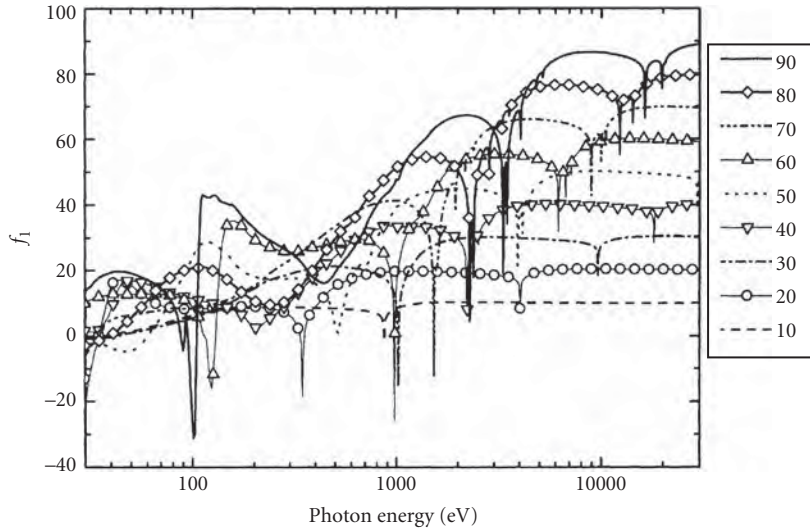


FIGURE 1 The atomic scattering factor f_1 as a function of photon energy from 30 to 30,000 eV for atomic number 10(Ne), 20(Ca), 30(Zn), 40(Zr), 50(Sn), 60(Nd), 70(Yb), 80(Hg), and 90(Th).

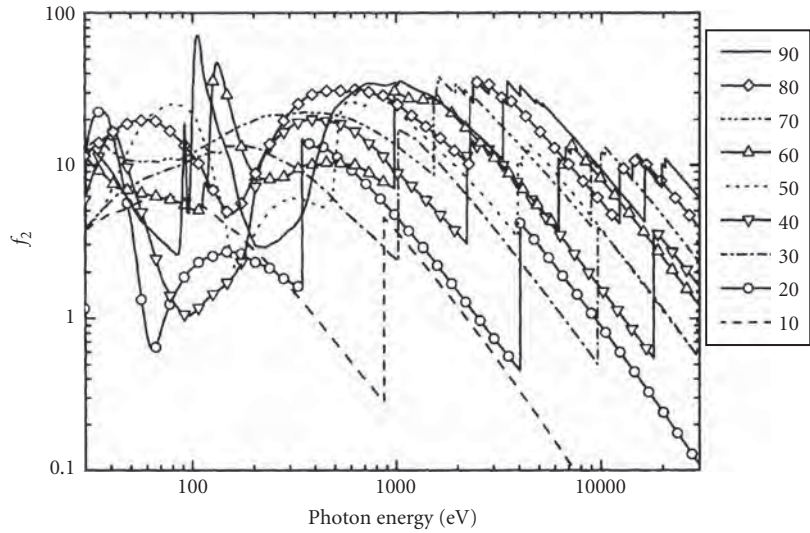


FIGURE 2 The atomic scattering factor f_2 as a function of photon energy from 30 to 30,000 eV for atomic number 10(Ne), 20(Ca), 30(Zn), 40(Zr), 50(Sn), 60(Nd), 70(Yb), 80(Hg), and 90(Th).

36.2 ELECTRON BINDING ENERGIES, PRINCIPAL K- AND L-SHELL EMISSION LINES, AND AUGER ELECTRON ENERGIES

TABLE 1 Electron Binding Energies in Electron Volts (eV) for the Elements in Their Natural Forms

Element	K1s	L ₁ 2s	L ₂ 2P _{1/2}	L ₃ 2P _{3/2}	M ₁ 3s	M ₂ 3P _{1/2}	M ₃ 3P _{3/2}	M ₄ 3d _{3/2}	M ₅ 3d _{5/2}	N ₁ 4s	N ₂ 4P _{1/2}	N ₃ 4P _{3/2}
1 H	13.6											
2 He	24.6*											
3 Li	54.7*											
4 Be	111.5*											
5 B	188*											
6 C	284.2*											
7 N	409.9*	37.3*										
8 O	543.1*	41.6*										
9 F	696.7*											
10 Ne	870.2*	48.5*	21.7*	21.6*								
11 Na	1070.8†	63.5†	30.4†	30.5*								
12 Mg	1303.0†	88.6*	49.6†	49.2†								
13 Al	1559.6	117.8*	72.9*	72.5*								
14 Si	1838.9	149.7*	99.8*	99.2*								
15 P	2145.5	189*	136*	135*								
16 S	2472	230.9*	163.6*	162.5*								
17 Cl	2822.4	270.2*	202*	200*								
18 Ar	3205.9*	326.3*	250.6*	248.4*	29.3*	15.9*	15.7*					
19 K	3608.4*	378.6*	297.3*	294.6*	34.8*	18.3*	18.3*					
20 Ca	4038.5*	438.4†	349.7†	346.2†	44.3†	25.4†	25.4†					
21 Sc	4492.8	498.0*	403.6*	398.7*	51.1*	28.3*	28.3*					
22 Ti	4966.4	560.9†	461.2†	453.8†	58.7†	32.6†	32.6†					
23 V	5465.1	626.7†	519.8†	512.1†	66.3†	37.2†	37.2†					
24 Cr	5989.2	695.7†	583.8†	574.1†	74.1†	42.2†	42.2†					
25 Mn	6539.0	769.1†	649.9†	638.7†	82.3†	47.2†	47.2†					
26 Fe	7112.0	844.6†	719.9†	706.8†	91.3†	52.7†	52.7†					
27 Co	7708.9	925.1†	793.3†	778.1†	101.0†	58.9†	58.9†					
28 Ni	8332.8	1008.6†	870.0†	852.7†	110.8†	68.0†	66.2†					
29 Cu	8978.9	1096.7†	952.3†	932.5†	122.5†	77.3†	75.1†					
30 Zn	9658.6	1196.2*	1044.9*	1021.8*	139.8*	91.4*	88.6*	10.2*	10.1*			
31 Ga	10367.1	1299.0*	1143.2†	1116.4†	159.5†	103.5†	103.5†	18.7†	18.7†			
32 Ge	11103.1	1414.6*	1248.1*	1217.0*	180.1*	124.9*	120.8*	29.0*	29.0*			
33 As	11866.7	1527.0*	1359.1*	1323.6*	204.7*	146.2*	141.2*	41.7*	41.7*			
34 Se	12657.8	1652.0*	1474.3*	1433.9*	229.6*	166.5*	160.7*	55.5*	54.6*			

(Continued)

TABLE 1 Electron Binding Energies in Electron Volts (eV) for the Elements in Their Natural Forms (*Continued*)

Element	K1s	L ₁ 2s	L ₂ 2p _{1/2}	L ₃ 2p _{3/2}	M ₁ 3s	M ₂ 3p _{1/2}	M ₃ 3p _{3/2}	M ₄ 3d _{3/2}	M ₅ 3d _{5/2}	N ₁ 4s	N ₂ 4p _{1/2}	N ₃ 4p _{3/2}
35 Br	13473.7	1782.0*	1596.0*	1549.9*	257*	189*	182*	70*	69*			
36 Kr	14325.6	1921.0	1730.9*	1678.4*	292.8*	222.2*	214.4	95.0*	93.8*	27.5*	14.1*	14.1*
37 Rb	15199.7	2065.1	1863.9	1804.4	326.7*	248.7*	239.1*	113.0*	112*	30.5*	16.3*	15.3*
38 Sr	16104.6	2216.3	2006.8	1939.6	358.7†	280.3†	270.0†	136.0†	134.2†	38.9†	20.3†	20.3†
39 Y	17038.4	2372.5	2155.5	2080.0	392.0*	310.6*	298.8*	157.7†	155.8†	43.8*	24.4*	23.1*
40 Zr	17997.6	2531.6	2306.7	2222.3	430.3†	343.5†	329.8†	181.1†	178.8†	50.6†	28.5†	27.7†
41 Nb	18985.6	2697.7	2464.7	2370.5	466.6†	376.1†	360.6†	205.0†	202.3†	56.4†	32.6†	30.8†
42 Mo	19999.5	2865.5	2625.1	2520.2	506.3†	411.6†	394.0†	231.1†	227.9†	63.2†	37.6†	35.5†
43 Tc	21044.0	3042.5	2793.2	2676.9	544*	445*	425*	257*	253*	68*	39†	39*
44 Ru	22117.2	3224.0	2966.9	2837.9	586.2†	483.5†	461.4†	284.2†	280.0†	75.0†	46.5†	43.2†
45 Rh	23219.9	3411.9	3146.1	3003.8	628.1†	521.3†	496.5†	311.9†	307.2†	81.4*	50.5†	47.3†
46 Pd	24350.3	3604.3	3330.3	3173.3	671.6†	559.9†	532.3†	340.5†	335.2†	87.6*	55.7†	50.9†
47 Ag	25514.0	3805.8	3523.7	3351.1	719.0†	603.8†	573.0†	374.0†	368.0†	97.0†	63.7†	58.3†
48 Cd	26711.2	4018.0	3727.0	3537.5	772.0†	652.6†	618.4†	411.9†	405.2†	109.8†	63.9†	63.9†
49 In	27939.9	4237.5	3938.0	3730.1	827.2†	703.2†	665.3†	451.4†	443.9†	122.7†	73.5†	73.5†
50 Sn	29200.1	4464.7	4156.1	3928.8	884.7†	756.5†	714.6†	493.2†	484.9†	137.1†	83.6†	83.6†
51 Sb	30491.2	4698.3	4380.4	4132.2	946†	812.7†	766.4†	537.5†	528.2†	153.2†	95.6†	95.6†
52 Te	31813.8	4939.2	4612.0	4341.4	1006†	870.8†	820.8†	583.4†	573.0†	169.4†	103.3†	103.3†
53 I	33169.4	5188.1	4852.1	4557.1	1072*	931*	875*	631*	620*	186*	123*	123*
54 Xe	34561.4	5452.8	5103.7	4782.2	1148.7*	1002.1*	940.6*	689.0*	676.4*	213.2*	146.7	145.5*
55 Cs	35984.6	5714.3	5359.4	5011.9	1211*	1071*	1003*	740.5*	726.6*	232.3*	172.4*	161.3*
56 Ba	37440.6	5988.8	5623.6	5247.0	1293*	1137*	1063*	795.7*	780.5*	253.5†	192	178.6†
57 La	38924.6	6266.3	5890.6	5482.7	1362*	1209*	1128*	853*	836*	247.7*	205.8	196.0*
58 Ce	40443.0	6548.8	6164.2	5723.4	1436*	1274*	1187*	902.4*	883.8*	291.0*	223.2	206.5*
59 Pr	41990.6	6834.8	6440.4	5964.3	1511.0	1337.4	1242.2	948.3*	928.8*	304.5	236.3	217.6
60 Nd	43568.9	7126.0	6721.5	6207.9	1575.3	1402.8	1297.4	1003.3*	980.4*	319.2*	243.3	224.6
61 Pm	45184.0	7427.9	7012.8	6459.3	—	1471.4	1356.9	1051.5	1026.9	—	242	242
62 Sm	46834.2	7736.8	7311.8	6716.2	1722.8	1540.7	1419.8	1110.9*	1083.4*	347.2*	265.6	247.4
63 Eu	48519.0	8052.0	7617.1	6976.9	1800.0	1613.9	1480.6	1158.6*	1127.5*	360	284	257
64 Gd	50239.1	8375.6	7930.3	7242.8	1880.8	1688.3	1544.0	1221.9*	1189.6*	378.6*	286	270.9
65 Tb	51995.7	8708.0	8251.6	7514.0	1967.5	1767.7	1611.3	1276.9*	1241.1*	396.0*	322.4*	284.1*
66 Dy	53788.5	9045.8	8580.6	7790.1	2046.8	1841.8	1675.6	1332.5	1292.6*	414.2*	333.5*	293.2*
67 Ho	55617.7	9394.2	8917.8	8071.1	2128.3	1922.8	1741.2	1391.5	1351.4	432.4*	343.5	308.2*
68 Er	57485.5	9751.3	9264.3	8357.9	2206.5	2005.8	1811.8	1453.3	1409.3	449.8*	366.2	320.2*
69 Tm	59398.6	10115.7	9616.9	8648.0	2306.8	2089.8	1884.5	1514.6	1467.7	470.9*	385.9*	332.6*
70 Yb	61332.3	10486.4	9978.2	8943.6	2398.1	2173.0	1949.8	1576.3	1527.8	480.5*	388.7*	339.7*
71 Lu	63313.8	10870.4	10348.6	9244.1	2491.2	2263.5	2023.6	1639.4	1588.5	506.8*	412.4*	359.2*
72 Hf	65350.8	11270.7	10739.4	9560.7	2600.9	2365.4	2107.6	1716.4	1661.7	538*	438.2†	380.7†
73 Ta	67416.4	11681.5	11136.1	9881.1	2708.0	2468.7	2194.0	1793.2	1735.1	563.4†	463.4†	400.9†
74 W	69525.0	12099.8	11544.0	10206.8	2819.6	2574.9	2281.0	1871.6	1809.2	594.1†	490.4†	423.6†

75	Re	71676.4	12526.7	11958.7	10535.3	2931.7	2681.6	2367.3	1948.9	1882.9	625.4	518.7 [†]	446.8 [†]
76	Os	73870.8	12968.0	12385.0	10870.9	3048.5	2792.2	2457.2	2030.8	1960.1	658.2 [†]	549.1 [†]	470.7 [†]
77	Ir	76111.0	13418.5	12824.1	11215.2	3173.7	2908.7	2550.7	2116.1	2040.4	691.1 [†]	577.8 [†]	495.8 [†]
78	Pt	78394.8	13879.9	13272.6	11563.7	3296.0	3026.5	2645.4	2201.9	2121.6	725.4 [†]	609.1 [†]	519.4 [†]
79	Au	80724.9	14352.8	13733.6	11918.7	3424.9	3147.8	2743.0	2291.1	2205.7	762.1 [†]	642.7 [†]	546.3 [†]
80	Hg	83102.3	14839.3	14208.7	12283.9	3561.6	3278.5	2847.1	2384.9	2294.9	802.2 [†]	680.2 [†]	576.6 [†]
81	Tl	85530.4	15346.7	14697.9	12657.5	3704.1	3415.7	2956.6	2485.1	2389.3	846.2 [†]	720.5 [†]	609.5 [†]
82	Pb	88004.5	15860.8	15200.0	13035.2	3850.7	3554.2	3066.4	2585.6	2484.0	891.8 [†]	761.9 [†]	643.5 [†]
83	Bi	90525.9	16387.5	15711.1	13418.6	3999.1	3696.3	3176.9	2687.6	2579.6	939 [†]	805.2 [†]	678.8 [†]
84	Po	93105.0	16939.3	16244.3	13813.8	4149.4	3854.1	3301.9	2798.0	2683.0	995*	851*	705*
85	At	95729.9	17493	16784.7	14213.5	4317	4008	3426	2908.7	2786.7	1042*	886*	740*
86	Rn	98404	18049	17337.1	14619.4	4482	4159	3538	3021.5	2892.4	1097*	929*	768*
87	Fr	101137	18639	17906.5	15031.2	4652	4327	3663	3136.2	2999.9	1153*	980*	810*
88	Ra	103921.9	19236.7	18484.3	15444.4	4822.0	4489.5	3791.8	3248.4	3104.9	1208*	1057.6*	879.1*
89	Ac	106755.3	19840	19083.2	15871.0	5002	4656	3909	3370.2	3219.0	1269*	1080*	890*
90	Th	109650.9	20472.1	19693.2	16300.3	5182.3	4830.4	4046.1	3490.8	3332.0	1330*	1168*	966.4 [†]
91	Pa	112601.4	21104.6	20313.7	16733.1	5366.9	5000.9	4173.8	3611.2	3441.8	1387*	1224*	1007*
92	U	115606.1	21757.4	20947.6	17166.3	5548.0	5182.2	4303.4	3727.6	3551.7	1439*	1271*	1043.0 [†]

Element	N ₄ d _{3/2}	N ₅ d _{5/2}	N ₆ 4f _{5/2}	N ₇ 4f _{7/2}	O ₁ 5s	O ₂ 5p _{1/2}	O ₃ 5p _{3/2}	O ₄ 5d _{3/2}	O ₅ 5d _{5/2}
48	Cd	11.7 [†]	10.7 [†]						
49	In	17.7 [†]	16.9 [†]						
50	Sn	24.9 [†]	23.9 [†]						
51	Sb	33.3 [†]	32.1 [†]						
52	Te	41.9 [†]	40.4 [†]						
53	I	50*	50*						
54	Xe	69.5*	67.5*	—	—	23.3*	13.4*	12.1*	
55	Cs	79.8*	77.5*	—	—	22.7	14.2*	12.1*	
56	Ba	92.6 [†]	89.9 [†]	—	—	30.3 [†]	17.0 [†]	14.8	
57	La	105.3*	102.5*	—	—	34.3*	19.3*	16.8*	
58	Ce	109*	—	—	—	37.8	19.8*	17.0*	
59	Pr	115.1*	115.1*	—	—	37.4	22.3	22.3	
60	Nd	120.5*	120.5*	—	—	37.5	21.1	21.1	
61	Pm	120	120	—	—	—	—	—	
62	Sm	129	129	—	—	37.4	21.3	21.3	
63	Eu	133	127.7*	—	—	31.8	22.0	22.0	
64	Gd	140.5	142.6*	—	—	43.5*	20	20	
65	Tb	150.5*	150.5*	—	—	45.6*	28.7*	22.6*	
66	Dy	153.6*	153.6*	—	—	49.9*	29.5	23.1	
67	Ho	160*	160*	—	—	49.3*	30.8*	24.1*	
68	Er	167.6*	167.6*	—	—	50.6*	31.4*	24.7*	
69	Tm	175.5*	175.5*	—	—	54.7*	31.8*	25.0*	

(Continued)

TABLE 1 Electron Binding Energies in Electron Volts (eV) for the Elements in Their Natural Forms (*Continued*)

Element	N ₄ 4d _{3/2}	N ₅ 4d _{3/2}	N ₆ 4f _{5/2}	N ₇ 4f _{7/2}	O ₁ 5s	O ₂ 5p _{1/2}	O ₃ 5p _{3/2}	O ₄ 5d _{3/2}	O ₅ 5d _{3/2}
70 Yb	191.2*	182.4*	—	—	52.0*	30.3*	24.1*		
71 Lu	206.1*	196.3†	8.9*	7.5*	57.3*	33.6*	26.7*		
72 Hf	220.0†	211.5†	15.9†	14.2†	64.2†	38*	29.9*		
73 Ta	237.9†	226.4†	23.5†	21.6†	69.7†	42.2*	32.7*		
74 W	255.9†	243.5†	33.6*	31.4†	75.6†	45.3*	36.8*		
75 Re	273.9†	260.5†	42.9*	40.5†	83†	45.6*	34.6*		
76 Os	293.1†	278.5†	53.4†	50.7†	84†	58*	44.5†		
77 Ir	311.9†	296.3†	63.8†	60.8†	95.2*	63.0*	48.0†		
78 Pt	331.6†	314.6†	74.5†	71.2†	101†	65.3*	51.7†		
79 Au	353.2†	335.1†	87.6†	83.9†	107.2*	74.2†	57.2†		
80 Hg	378.2†	358.8†	104.0†	99.9†	127†	83.1†	64.5†	9.6†	7.8†
81 Tl	405.7†	385.0†	122.2†	117.8†	136*	94.6†	73.5†	14.7†	12.5†
82 Pb	434.3†	412.2†	141.7†	136.9†	147*	106.4†	83.3†	20.7†	18.1†
83 Bi	464.0†	440.1†	162.3†	157.0†	159.3*	119.0†	92.6†	26.9†	23.8†
84 Po	500*	473*	184*	184*	177*	132*	104*	31*	31*
85 At	533*	507*	210*	210*	195*	148*	115*	40*	40*
86 Rn	567*	541*	238*	238*	214*	164*	127*	48*	48*
87 Fr	603*	577*	268*	268*	234*	182*	140*	58*	58*
88 Ra	635.9*	602.7*	299*	299*	254*	200*	153*	68*	68*
89 Ac	675*	639*	319*	319*	272*	215*	167*	80*	80*
90 Th	712.1†	675.2†	342.4†	333.1	290*	229*	182*	92.5†	85.4†
91 Pa	743*	708*	371*	360*	310*	232*	232*	94*	94*
92 U	778.3†	736.2†	388.2*	377.4†	321*	257*	192*	102.8†	94.2†

A compilation by G. P. Williams of Brookhaven National Laboratory, "Electron Binding Energies," in *X-Ray Data Booklet*, Lawrence Berkeley National Laboratory Pub-490 Rev. 2 (2001), based largely on values given by J. A. Bearden and A. F. Barr, "Re-evaluation of X-Ray Atomic Energy Levels," *Rev. Mod. Phys.* **39**:125 (1967); corrected in 1998 by E. Gullikson (LBNL, unpublished). The energies are given in electron volts relative to the vacuum level for the rare gases and for H₂, N₂, O₂, F₂, and Cl₂, relative to the Fermi level for the metals; and relative to the top of the valence bands for semiconductors.

*From M. Cardona and L. Lay (eds.), *Photoemission in Solids I: General Principles*, Springer-Verlag, Berlin (1978).

†From J. C. Fuggle and N. Mårtensson, "Core-Level Binding Energies in Metals," *J. Electron. Spectrosc. Relat. Phenom.* **21**:275 (1980).

For further updates, consult the Web site <http://xdb.lbl.gov/>.

TABLE 2 Photon Energies, in Electronvolts (eV), of Principal K- and L-Shell Emission Lines*

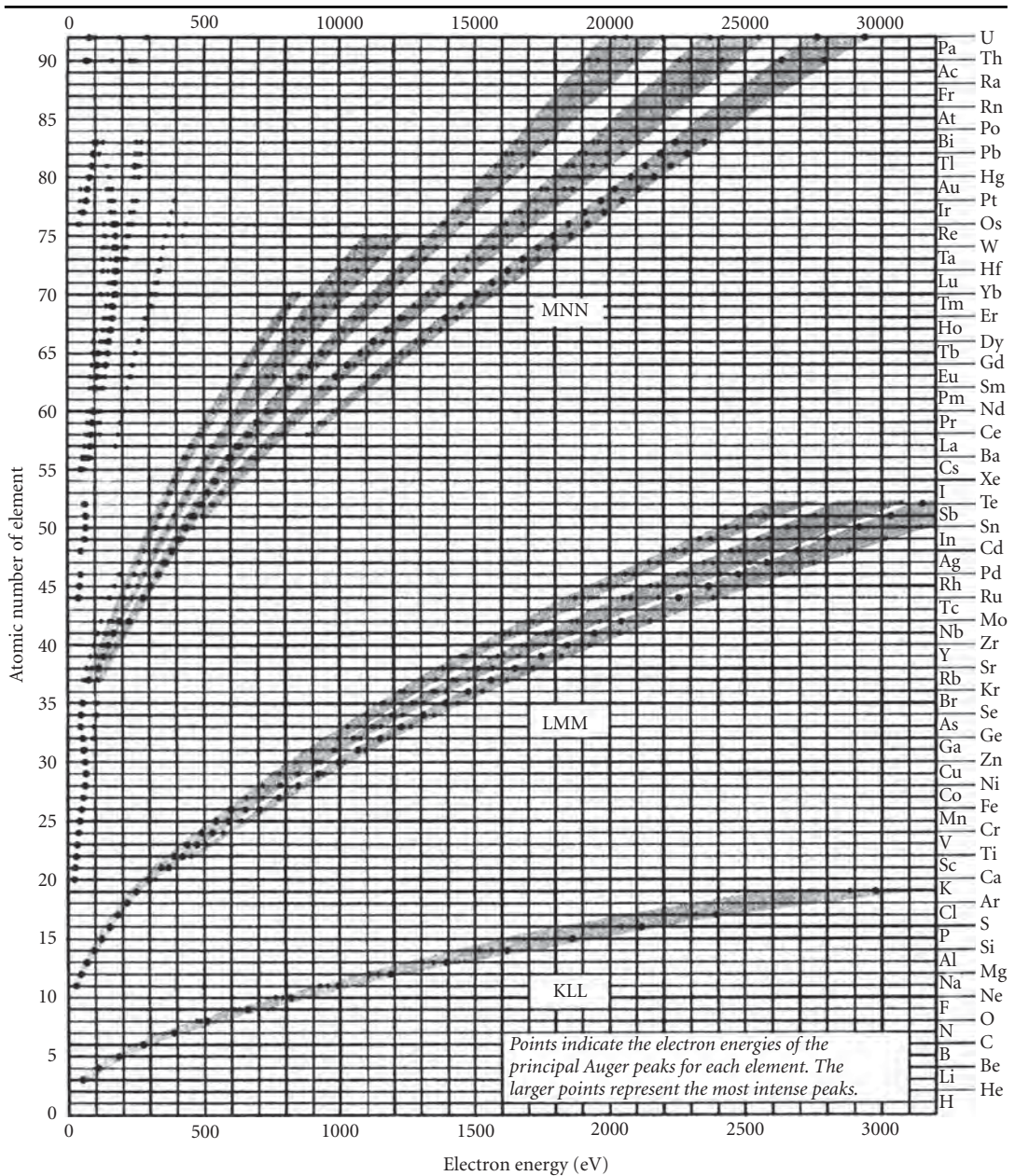
Element	$K\alpha_1$	$K\alpha_2$	$K\beta_1$	$L\alpha_1$	$L\alpha_2$	$L\beta_1$	$L\beta_2$	$L\gamma_1$
3 Li	54.3							
4 Be	108.5							
5 B	183.3							
6 C	277							
7 N	392.4							
8 O	524.9							
9 F	676.8							
10 Ne	848.6	848.6						
11 Na	1,040.98	1,040.98	1,071.1					
12 Mg	1,253.60	1,253.60	1,302.2					
13 Al	1,486.70	1,486.27	1,557.45					
14 Si	1,739.98	1,739.38	1,835.94					
15 P	2,013.7	2,012.7	2,139.1					
16 S	2,307.84	2,306.64	2,464.04					
17 Cl	2,622.39	2,620.78	2,815.6					
18 Ar	2,957.70	2,955.63	3,190.5					
19 K	3,313.8	3,311.1	3,589.6					
20 Ca	3,691.68	3,688.09	4,012.7	341.3	341.3	344.9		
21 Sc	4,090.6	4,086.1	4,460.5	395.4	395.4	399.6		
22 Ti	4,510.84	4,504.86	4,931.81	452.2	452.2	458.4		
23 V	4,952.20	4,944.64	5,427.29	511.3	511.3	519.2		
24 Cr	5,414.72	5,405.509	5,946.71	572.8	572.8	582.8		
25 Mn	5,898.75	5,887.65	6,490.45	637.4	637.4	648.8		
26 Fe	6,403.84	6,390.84	7,057.98	705.0	705.0	718.5		
27 Co	6,930.32	6,915.30	7,649.43	776.2	776.2	791.4		
28 Ni	7,478.15	7,460.89	8,264.66	851.5	851.5	868.8		
29 Cu	8,047.78	8,027.83	8,905.29	929.7	929.7	949.8		
30 Zn	8,638.86	8,615.78	9,572.0	1,011.7	1,011.7	1,034.7		
31 Ga	9,251.74	9,224.82	10,264.2	1,097.92	1,097.92	1,124.8		
32 Ge	9,886.42	9,855.32	10,982.1	1,188.00	1,188.00	1,218.5		
33 As	10,543.72	10,507.99	11,726.2	1,282.0	1,282.0	1,317.0		
34 Se	11,222.4	11,181.4	12,495.9	1,379.10	1,379.10	1,419.23		
35 Br	11,924.2	11,877.6	13,291.4	1,480.43	1,480.43	1,525.90		
36 Kr	12,649	12,598	14,112	1,586.0	1,586.0	1,636.6		
37 Rb	13,395.3	13,335.8	14,961.3	1,694.13	1,692.56	1,752.17		
38 Sr	14,165	14,097.9	15,835.7	1,806.56	1,804.74	1,871.72		
39 Y	14,958.4	14,882.9	16,737.8	1,922.56	1,920.47	1,995.84		
40 Zr	15,775.1	15,690.9	17,667.8	2,042.36	2,039.9	2,124.4	2,219.4	2,302.7
41 Nb	16,615.1	16,521.0	18,622.5	2,165.89	2,163.0	2,257.4	2,367.0	2,461.8
42 Mo	17,479.34	17,374.3	19,608.3	2,293.16	2,289.85	2,394.81	2,518.3	2,623.5
43 Te	18,367.1	18,250.8	20,619	2,424.0	—	2,536.8	—	—
44 Ru	19,279.2	19,150.4	21,656.8	2,558.55	2,554.31	2,683.23	2,836.0	2,964.5
45 Rh	20,216.1	20,073.7	22,723.6	2,696.74	2,692.05	2,834.41	3,001.3	3,143.8
46 Pd	21,177.1	21,020.1	23,818.7	2,838.61	2,833.29	2,990.22	3,171.79	3,328.7
47 Ag	22,162.92	21,990.3	24,942.4	2,984.31	2,978.21	3,150.94	3,347.81	3,519.59
48 Cd	23,173.6	22,984. j	26,095.5	3,133.73	3,126.91	3,316.57	3,528.12	3,716.86
49 In	24,209.7	24,002.0	27,275.9	3,286.94	3,279.29	3,487.21	3,713.81	3,920.81
50 Sn	25,271.3	25,044.0	28,486.0	3,443.98	3,435.42	3,662.80	3,904.86	4,131.12
51 Sb	26,359.1	26,110.8	29,725.6	3,604.72	3,595.32	3,843.57	4,100.78	4,347.79
52 Te	27,472.3	27,201.7	30,995.7	3,769.33	3,758.8	4,029.58	4,301.7	4,570.9

(Continued)

TABLE 2 Photon Energies, in Electronvolts (eV), of Principal K- and L-Shell Emission Lines* (*Continued*)

Element	K α_1	K α_2	K β_1	L α_1	L α_2	L β_1	L β_2	L γ_1
53 I	28,612.0	28,317.2	32,294.7	3,937.65	3,926.04	4,220.72	4,507.5	4,800.9
54 Xe	29,779	29,458	33,624	4,109.9	—	—	—	—
55 Cs	30,972.8	30,625.1	34,986.9	4,286.5	4,272.2	4,619.8	4,935.9	5,280.4
56 Ba	32,193.6	31,817.1	36,378.2	4,466.26	4,450.90	4,827.53	5,156.5	5,531.1
57 La	33,441.8	33,034.1	37,801.0	4,650.97	4,634.23	5,042.1	5,383.5	5,788.5
58 Ce	34,719.7	34,278.9	39,257.3	4,840.2	4,823.0	5,262.2	5,613.4	6,052
59 Pr	36,026.3	35,550.2	40,748.2	5,033.7	5,013.5	5,488.9	5,850	6,322.1
60 Nd	37,361.0	36,847.4	42,271.3	5,230.4	5,207.7	5,721.6	6,089.4	6,602.1
61 Pm	38,724.7	38,171.2	43,826	5,432.5	5,407.8	5,961	6,339	6,892
62 Sm	40,118.1	39,522.4	45,413	5,636.1	5,609.0	6,205.1	6,586	7,178
63 Eu	41,542.2	40,901.9	47,037.9	5,845.7	5,816.6	6,456.4	6,843.2	7,480.3
64 Gd	42,996.2	42,308.9	48,697	6,057.2	6,025.0	6,713.2	7,102.8	7,785.8
65 Tb	44,481.6	43,744.1	50,382	6,272.8	6,238.0	6,978	7,366.7	8,102
66 Dy	45,998.4	45,207.8	52,119	6,495.2	6,457.7	7,247.7	7,635.7	8,418.8
67 Ho	47,546.7	46,699.7	53,877	6,719.8	6,679.5	7,525.3	7,911	8,747
68 Er	49,127.7	48,221.1	55,681	6,948.7	6,905.0	7,810.9	8,189.0	9,089
69 Tm	50,741.6	49,772.6	57,517	7,179.9	7,133.1	8,101	8,468	9,426
70 Yb	52,388.9	51,354.0	5,937	7,415.6	7,367.3	8,401.8	8,758.8	9,780.1
71 Lu	54,069.8	52,965.0	61,283	7,655.5	7,604.9	8,709.0	9,048.9	10,143.4
72 Hf	55,790.2	54,611.4	63,234	7,899.0	7,844.6	9,022.7	9,347.3	10,515.8
73 Ta	57,532	56,277	65,223	8,146.1	8,087.9	9,343.1	9,651.8	10,895.2
74 W	59,318.24	57,981.7	67,244.3	8,397.6	8,335.2	9,672.35	9,961.5	11,285.9
75 Re	61,140.3	59,717.9	69,310	8,652.5	8,586.2	10,010.0	10,275.2	11,685.4
76 Os	63,000.5	61,486.7	71,413	8,911.7	8,841.0	10,355.3	10,598.5	12,095.3
77 Ir	64,895.6	63,286.7	73,560.8	9,175.1	9,099.5	10,708.3	10,920.3	12,512.6
78 Pt	66,832	65,112	75,748	9,442.3	9,361.8	11,070.7	11,250.5	12,942.0
79 Au	68,803.7	66,989.5	77,984	9,713.3	9,628.0	11,442.3	11,584.7	13,381.7
80 Hg	70,819	68,895	80,253	9,988.8	9,897.6	11,822.6	11,924.1	13,830.1
81 Tl	72,871.5	70,831.9	82,576	10,268.5	10,172.8	12,213.3	12,271.5	14,291.5
82 Pb	74,969.4	72,804.2	84,936	10,551.5	10,449.5	12,613.7	12,622.6	14,764.4
83 Bi	77,107.9	74,814.8	87,343	10,838.8	10,730.91	13,023.5	12,979.9	15,247.7
84 Po	79,290	76,862	8,980	11,130.8	11,015.8	13,447	13,340.4	15,744
85 At	8,152	7,895	9,230	11,426.8	11,304.8	13,876	—	16,251
86 Rn	8,378	8,107	9,487	11,727.0	11,597.9	14,316	—	16,770
87 Fr	8,610	8,323	9,747	12,031.3	11,895.0	14,770	1,445	17,303
88 Ra	8,847	8,543	10,013	12,339.7	12,196.2	15,235.8	14,841.4	17,849
89 Ac	90,884	8,767	10,285	12,652.0	12,500.8	15,713	—	18,408
90 Th	93,350	89,953	105,609	12,968.7	12,809.6	16,202.2	15,623.7	18,982.5
91 Pa	95,868	92,287	108,427	13,290.7	13,122.2	16,702	16,024	19,568
92 U	98,439	94,665	111,300	13,614.7	13,438.8	17,220.0	16,428.3	20,167.1
93 Np	—	—	—	13,944.1	13,759.7	17,750.2	16,840.0	20,784.8
94 Pu	—	—	—	14,278.6	14,084.2	18,293.7	17,255.3	21,417.3
95 Am	—	—	—	14,617.2	14,411.9	18,852.0	17,676.5	22,065.2

*Photon energies in electronvolts (eV) of some characteristic emission lines of the elements of atomic number $3 \leq Z \leq 95$, as compiled by J. Kortright. "Characteristic X-Ray Energies," in *X-Ray Data Booklet* (Lawrence Berkeley National Laboratory Pub-490, Rev. 2, 1999). Values are largely based on those given by J. A. Bearden, "X-Ray Wavelengths," *Rev. Mod. Phys.* **39**:78 (1967), which should be consulted for a more complete listing. Updates may also be noted at the Web site www.cxro.lbl.gov/optical_constants.

TABLE 3 Curves Showing Auger Energies,* in Electronvolts (eV), for Elements of Atomic Number $3 \leq Z \leq 92$ 

*Only dominant energies are given, and only for principal Auger peaks. The literature should be consulted for detailed tabulations, and for shifted values in various common compounds.²⁻⁴ (Courtesy of Physical Electronics, Inc.²)

36.3 REFERENCES

1. B. L. Henke, E. M. Gullikson, and J. C. Davis. "X-Ray Interactions: Photoabsorption, Scattering, Transmission, and Reflection at $E = 50\text{--}3000$ eV, $Z = 1\text{--}92$." *Atomic Data and Nuclear Data Tables* **54**(2):181–342 (July 1993).
2. K. D. Childs, B. A. Carlson, L. A. Vanier, J. F. Moulder, D. F. Paul, W. F. Stickle, and D. G. Watson. *Handbook of Auger Electron Spectroscopy*; C. L. Hedberg (ed.), Physical Electronics, Eden Prairie, MN, 1995.
3. J. F. Moulder, W. F. Stickle, P. E. Sobol, and K. D. Bomben, *Handbook of X-Ray Photoelectron Spectroscopy*, Physical Electronics, Eden Prairie, MN 1995.
4. D. Briggs, *Handbook of X-Ray and Ultraviolet Photoelectron Spectroscopy*, Heyden, London, 1977.

SUBPART

5.2

REFRACTIVE AND
INTERFERENCE OPTICS

This page intentionally left blank

REFRACTIVE X-RAY LENSES

Bruno Lengeler

*Physikalisches Institut
RWTH Aachen University
Aachen, Germany*

Christian G. Schroer

*Institute of Structural Physics
TU Dresden
Dresden, Germany*

37.1 INTRODUCTION

The last ten years have seen a remarkable progress in the development of new x-ray optics and in the improvement of existing devices. In this chapter, we describe the properties of one type of these new optics: refractive x-ray lenses.

For a long time these lenses were considered as not feasible, due to the weak refraction and the relatively strong absorption of x rays in matter. However, in 1996 it was shown experimentally that focusing by x-ray lenses is possible if the radius of curvature R of an individual lens is chosen to be small (e.g., below 0.5 mm, cf. Fig. 1a), if many such lenses are stacked behind one another in a row, and if a lens material with low atomic number Z , such as aluminium, is chosen.^{1,2} The first lenses of this type consisted of a row of holes, 1 mm in diameter, drilled in a block of aluminium.

In the meantime, many different types of refractive lenses made of various materials have been developed.^{3–19} One of the most important developments was to make these optics aspherical,⁴ reducing spherical aberration to a minimum and thus making these optics available for high-resolution x-ray microscopy. As each individual lens is thin in the optical sense, the ideal aspherical shape is a paraboloid of rotation. In the following, we focus on two types of high resolution x-ray optics, rotationally parabolic lenses made of beryllium and aluminium^{6,7,13,14,16,20} and cylindrically parabolic lenses with particularly short focal length.^{21,22} An example of the first type of lenses developed and made at Aachen University is described in Secs. 37.2 to 37.5. These lenses allow x-ray imaging nearly free of distortions and can be used as an objective lens in an x-ray microscope for efficient focusing in scanning microscopy, and for a variety of beam conditioning applications at third generation synchrotron radiation sources. The second type of lenses, so-called nanofocusing lenses (NFLs), have a short focal distance and large numerical aperture and are thus particularly suited to focus hard x rays to sub-100 nm dimensions for scanning microscopy applications. While focusing of hard x rays down to 50 nm has been demonstrated experimentally, these optics have the potential of focusing hard x rays to about 10 nm^{21,22} and perhaps below.²³ The development of nanofocusing lenses currently pursued at TU Dresden is described in Sec. 37.6.

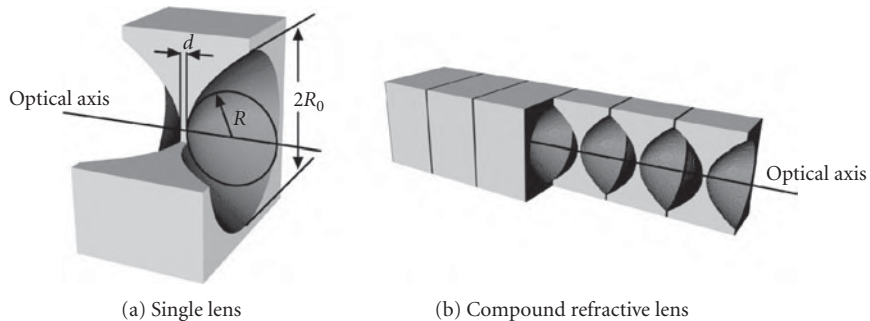


FIGURE 1 (a) Individual refractive x-ray lens with rotationally parabolic profile and (b) stack of individual lenses forming a refractive x-ray lens. (Reused with permission from Ref. 24.)

37.2 REFRACTIVE X-RAY LENSES WITH ROTATIONALLY PARABOLIC PROFILE

Refractive x-ray lenses with rotationally parabolic profiles^{6,7,13,16} allow for focusing in both directions, free of spherical aberration and other distortions. Aluminium and beryllium are the lens materials most commonly used. Beryllium is especially suitable for x-ray energies between 7 and 40 keV due to its low attenuation of x rays (low atomic number $Z = 4$). Between about 40 and 90 keV aluminium ($Z = 13$) is more appropriate as a lens material. Having been able to solve the problems with handling and plastically deforming beryllium, Be lenses can be manufactured with rotationally parabolic profiles.^{13,14} Figure 1a shows a schematic drawing of an individual lens. Note the concave shape of a focusing lens, which is a result of the diffractive part $1 - \delta$ of the index of refraction n being smaller than 1 in the x-ray range. In Fig. 1b a number N of individual lenses is stacked behind each other to form a refractive x-ray lens. Figure 2 shows a stack of Be lenses in their casing with protective atmosphere.

In the thin lens approximation, the focal length of a stack of N lenses is $f_0 = R/2N\delta$. Here, R is the radius of curvature at the apex of the paraboloid (cf. Fig. 1a). For paraboloids, R and the geometric aperture $2R_0$ are independent of one another, in contrast to the case for spherical lenses. Most lenses up to now had the parameters $R = 0.2$ mm and $2R_0 \approx 1$ mm. Up to several hundred lenses can be aligned

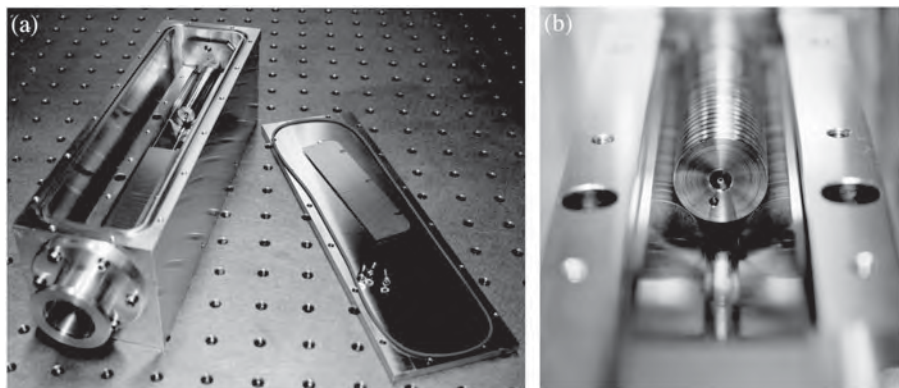


FIGURE 2 (a) Housing with partly assembled Be lens and (b) stack of Be lenses. Each individual lens is centered inside of a hard metal coin. The lenses are aligned along the optical axis by stacking the coins in a high precision v-groove. (See also color insert.)

in a lens stack in such a way that the optical axes of the individual lenses agree on the micrometer scale. The form fidelity of the paraboloids is better than $0.3 \mu\text{m}$ and surface roughness is below $0.1 \mu\text{m}$.

In the meantime, lenses with different radii of curvature R ($R = 50, 100, 200, 300, 500, 1000$, and $1500 \mu\text{m}$) have been developed to optimize the optics for various applications. They are available in three lens materials, i.e., beryllium, aluminium, and nickel. Lenses with small radii R are especially suited for microscopy applications with high lateral resolution, whereas those with a large radius R are designed for beam conditioning purposes, such as prefocusing and collimation.

In general, the total length L of a lens stack is not negligible compared to the focal length f . Then, a correction has to be applied to the thin lens approximation for the focal length. For a thick lens the focal length is

$$f = f_0 \sqrt{\frac{L}{f_0}} \frac{1}{\sin \sqrt{L/f_0}} \quad (1)$$

as measured from the principal planes located at

$$H_{1,2} = \pm \left[f_0 \sqrt{\frac{L}{f_0}} \frac{1 - \cos \sqrt{\frac{L}{f_0}}}{\sin \sqrt{\frac{L}{f_0}}} - \frac{L}{2} \right] \quad (2)$$

behind and before the center of the lens, respectively.

The attenuation of x rays in matter is a key parameter in the design of refractive x-ray lenses. As the thickness of the lens material increases with increasing distance from the optical axis, the lens becomes more and more absorbing toward its periphery. Thus a refractive x-ray lens has no sharp aperture, but a Gaussian transmission that is responsible for the diffraction at the lens. As a result, we can assign an effective aperture D_{eff} to the lens that is smaller than the geometric aperture $2R_0^{6,7}$ and that determines the diffraction at the lens, its numerical aperture, and the achievable diffraction-limited spot size.

At low energies (below about 10 keV for beryllium), the attenuation is dominated by photoabsorption. The mass photoabsorption coefficient τ/ρ varies approximately like Z^3/E^3 with atomic number Z and with photon energy E . When τ/ρ drops to below about $0.15 \text{ cm}^2/\text{g}$ at higher x-ray energies, the mass attenuation coefficient $\mu = \tau + \mu_c$ is dominated by Compton scattering (μ_c) and stays more or less constant, independent of energy and atomic number Z . For beryllium the cross-over between the photoabsorption and Compton scattering dominated attenuation is at about 17 keV. The performance of beryllium lenses is optimal in this energy range.

Compton scattering ultimately limits the performance (lateral resolution) of refractive x-ray lenses. Compton scattering has a twofold detrimental influence. Photons which are Compton scattered no longer contribute to the image formation. In addition, they generate a background which reduces the signal-to-background-ratio in the image.

Synchrotron radiation sources of the third generation can create a considerable heat load in the first optical element hit by the beam. This is expected to be even more true at x-ray free-electron laser sources which are being developed at present. The compatibility with such a high heat load was tested for refractive lenses made of beryllium at the undulator beamline ID10 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The power density and power of the white beam generated by 3 undulators in a row was about $100 \text{ W}/\text{mm}^2$ and 40 W, respectively. A stack of 12 Be lenses was exposed to the beam. The lenses were housed in an evacuated casing. They were indirectly cooled via a thermal link to a copper plate which in turn was water cooled. The temperature was measured by three thermocouples, one at each end and one at the center of the lens stack. The highest temperature was measured at the center, increasing within a few minutes to 65°C and staying constant afterward, except for small variations due to changes of the electron current in the ring. A temperature of 65°C poses no problem for Be lenses. The melting point of Be is at 1285°C and recrystallization of Be occurs only above 600°C . At present, rotationally parabolic Be lenses have been installed in the front ends of several undulator beamlines at ESRF, being routinely used in the undulator “white” beam. At

the present undulator beamlines no deterioration of Be or Al x-ray lenses has been observed in the monochromatic beam, even after many years of operation. In terms of stability metallic lens materials are far superior to insulators, like plastics or glass. The high density of free electrons in metals prevents radiation damage by bond breaking or local charging.

The heat load resistance of these optics is of utmost importance for focusing applications at future x-ray free-electron lasers. Model calculations suggest that these optics are stable in the hard x-ray beam (8 to 12 keV) generated by x-ray free-electron lasers, such as the LCLS in Stanford and the future European X-Ray Free-Electron Laser Project XFEL in Hamburg.^{24–27}

37.3 IMAGING WITH PARABOLIC REFRACTIVE X-RAY LENSES

Refractive x-ray lenses with parabolic profile are especially suited for hard x-ray full-field microscopy since they are relatively free of distortions compared to crossed lenses with cylindrical symmetry. For this purpose, a refractive x-ray lens is placed a distance L_1 behind the object that is illuminated from behind by monochromatic synchrotron radiation. The image of the object is formed at a distance $L_2 = L_1 f / (L_1 - f)$ behind the lens on a position-sensitive detector. To achieve large magnifications $M = L_2 / L_1 = f / (L_1 - f)$ up to 100, L_1 should be chosen to be slightly larger than the focal distance f . Figure 3 shows the image of a Ni mesh (periodicity 12.7 μm) imaged with a Be lens ($N = 91$, $f = 493$ mm, $L_2 / L_1 = 10$) at 12 keV onto high resolution x-ray film. Details of the contrast formation are described in Refs. 28 and 29. There are no apparent distortions visible in the image. This is a consequence of the parabolic lens profile. Figure 4 compares imaging with parabolic and spherical lenses in a numerical simulation. Spherical aberration dominates the image formed by the spherical lens, clearly demonstrating the need for a lens surface in the form of a paraboloid of rotation. A comparison of Fig. 3 with Fig. 4a shows that the experimental result is very close to a numerical calculation with idealized parabolic lenses.

The first x-ray microscope of this kind was built using aluminium lenses.⁶ However, using beryllium as a lens material rather than aluminium has several advantages. The reduced attenuation inside the lens results in a larger effective aperture that leads to a higher spatial resolution and a larger field of view. In addition, the efficiency of the setup is improved, since the transmission of the lens is higher.

For a refractive lens with a parabolic profile the lateral resolution is limited by the diffraction at its Gaussian aperture, giving rise to a Gaussian shape of the Airy disc.⁷ The full width at half maximum of the Airy disc is given by

$$d_t = 0.75 \frac{\lambda}{2NA} = 0.75 \frac{\lambda L_1}{D_{\text{eff}}} \quad (3)$$

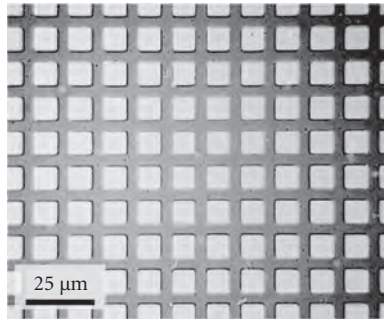


FIGURE 3 Hard x-ray micrograph of a Ni mesh.¹⁴ (Reused with permission from Ref. 20.)

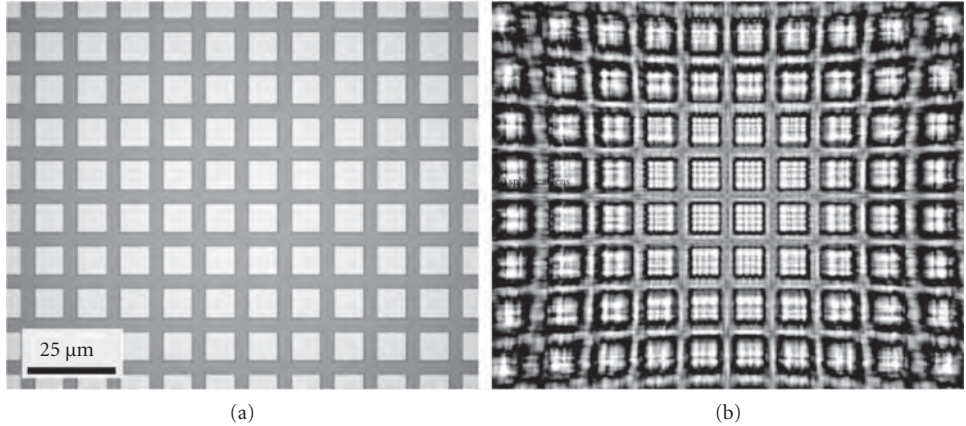


FIGURE 4 Numerical simulation of the imaging process using (a) parabolic and (b) spherical lens.

The numerical aperture NA is defined by $\sin \alpha$, where 2α is the angle spanned by the effective aperture D_{eff} of the lens as seen from an object point.⁷ This result is well known from optics, the factor 0.75 being different from the usual factor 1.22. The difference can be traced back to the fact that in normal optics apertures are sharply delimited whereas for x-ray lenses the attenuation changes smoothly as described earlier. The effective aperture D_{eff} is limited by x-ray attenuation and that is ultimately limited by Compton scattering. An estimate of d_t for lenses with large apertures shows that it scales with

$$d_t = a\lambda \sqrt{\frac{\mu f}{\delta}}$$

where a is a factor of order one. This implies that a low value for d_t needs a small focal length f and a low mass attenuation coefficient μ , in other words a low Z material. Since δ is proportional to λ^2 the main x-ray energy dependence enters via μ . With the present day technology for fabrication of refractive lenses with rotationally parabolic profile, focal lengths between 10 and 20 cm can be achieved for energies between 10 and 20 keV, resulting for Be lenses in a lateral resolution down to about 50 nm. We estimate that it will be difficult to reach values below 30 nm.

The main strength of the x-ray microscope is the large penetration depth of hard x rays in matter that allows one to investigate non-destructively inner structures of an object. In combination with tomographic techniques, it allows one to reconstruct the three-dimensional inner structure of the object with submicrometer resolution.³⁰ In addition, full-field imaging in demagnifying geometry can be used for hard x-ray lithography.³¹ The high quality of refractive lenses is also demonstrated by the preservation of the lateral coherence.^{16,32}

37.4 MICROFOCUSING WITH PARABOLIC REFRACTIVE X-RAY LENSES

Refractive x-ray lenses with parabolic profiles can also be used for generating a (sub-)micrometer focal spot for x-ray microanalysis and tomography. For that purpose, the synchrotron radiation source is imaged by the lens onto the sample in a strongly demagnifying way, i.e., the source-lens distance L_1 is chosen to be much larger than the lens-sample distance L_2 . At a synchrotron radiation source the horizontal source size is typically larger than the vertical one. As the lens images this horizontally elongated source to the sample position, the focal spot is larger in the horizontal direction than in the vertical direction. With Be lenses ($R = 200 \mu\text{m}$), a vertical spot size well below $1 \mu\text{m}$ is

routinely achieved, while the horizontal spot size is typically limited to a few micrometers by the horizontal source size and the demagnification of the setup. The diffraction limit of these optics is usually not reached in typical microfocusing geometries, i.e., 40 to 70 m from a typical undulator source at a synchrotron radiation source of the third generation. The spot size is dominated by the geometric image of the source and diffraction at the lens aperture and aberrations are negligible.

By means of new Be lenses with smaller radii of curvature, e.g., $R = 50\ \mu\text{m}$, a focal length of 15 cm can be reached, thus resulting in a demagnification of the source size by about a factor 400 in 60-m distance from the source. At a low- β undulator source, this demagnification allows one to reach the sub-micrometer regime also in the horizontal direction. Close to diffraction-limited focusing, however, is still only possible at very long beamlines. For example, at a distance of $L_1 = 145\ \text{m}$ from a low- β source at the ESRF [effective source size $60 \times 125\ \mu\text{m}$ ($V \times H$)], a microbeam of $60 \times 125\ \text{nm}^2$ is expected, approaching the diffraction limit of these optics in the vertical direction. To generate foci well below 100 nm at short distances from the source, focal distances in the centimeter range are needed to generate large demagnifications. This can be achieved with the nanofocusing lenses described in Sec. 37.6.^{21,22}

Hard x-ray microbeams find a large number of applications in scanning microscopy and have been used for a variety of experiments. They include, for example, microdiffraction,³³ microfluorescence mapping³⁴ and tomography,^{35–37} and x-ray absorption spectroscopic³⁸ and small-angle x-ray scattering tomography.³⁹ In materials science there is a great interest in using very hard x rays above about 80 keV, because many samples with high Z metallic components and thicknesses of many millimeters show strong x-ray absorption. At 80 keV parabolic aluminium lenses (preferably with $R = 50\ \mu\text{m}$) are well suited.⁴⁰ For higher x-ray energies, lenses made of nickel become advantageous, due to the strong refraction in nickel resulting from its relatively high density ($\rho_{\text{Ni}} = 8.9\ \text{g/cm}^3$). For energies above 80 keV, parabolic cylinder lenses made with a LIGA technique have been successfully tested.⁴¹ Rotationally parabolic nickel lenses are in the process of development. The challenge is to produce lenses with minimal thickness d (cf. Fig. 1a) in order to minimize absorption in the stack of lenses. A value of $d = 10\ \mu\text{m}$ is tolerable and feasible.

Be refractive lenses appear to be well suited to focus the beam from an x-ray free-electron-laser.²⁴

37.5 PREFOCUSING AND COLLIMATION WITH PARABOLIC REFRACTIVE X-RAY LENSES

Most experimental setups on synchrotron radiation beamlines are located between 40 and 150 m from the source. Depending on the experiment, the beam size, flux, divergence, or lateral coherence length may not be optimal at the position of the experiment. Using appropriate lenses upstream of the experiment, a given parameter can be optimized.

For example, the divergence of the beam may lead to a significant reduction of the flux at the sample position, in particular at a low- β undulator source. In this case, the beam can be moderately focused with refractive lenses with large radius of curvature and thus large geometric and effective aperture. For instance, lenses with $R = 1500\ \mu\text{m}$ have a geometric aperture $2R_0$ of 3 mm. In a one-to-one imaging geometry at 50 m from the source they have an angular acceptance of $85\ \mu\text{rad}$ at 17 keV, thus capturing a large fraction of the beam in horizontal direction. This opens excellent possibilities to increase the photon flux, in particular as the lenses can be easily moved in and out of the beam without affecting the optical axis and the alignment of the experiment. In the new future, parabolic cylinder lenses made of Be and Al will also become available. With a height of 3.5 mm and radii of curvature between 200 and $1500\ \mu\text{m}$, they can be used for one-dimensional focusing and collimation.

37.6 NANOFOCUSING REFRACTIVE X-RAY LENSES

High quality magnified imaging with x rays requires optical components free of distortion, like rotationally parabolic refractive x-ray lenses. However, for technical reasons, their radii of curvature cannot be made smaller than about $50\ \mu\text{m}$. This limits the focal distance from below and thus the achievable

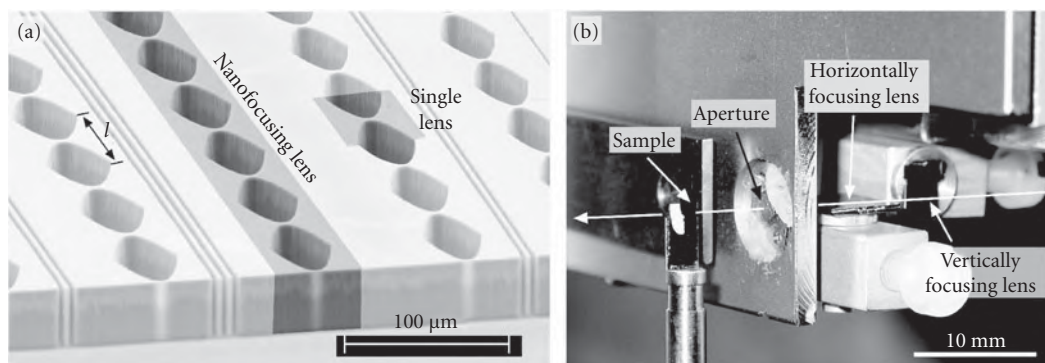


FIGURE 5 (a) Array of nanofocusing lenses made of silicon. A large number of single lenses are aligned behind each other to form a nanofocusing lens. Several nanofocusing lenses with different radius of curvature R are placed in parallel onto the same substrate. (b) Scanning microprobe setup with two crossed nanofocusing lenses. An aperture defining pinhole is placed behind the second lens. (See also color insert.)

demagnification in microfocus experiments. Therefore, another approach has been pursued for the generation of particularly small focal spots. These are nanofocusing cylinder lenses with parabolic profile and a focal length of the order of 1 cm.^{21,22} This is achieved by choosing the radius of curvature of the parabolas as small as 1 to 5 μm. Figure 5a shows an array of nanofocusing lenses made of silicon. When two lenses are used in crossed geometry as shown in Fig. 5b, two-dimensional focusing can be achieved. So far, focal spot sizes down to $47 \times 55 \text{ nm}^2$ (H $\times$ V) have been reached at $L_1 = 47 \text{ m}$ from the low- β undulator source at beamline ID13 of the ESRF ($E = 21 \text{ keV}$).²²

While this spot size is close to the ideal performance of silicon lenses in this particular imaging geometry, significant improvements can be made in the future by further optimization of the optics and the imaging geometry. Figure 6 shows the optimal diffraction limit as a function of x-ray energy for different lens materials. Over a wide range of energies (from $E = 8 \text{ keV}$ to over $E = 100 \text{ keV}$, a diffraction

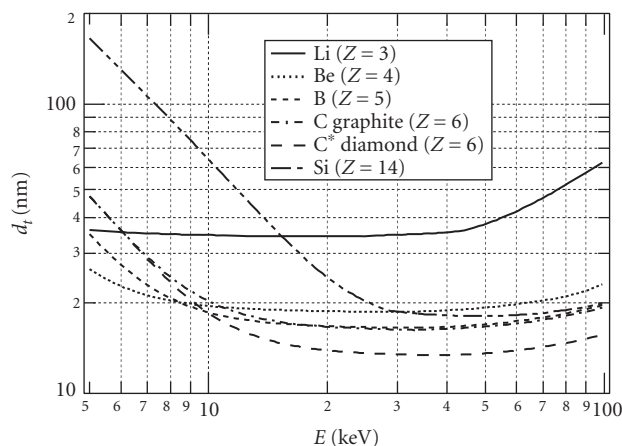


FIGURE 6 Minimal diffraction limits of nanofocusing lenses made of different lens materials and having a working distance of 1 mm. The radius of curvature R and the length of a single lens l were varied within a range accessible by modern microfabrication techniques.²¹ (Copyright 2003 by the American Institute of Physics. Reused with permission from Ref. 21.)

limit below 20 nm is expected. Best performance is obtained for low Z materials with high density. The reason for this is that attenuation is no longer limiting the aperture of nanofocusing lenses for low Z materials, as the overall length of the lens is short. For a given focal length, the geometric aperture is, however, limited by the refractive strength per unit length inside the lens. The higher δ , the larger can be the radius of curvature R and thus the geometric aperture $R_0 = \sqrt{R(l-d)}$, if the thickness l of an individual lens is kept constant (cf. Fig. 5a). In the limit, the numerical aperture approaches $\sqrt{2\delta}$ that coincides with the critical angle of total reflection. At highly brilliant sources, such as the ESRF, these diffraction limits are expected to be reached with fluxes above 10^9 photons per second.

For these optics, prefocusing as described in Sec. 37.5 is of utmost importance to obtain optimal performance. Optimal diffraction-limited focusing is obtained when the lateral coherence length at the optic is slightly larger than the effective aperture. This requirement is usually not fulfilled at the position of the experiment. By appropriate prefocusing, the lateral coherence length can be adapted to the aperture of the NFL, thus optimally focusing the coherent flux from the source onto the sample. This scheme is pursued in modern hard x-ray scanning microscopes, both at ESRF and at the future synchrotron radiation source PETRA III at DESY in Hamburg, Germany.

For refractive lenses made of identical single lenses, the numerical aperture is fundamentally limited by the critical angle of total reflection $\sqrt{2\delta}$. This limitation can be overcome with refractive optics by gradually (adiabatically) adjusting the aperture of the individual lenses to the converging beam inside the optic. For these so-called adiabatically focusing lenses, the numerical aperture can exceed $\sqrt{2\delta}$ leading to diffraction limits well below 10 nm.²³

The main applications of nanofocusing lenses lie in scanning microscopy and microanalysis with hard x rays. They allow one to perform x-ray analytical techniques, such as diffraction,⁴² fluorescence analysis, and absorption spectroscopy, with high spatial resolution. Also, coherent x-ray diffraction imaging greatly benefits from focusing the coherent beam with NFLs.⁴³ The current performance of a hard x-ray scanning microscope based on nanofocusing lenses is illustrated in Fig. 7. In collaboration with W. H. Schröder from the Research Center Jülich the distribution of physiologically relevant ions and heavy metals was mapped inside the tip of a leaf hair (trichome) of the model plant *Arabidopsis thaliana*. Figure 7c shows the two-dimensional map of a variety of elements obtained by scanning the

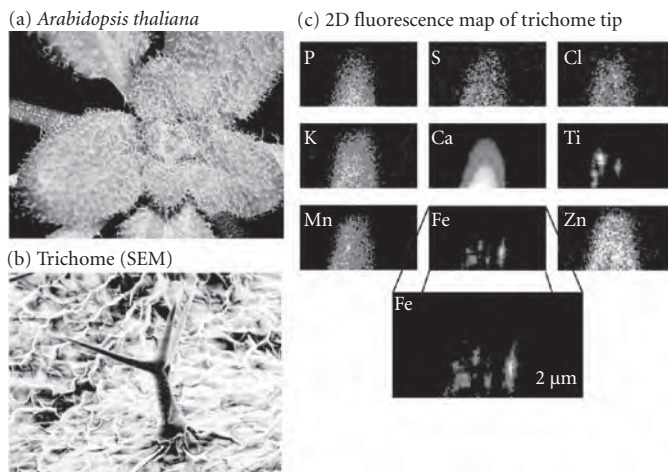


FIGURE 7 (a) Photograph of the plant *Arabidopsis thaliana*, (b) secondary electron micrograph of a leaf hair (trichome), and (c) two-dimensional fluorescence map of the tip of the trichome at 100-nm spatial resolution. While most elements are homogeneously distributed, iron (Fe) and titanium (Ti) are localized on the level of 100 nm. (See also color insert.) (Sample provided by W. H. Schröder, Research Center Jülich.)

tip of a trichome with a hard x-ray nanobeam ($E = 15$ keV). The step size was 100 nm in both dimensions, clearly showing a strong localization of iron and titanium. While the reason for this localization remains unknown, it impressively demonstrates the high spatial resolution obtained with nanofocusing lenses. While these optics are ideal for microbeam applications, they are not well suited for high quality full-field imaging due to distortions in the image due to the crossing to two cylinder lenses with different focal lengths.

37.7 CONCLUSION

Since their first experimental realization about one decade ago, refractive x-ray lenses have developed into a high-quality x-ray optic. Similar to glass lenses for visible light, they have a broad range of applications and can be used in very much the same way. Due to their good imaging properties, refractive optics are particularly suited for hard x-ray microscopy and microanalysis. Due to the weak refraction of hard x rays in matter, they are generally slim, operating in the paraxial regime with typical numerical apertures below a few times 10^{-3} . They can be used in the whole hard x ray range from about five to several hundred keV. As the refractive index depends on energy, refractive lenses are chromatic. Thus, they are mostly used with monochromatic radiation. Today, spatial resolutions down to 50 nm have been reached in hard x-ray microscopy. Potentially, these optics can generate hard x-ray beams down to below 10 nm. Their straight optical path makes them easy to use and align and enhances the stability of x-ray microscopes, as angular instabilities do not affect the focus. In addition, they are extremely robust, both mechanically and thermally. Therefore, they can be used as front end optics at third-generation synchrotron radiation sources and are good candidates to focus the radiation from free-electron lasers. Today, refractive x-ray lenses are routinely used at many beamlines of different synchrotron radiation sources.

37.8 REFERENCES

1. A. Snigirev, V. Kohn, I. Snigireva, and B. Lengeler, "A Compound Refractive Lens for Focusing High Energy X-Rays," *Nature (London)* **384**:49 (1996).
2. B. Lengeler, J. Tümmeler, A. Snigirev, I. Snigireva, and C. Raven, "Transmission and Gain of Singly and Doubly Focusing Refractive X-Ray Lenses," *J. Appl. Phys.* **84**:5855–5861 (1998).
3. A. Snigirev, B. Filseth, P. Elleaume, T. Klocke, V. Kohn, B. Lengeler, I. Snigireva, A. Souvorov, and J. Tümmeler, "Refractive Lenses for High Energy X-Ray Focusing," In A. M. K. A. T. Macrander, ed., "High Heat Flux and Synchrotron Radiation Beamlines," *Proc. SPIE*, **3151**:164–170 (1997).
4. B. Lengeler, "Linsensysteme für Röntgenstrahlen," *Spektrum der Wissenschaft* 25–30 (1997).
5. A. Snigirev, V. Kohn, I. Snigireva, A. Souvorov, and B. Lengeler, "Focusing High-Energy X-Rays by Compound Refractive Lenses," *Appl. Opt.* **37**:653–662 (1998).
6. B. Lengeler, C. G. Schroer, M. Richwin, J. Tümmeler, M. Drakopoulos, A. Snigirev, and I. Snigireva, "A Microscope for Hard X-Rays Based on Parabolic Compound Refractive Lenses," *Appl. Phys. Lett.* **74**:3924–3926 (1999).
7. B. Lengeler, C. Schroer, J. Tümmeler, B. Benner, M. Richwin, A. Snigirev, I. Snigireva, and M. Drakopoulos, "Imaging by Parabolic Refractive Lenses in the Hard X-Ray Range," *J. Synchrotron Rad.* **6**:1153–1167 (1999).
8. Y. Kohmura, M. Awaji, Y. Suzuki, T. Ishikawa, Y. I. Dudchik, N. N. Kolchewsky, and F. F. Komarow, "X-Ray Focusing Test and X-Ray Imaging Test by a Microcapillary X-Ray Lens at an Undulator Beamline," *Rev. Sci. Instrum.* **70**:4161–4167 (1999).
9. J. T. Cremer, M. A. Piestrup, H. R. Beguiristain, C. K. Gary, R. H. Pantell, and R. Tatchyn, "Cylindrical Compound Refractive X-Ray Lenses Using Plastic Substrates," *Rev. Sci. Instrum.* **70** (1999).
10. B. Cederström, R. N. Cahn, M. Danielsson, M. Lundqvist, and D. R. Nygren, "Focusing Hard X-Rays with Old LP's," *Nature* **404**:951 (2000).
11. V. Aristov, M. Grigoriev, S. Kuznetsov, et al., "X-Ray Refractive Planar Lens with Minimized Absorption," *Appl. Phys. Lett.* **77**:4058–4060 (2000).

12. E. M. Dufresne, D. A. Arms, R. Clarke, N. R. Pereira, S. B. Dierker, and D. Foster, "Lithium Metal for X-Ray Refractive Optics," *Appl. Phys. Lett.* **79**:4085–4087 (2001).
13. B. Lengeler, C. G. Schroer, B. Benner, A. Gerhardus, T. F. Günzler, M. Kuhlmann, J. Meyer, and C. Zimprich, "Parabolic Refractive X-Ray Lenses," *J. Synchrotron Rad.* **9**:119–124 (2002).
14. C. G. Schroer, M. Kuhlmann, B. Lengeler, T. F. Günzler, O. Kurapova, B. Benner, C. Rau, A. S. Simionovici, A. Snigirev, and I. Snigireva, "Beryllium Parabolic Refractive X-Ray Lenses," In D. C. Mancini, ed., "Design and Microfabrication of Novel X-Ray Optics," *Proc. SPIE*, **4783**:10–18, SPIE, Bellingham (2002).
15. B. Cederström, M. Lundqvist, and C. Ribbing, "Multi-Prism X-Ray Lens," *Appl. Phys. Lett.* **81**:1399–1401 (2002).
16. B. Lengeler, C. G. Schroer, M. Kuhlmann, B. Benner, T. F. Günzler, O. Kurapova, A. Somogyi, A. Snigirev, and I. Snigireva, "Beryllium Parabolic Refractive X-Ray Lenses," In T. Warwick, J. Arthur, H. A. Padmore, and J. Stöhr, eds., Synchrotron Radiation Instrumentation, *Proc. AIP Conference*, **705**:748–751 (2004).
17. B. Nöhammer, J. Hoszowska, A. K. Freund, and C. David, "Diamond Planar Refractive Lenses for Third- and Forth-Generation X-Ray Sources," *J. Synchrotron Rad.* **10**:168–171 (2003).
18. V. Nazmov, E. Reznikova, M. Boerner, et al., "Refractive Lenses Fabricated by Deep SR Lithography and LIGA Technology for X-Ray Energies from 1 keV to 1 MeV," In T. Warwick, J. Arthur, H. A. Padmore, and J. Stöhr, eds., Synchrotron Radiation Instrumentation, *Proc. AIP Conference*, **705**:752–755 (2004).
19. V. Nazmov, E. Reznikova, A. Somogyi, J. Mohr, and V. Saile, "Planar Sets of Cross X-Ray Refractive Lenses from SU-8 Polymer," In A. S. Snigirev and D. C. Mancini, eds., "Design and Microfabrication of Novel X-Ray Optics II," *Proc. SPIE*, **5539**:235–243 (2004).
20. B. Lengeler, C. G. Schroer, M. Kuhlmann, B. Benner, T. F. Günzler, O. Kurapova, F. Zontone, A. Snigirev, and I. Snigireva, "Refractive X-Ray Lenses," *J. Phys. D: Appl. Phys.* **38**:A218–A222 (2005).
21. C. G. Schroer, M. Kuhlmann, U. T. Hunger, et al., "Nanofocusing Parabolic Refractive X-Ray Lenses," *Appl. Phys. Lett.* **82**:1485–1487 (2003).
22. C. G. Schroer, O. Kurapova, J. Patommel, et al., "Hard X-Ray Nanoprobe Based on Refractive X-Ray Lenses," *Appl. Phys. Lett.* **87**:124103 (2005).
23. C. G. Schroer and B. Lengeler, "Focusing Hard X Rays to Nanometer Dimensions by Adiabatically Focusing Lenses," *Phys. Rev. Lett.* **94**:054802 (2005).
24. C. G. Schroer, J. Tümmler, B. Lengeler, M. Drakopoulos, A. Snigirev, and I. Snigireva, "Compound Refractive Lenses: High Quality Imaging Optics for the XFEL," In D. M. Mills, H. Schulte-Schrepping, and J. R. Arthur, eds., "X-Ray FEL Optics and Instrumentation," *Proceedings of the SPIE*, **4143**:60–68 (2001).
25. G. Materlik and T. Tschentscher, "TESLA Technical Design Report, Part V, The X-ray Free Electron Laser," *Tech. Rep.* DESY 2001-011, DESY, Hamburg (2001).
26. R. M. Bionta, "Controlling Dose to Low Z Solids at LCLS," *Tech. Rep.* Lawrence Livermore National Laboratory, UCRL-ID-137222, January 3, 2000, LCLS-TN-00-4 (2000).
27. C. G. Schroer, B. Benner, M. Kuhlmann, O. Kurapova, B. Lengeler, F. Zontone, A. Snigirev, I. Snigireva, and H. Schulte-Schrepping, "Focusing Hard X-Ray FEL Beams with Parabolic Refractive Lenses," In S. G. Biedron, W. Eberhardt, T. Ishikawa, and R. O. Tatchyn, eds., "Fourth Generation X-Ray Sources and Optics II," *Proceedings of the SPIE*, **5534**:116–124 (2004).
28. C. G. Schroer, B. Benner, T. F. Günzler, M. Kuhlmann, B. Lengeler, C. Rau, T. Weitkamp, A. Snigirev, and I. Snigireva, "Magnified Hard X-Ray Microtomography: Toward Tomography with Sub-Micron Resolution," In U. Bonse, ed., "Developments in X-Ray Tomography III," *Proceedings of the SPIE*, **4503**:23–33 (2002).
29. V. Kohn, I. Snigireva, and A. Snigirev, "Diffraction Theory of Imaging with X-Ray Compound Refractive Lens," *Opt. Commun.* **216**:247–260 (2003).
30. C. G. Schroer, J. Meyer, M. Kuhlmann, B. Benner, T. F. Günzler, B. Lengeler, C. Rau, T. Weitkamp, A. Snigirev, and I. Snigireva, "Nanotomography Based on Hard X-Ray Microscopy with Refractive Lenses," *Appl. Phys. Lett.* **81**:1527–1529 (2002).
31. C. G. Schroer, B. Benner, T. F. Günzler, et al., "High Resolution Imaging and Lithography With Hard X-Rays Using Parabolic Compound Refractive Lenses," *Rev. Sci. Instrum.* **73**:1640 (2002).
32. B. Lengeler, C. G. Schroer, M. Kuhlmann, B. Benner, T. F. Günzler, O. Kurapova, F. Zontone, A. Snigirev, and I. Snigireva, "Beryllium Parabolic Refractive X-Ray Lenses," In A. S. Snigirev and D. C. Mancini, eds., "Design and Microfabrication of Novel X-Ray Optics II," *Proceedings of the SPIE*, **5539**:1–9 (2004).

33. O. Castelnau, M. Drakopoulos, C. G. Schroer, I. Snigireva, A. Snigirev, and T. Ungar, "Dislocation Density Analysis in Single Grains of Steel by X-Ray Scanning Microdiffraction," *Nucl. Instrum. Methods A* **467–468**: 1245–1248 (2001).
34. S. Bohic, A. Simionovici, A. Snigirev, R. Ortega, G. Devès, D. Heymann, and C. G. Schroer, "Synchrotron Hard X-Ray Microprobe: Fluorescence Imaging of Single Cells," *Appl. Phys. Lett.* **78**:3544–3546 (2001).
35. A. S. Simionovici, M. Chukalina, C. Schroer, M. Drakopoulos, A. Snigirev, I. Snigireva, B. Lengeler, K. Janssens, and F. Adams, "High-Resolution X-Ray Fluorescence Microtomography of Homogeneous Samples," *IEEE Trans. Nucl. Sci.* **47**:2736–2740 (2000).
36. C. G. Schroer, J. Tümmler, T. F. Günzler, B. Lengeler, W. H. Schröder, A. J. Kuhn, A. S. Simionovici, A. Snigirev, and I. Snigireva, "Fluorescence Microtomography: External Mapping of Elements Inside Biological Samples," In F. P. Doty, H. B. Barber, H. Roehrig, and E. J. Morton, eds., "Penetrating Radiation Systems and Applications II," *Proceedings of the SPIE*, **4142**:287–296 (2000).
37. C. G. Schroer, "Reconstructing X-Ray Fluorescence Microtomograms," *Appl. Phys. Lett.* **79**:1912–1914 (2001).
38. C. G. Schroer, M. Kuhlmann, T. F. Günzler, et al., "Mapping the Chemical States of an Element Inside a Sample Using Tomographic X-Ray Absorption Spectroscopy," *Appl. Phys. Lett.* **82**:3360–3362 (2003).
39. C. G. Schroer, M. Kuhlmann, S. V. Roth, R. Gehrke, N. Stribeck, A. Almendarez-Camarillo, and B. Lengeler, "Mapping the Local Nanostructure Inside a Specimen by Tomographic Small Angle X-Ray Scattering," *Appl. Phys. Lett.* **88**:164102 (2006).
40. H. Reichert, V. Honkimaki, A. Snigirev, S. Engemann, and H. Dosch, "A New X-ray Transmission-Reflection Scheme for the Study of Deeply Buried Interfaces Using High Energy Microbeams," *Physica B* **336**:46–55 (2003).
41. V. Nazmov, E. Resnikova, A. Last, J. Mohr, V. Saile, R. Simon, and M. DiMichiel, "X-Ray Lenses Fabricated by LIGA Technology," In J. -Y. Choi and S. Rah, eds., "Synchrotron Radiation Instrumentation: Ninth International Conference on Synchrotron Radiation Instrumentation," *AIP Conference Proceedings*, **879**:770–773 (2007).
42. M. Hanke, M. Dubsclaff, M. Schmidbauer, T. Boeck, S. Schöder, M. Burghammer, C. Riekkel, J. Patommel, and C. G. Schroer, "Scanning X-Ray Diffraction with 200 nm Spatial Resolution," *Applied Physics Letters* **92**:193109 (2008).
43. C. G. Schroer, P. Boye, J. Feldkamp, J. Patommel, A. Schropp, A. Schwab, S. Stephan, M. Burghammer, S. Schoder, and C. Riekkel, "Coherent X-Ray Diffraction Imaging with Nanofocused Illumination," *Phys. Rev. Lett.* **101**:090801 (2008).

This page intentionally left blank

GRATINGS AND MONOCHROMATORS IN THE VUV AND SOFT X-RAY SPECTRAL REGION

Malcolm R. Howells

*Advanced Light Source
Lawrence Berkeley National Laboratory
Berkeley, California*

38.1 INTRODUCTION

Spectroscopy in the photon energy region from the visible to about 1 to 2 keV is generally done using reflection gratings. In the region above 40 eV, reasonable efficiency is only obtained at grazing angles and in this article we concentrate mainly on that case. Flat gratings were the first to be used and even today are still important. However, the advantages of spherical ones were recognized very early.¹ The first type of focusing grating to be analyzed theoretically was that formed by the intersection of a substrate surface with a set of parallel equispaced planes: the so-called “Rowland grating.” The theory of the spherical case was established first,^{1–3} and was described comprehensively in the 1945 paper of Beutler.⁴ Treatments of toroidal⁵ and ellipsoidal⁶ gratings came later, and the field has been reviewed by Welford,⁷ Samson,⁸ Hunter,⁹ and Namioka.¹⁰

The major developments in the last three decades have been in the use of nonuniformly spaced grooves. The application of holography to spectroscopic gratings was first reported by Rudolph and Schmahl^{11,12} and by Labeyrie and Flamand.¹³ Its unique opportunities for optical design were developed initially by Jobin-Yvon¹⁴ and by Namioka and coworkers.^{15,16} A different approach was followed by Harada¹⁷ and others, who developed the capability to produce gratings with variable-line spacing through the use of a computer-controlled ruling engine. The application of this class of gratings to spectroscopy has been developed still more recently, principally by Hettrick.¹⁸

In this chapter we will give a treatment of grating theory up to fourth order in the optical path, which is applicable to any substrate shape and any groove pattern that can be produced by holography or by ruling straight grooves with (possibly) variable spacing. The equivalent information is available up to sixth order at the website of the Center for X-Ray Optics at the Lawrence Berkeley National Laboratory.¹⁹

38.2 DIFFRACTION PROPERTIES

Notation and Sign Convention

We adopt the notation of Fig. 1 in which α and β have opposite signs if they are on opposite sides of the normal.

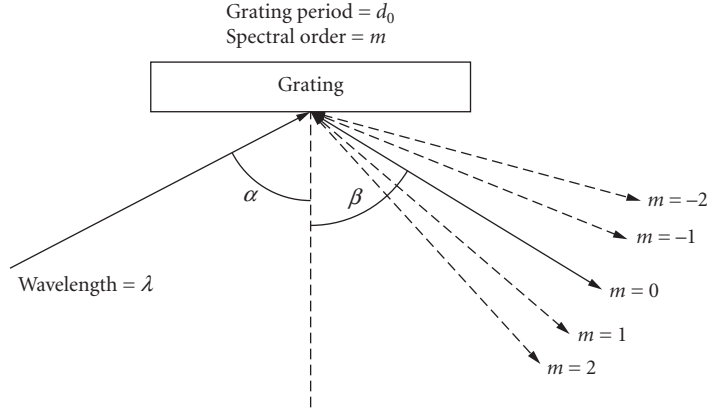


FIGURE 1 Grating equation notation.

Grating Equation

The grating equation may be written

$$m\lambda = d_0(\sin\alpha + \sin\beta) \quad (1)$$

The angles α and β are both arbitrary, so it is possible to impose various conditions relating them. If this is done, then for each λ , there will be a unique α and β . The following conditions are used:

1. *On-blaze condition:*

$$\alpha + \beta = 2\theta_B \quad (2)$$

where θ_B is the blaze angle (the angle of the sawtooth). The grating equation is then

$$m\lambda = 2d_0 \sin\theta_B \cos(\beta + \theta_B) \quad (3)$$

2. *Fixed in and out directions:*

$$\alpha - \beta = 2\theta \quad (4)$$

where 2θ is the (constant) included angle. The grating equation is then

$$m\lambda = 2d_0 \cos\theta \sin(\theta + \beta) \quad (5)$$

In this case, the wavelength scan ends when α or β reaches 90° , which occurs at the horizon wavelength $\lambda_H = 2d_0 \cos^2\theta$.

3. *Constant incidence angle:* Equation (1) gives β directly.
4. *Constant focal distance (of a plane grating):*

$$\frac{\cos\beta}{\cos\alpha} = \text{a constant } c_{ff} \quad (6)$$

leading to a grating equation

$$1 - \left(\frac{m\lambda}{d_0} - \sin \beta \right)^2 = \frac{\cos^2 \beta}{c_{ff}^2} \quad (7)$$

Equations (3), (5), and (7) give β (and thence α) for any λ . Examples where the above α - β relationships may be used are as follows:

1. Kunz et al. plane-grating monochromator (PGM),²⁰ Hunter et al. double PGM,²¹ collimated-light SX700.²²
2. Toroidal-grating monochromators (TGMs),^{23,24} spherical-grating monochromators (SGMs, also known as the *Dragon* system),²⁵ Seya-Namioka,^{26,27} most aberration-reduced holographic SGMs,²⁸ and certain PGMs.^{18,29,30} The variable-angle SGM³¹ follows Eq. (4) approximately.
3. Spectrographs, *Grasshopper* monochromator.³²
4. SX700 PGM³³ and variants.^{22,34}

38.3 FOCUSING PROPERTIES³⁵

Calculation of the Path Function F

Following normal practice, we provide an analysis of the imaging properties of gratings by means of the path function F .¹⁶ For this purpose we use the notation of Fig. 2, in which the zeroth groove (of width d_0) passes through the grating pole O , while the n th groove passes through the variable point $P(\xi, w, l)$.

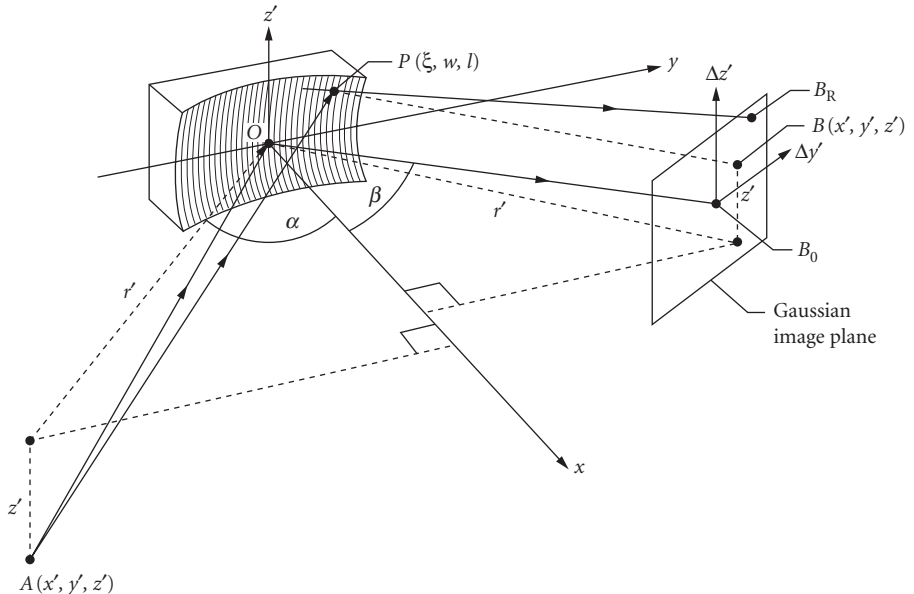


FIGURE 2 Focusing properties notation.

F is expressed as

$$F = \sum_{ijk} F_{ijk} w^i l^j$$

where

$$F_{ijk} = z^k C_{ijk}(\alpha, r) + z'^k C_{ijk}(\beta, r') + \frac{m\lambda}{d_0} f_{ijk} \quad (8)$$

and the f_{ijk} term, originating from the groove pattern, is given by one of the following expressions:

$$f_{ijk} = \begin{cases} 1 \text{ when } ijk = 100, 0 \text{ otherwise} & \text{Rowland} \\ \frac{d_0}{\lambda_0} [z_C^k C_{ijk}(\gamma, r_C) \pm z_D^k C_{ijk}(\delta, r_D)] & \text{holographic} \\ n_{ijk} & \text{varied line spacing} \end{cases} \quad (9)$$

The holographic groove pattern in Eq. (9) is assumed to be made using two coherent point sources C and D with cylindrical polar coordinates (r_C, γ, z_C) , (r_D, δ, z_D) relative to O . The lower (upper) sign refers to C and D , both real or both virtual (one real and one virtual), for which case the equiphase surfaces are confocal hyperboloids (ellipses) of revolution about CD . The grating with varied line spacing $d(w)$ is assumed to be ruled according to $d(w) = d_0(1 + v_1 w + v_2 w^2 + \dots)$. We consider all the gratings to be ruled on the general surface $x = \sum_{ij} a_{ij} w^i l^j$ and the a_{ij} coefficients³⁶ are given for the important substrate shapes in Tables 1 and 2.

TABLE 1 Ellipsoidal Mirror $a_{ij}'s^{*,36}$

$a_{20} = \frac{\cos \theta}{4} \left(\frac{1}{r} + \frac{1}{r'} \right)$	$a_{02} = \frac{a_{20}}{\cos^2 \theta}$	$a_{22} = \frac{a_{20}(2A^2 + C)}{2\cos^2 \theta}$
$a_{30} = a_{20}A$	$a_{12} = \frac{a_{20}A}{\cos^2 \theta}$	$a_{04} = \frac{a_{20}C}{8\cos^2 \theta}$
$a_{40} = \frac{a_{20}(4A^2 + C)}{4}$		
Other $a_{ij}'s$ with $i + j \leq 4$ are zero.		

^{*} r , r' and θ are the object distance, image distance, and incidence angle to the normal, respectively, and

$$A = \frac{\sin \theta}{2} \left(\frac{1}{r} - \frac{1}{r'} \right), \quad C = A^2 + \frac{1}{rr'}$$

The $a_{ij}'s$ for spheres; circular, parabolic, or hyperbolic cylinders; paraboloids; and hyperboloids can also be obtained from Tables 1 and 2 by suitable choices of the input parameters r , r' , and θ .

TABLE 2 Toroidal Mirror $a_{ij}'s^{*,36}$

$a_{20} = \frac{1}{2R}$	$a_{02} = \frac{1}{2\rho}$	$a_{04} = \frac{1}{8R^3}$
$a_{30} = \frac{1}{8R^3}$	$a_{22} = \frac{1}{4\rho R^2}$	
Other $a_{ij}'s$ with $i + j \leq 4$ are zero		

^{*} R and ρ are the major and minor radii of the bicycle-tire toroid.

TABLE 3 Coefficients C_{ijk} of the Expansion of F^* ,¹⁶

$C_{011} = -\frac{1}{r}$	$C_{020} = \frac{S}{2}$	
$C_{022} = -\frac{S}{4r^2} - \frac{1}{2r^3}$	$C_{031} = \frac{S}{2r^2}$	$C_{100} = -\sin \alpha$
$C_{040} = \frac{4a_{02}^2 - S^2}{8r} - a_{04} \cos \alpha$	$C_{200} = \frac{T}{2}$	$C_{111} = -\frac{\sin \alpha}{r^2}$
$C_{120} = \frac{S \sin \alpha}{2r} - a_{12} \cos \alpha$	$C_{102} = \frac{\sin \alpha}{2r^2}$	$C_{202} = -\frac{T}{4r^2} + \frac{\sin^2 \alpha}{2r^3}$
$C_{300} = -a_{30} \cos \alpha + \frac{T \sin \alpha}{2r}$	$C_{211} = \frac{T}{2r^2} - \frac{\sin^2 \alpha}{r^3}$	
$C_{220} = -a_{22} \cos \alpha + \frac{1}{4r} (4a_{20}a_{02} - TS - 2a_{12} \sin 2\alpha) + \frac{S \sin^2 \alpha}{2r^2}$		
$C_{400} = -a_{40} \cos \alpha + \frac{1}{8r} (4a_{20}^2 - T^2 - 4a_{30} \sin 2\alpha) + \frac{T \sin^2 \alpha}{2r^2}$		

^aThe coefficients for which $i \leq 4, j \leq 4, k \leq 2, i + j + k \leq 4$, and $j + k$ is even are included in these tables.

TABLE 4 Coefficients n_{ijk} of the Expansion of F

$n_{ijk} = 0$ for $j, k \neq 0$	
$n_{100} = 1$	$n_{300} = \frac{v_1^2 - v_2}{3}$
$n_{200} = \frac{-v_1}{2}$	$n_{400} = \frac{-v_1^3 + 2v_1v_2 - v_3}{4}$

The coefficient F_{ijk} is related to the strength of the i, j, k aberration of the wavefront diffracted by the grating. The coefficients C_{ijk} and n_{ijk} are given in Tables 3 and 4 in which the following notation is used:

$$T = T(r, \alpha) = \frac{\cos^2 \alpha}{r} - 2a_{20} \cos \alpha, S = S(r, \alpha) = \frac{1}{r} - 2a_{02} \cos \alpha \quad (10)$$

Determination of the Gaussian Image Point

By definition the principal ray AOB_0 arrives at the Gaussian image point $[B_0(r'_0, \beta_0, z'_0)]$ in Fig. 2. Its direction is given by Fermat's principle which implies $(\partial F / \partial w)_{w=0, l=0} = 0, (\partial F / \partial l)_{w=0, l=0} = 0$, from which

$$\frac{m\lambda}{d_0} = \sin \alpha + \sin \beta_0 \quad \frac{z}{r} + \frac{z'_0}{r'_0} = 0 \quad (11)$$

which are the grating equation and the law of magnification in the vertical direction. The tangential focal distance r'_0 is obtained by setting the focusing term F_{200} equal to zero and is given by

$$T(r, \alpha) + T(r'_0, \beta_0) = \begin{cases} 0 & \text{Rowland} \\ -\frac{m\lambda}{\lambda_0} [T(r_C, \gamma) \pm T(r_D, \delta)] & \text{holographic} \\ \frac{v_1 m \lambda}{d_0} & \text{varied line spacing} \end{cases} \quad (12)$$

Equations (11) and (12) determine the Gaussian image point B_0 and, in combination with the sagittal focusing condition ($F_{020} = 0$), describe the focusing properties of grating systems under the paraxial approximation. For a Rowland spherical grating the focusing condition [Eq. (12)] is

$$\left(\frac{\cos^2 \alpha}{r} - \frac{\cos \alpha}{R} \right) + \left(\frac{\cos^2 \beta}{r'_0} - \frac{\cos \beta}{R} \right) = 0 \quad (13)$$

which has the following important special cases:

1. A plane grating ($R = \infty$) implying $r'_0 = -r \cos^2 \beta / \cos^2 \alpha = -r/c_{ff}^2$, so that the focal distance and magnification are fixed if c_{ff} is held constant.³⁷
2. Object and image on the Rowland circle; $r = R \cos \alpha$, $r'_0 = R \cos \beta$, and $M = -1$.
3. $\beta = 0$ (Wadsworth condition).

The tangential focal distances of TGMs and SGMs with or without moving slits are also determined by Eq. (13).

In an aberrated system, the outgoing ray will arrive at the Gaussian image plane at a point B_R displaced from the Gaussian image point B_0 by the ray aberrations $\Delta y'$ and $\Delta z'$ (Fig. 2). The latter are given by³⁸⁻⁴⁰

$$\Delta y' = \frac{r'_0}{\cos \beta_0} \frac{\partial F}{\partial w} \quad \Delta z' = r'_0 \frac{\partial F}{\partial l} \quad (14)$$

where F is to be evaluated for $A = (r, \alpha, z)$, $B = (r'_0, \beta_0, z'_0)$. By means of the series expansion of F , these equations allow the ray aberrations to be calculated separately for each aberration type, as follows:

$$\Delta y'_{ijk} = \frac{r'_0}{\cos \beta_0} F_{ijk} i w^{i-1} l^j \quad \Delta z'_{ijk} = r'_0 F_{ijk} w^i j l^{j-1} \quad (15)$$

Moreover, provided the aberrations are not too large, they are additive, so that they may either reinforce or cancel.

38.4 DISPERSION PROPERTIES

Angular Dispersion

$$\left(\frac{\partial \lambda}{\partial \beta} \right)_\alpha = \frac{d \cos \beta}{m} \quad (16)$$

Reciprocal Linear Dispersion

$$\left(\frac{\partial \lambda}{\partial (\Delta y')} \right)_{\alpha} = \frac{d \cos \beta}{mr'} \equiv \frac{10^{-3} d [\text{\AA}] \cos \beta}{mr' [\text{m}]} \text{\AA/mm} \quad (17)$$

Magnification (M)

$$M(\lambda) = - \frac{\cos \alpha}{\cos \beta} \frac{r'}{r} \quad (18)$$

Phase-Space Acceptance (ε)

$$\varepsilon = N \Delta \lambda_{S_1} = N \Delta \lambda_{S_2} \quad (\text{assuming } S_2 = M S_1) \quad (19)$$

where N is the number of participating grooves.

38.5 RESOLUTION PROPERTIES

The following are the main contributions to the width of the instrumental line spread function (an estimate of the total width is the vector sum):

1. Entrance slit (width S_1):

$$\Delta \lambda_{S_1} = \frac{S_1 d \cos \alpha}{mr} \quad (20)$$

2. Exit slit (width S_2):

$$\Delta \lambda_{S_2} = \frac{S_2 d \cos \beta}{mr'} \quad (21)$$

3. Aberrations (of a perfectly made grating):

$$\Delta \lambda_A = \frac{\Delta y' d \cos \beta}{mr'} = \frac{d}{m} \left(\frac{\partial F}{\partial w} \right) \quad (22)$$

4. Slope error $\Delta \phi$ (of an imperfectly made grating):

$$\Delta \lambda_{SE} = \frac{d(\cos \alpha + \cos \beta) \Delta \phi}{m} \quad (23)$$

Note that, provided the grating is large enough, diffraction at the entrance slit always guarantees a coherent illumination of enough grooves to achieve the slit-width limited resolution. In such cases, a diffraction contribution to the width need not be added to those listed.

38.6 EFFICIENCY

The most accurate way to calculate grating efficiencies is by the full electromagnetic theory for which code is available from Neviere.^{41,42} However, approximate scalar-theory calculations are often useful and, in particular, provide a way to choose the groove depth (h) of a laminar grating. According to Bennett,⁴³ the best value of the groove-width-to-period ratio (r) is the one for which the area of the usefully illuminated groove bottom is equal to that of the top. The scalar theory efficiency of a laminar grating with $r = 0.5$ is given by Franks et al.⁴⁴ as the following:

$$E_0 = \frac{R}{4} \left[1 + 2(1 - P) \cos \left(\frac{4\pi h \cos \alpha}{\lambda} \right) + (1 - P)^2 \right]$$

$$E_m = \begin{cases} R[1 - 2 \cos Q^+ \cos(Q^- + \delta) + \cos^2 Q^+] / m^2 \pi^2 & m = \text{odd} \\ R \cos^2 Q^+ / m^2 \pi^2 & m = \text{even} \end{cases} \quad (24)$$

where

$$P = \frac{4h \tan \alpha}{d_0} \quad Q^\pm = \frac{m\pi h}{d_0} (\tan \alpha \pm \tan \beta) \quad \delta = \frac{2\pi h}{\lambda} (\cos \alpha + \cos \beta)$$

and R is effective reflectance given by $R = \sqrt{R(\alpha)R(\beta)}$ where $R(\alpha)$ and $R(\beta)$ are the intensity reflectances at α and β , respectively.

38.7 REFERENCES

1. H. A. Rowland, "On Concave Gratings for Optical Purposes," *Phil. Mag.* **16**(5th ser.):197–210 (1883).
2. J. E. Mack, J. R. Stehn, and B. Edlen, "On the Concave Grating Spectrograph, Especially at Large Angles of Incidence," *J. Opt. Soc. Am.* **22**:245–264 (1932).
3. H. A. Rowland, "Preliminary Notice of the Results Accomplished in the Manufacture and Theory of Gratings for Optical Purposes," *Phil. Mag.* **13**(supp.) (5th ser.):469–474 (1882).
4. H. G. Beutler, "The Theory of the Concave Grating," *J. Opt. Soc. Am.* **35**:311–350 (1945).
5. H. Haber, "The Torus Grating," *J. Opt. Soc. Am.* **40**:153–165 (1950).
6. T. Namioka, "Theory of the Ellipsoidal Concave Grating: I," *J. Opt. Soc. Am.* **51**:4–12 (1961).
7. W. Welford, "Aberration Theory of Gratings and Grating Mountings," in E. Wolf (ed.), *Progress in Optics*, vol. 4, North-Holland, Amsterdam, pp. 243–282, 1965.
8. J. A. R. Samson, *Techniques of Vacuum Ultraviolet Spectroscopy*, John Wiley & Sons, New York, 1967.
9. W. R. Hunter, "Diffraction Gratings and Mountings for the Vacuum Ultraviolet Spectral Region," in *Spectrometric Techniques*, G. A. Vanasse (ed.), vol. IV, Academic Press, Orlando, pp. 63–180, 1985.
10. T. Namioka and K. Ito, "Modern Developments in VUV Spectroscopic Instrumentation," *Physica Scripta* **37**:673–681 (1988).
11. D. Rudolph and G. Schmahl, "Verfahren zur Herstellung von Röntgenlinsen und Beugungsgittern," *Umsch. Wiss. Tech.* **67**:225 (1967).
12. D. Rudolph and G. Schmahl, "Holographic Gratings," in *Progress in Optics*, E. Wolf (ed.), vol. 14, North-Holland, Amsterdam, pp. 196–244, 1977.
13. A. Laberie and J. Flamand, "Spectrographic Performance of Holographically Made Diffraction Grating," *Opt. Comm.* **1**:5–8 (1969).

14. G. Pieuchard and J. Flamand, "Concave Holographic Gratings for Spectrographic Applications," Final report on NASA contract number NASW-2146, GSFC 283-56,777, Jobin Yvon, Longjumeau, France, 1972.
15. T. Namioka, H. Noda, and M. Seya, "Possibility of Using the Holographic Concave Grating in Vacuum Monochromators," *Sci. Light* **22**:77–99 (1973).
16. H. Noda, T. Namioka, and M. Seya, "Geometrical Theory of the Grating," *J. Opt. Soc. Am.* **64**:1031–1036 (1974).
17. T. Harada and T. Kita, "Mechanically Ruled Aberration-Corrected Concave Gratings," *Appl. Opt.* **19**:3987–3993 (1980).
18. M. C. Hettrick, "Aberration of Varied Line-Space Grazing Incidence Gratings," *Appl. Opt.* **23**:3221–3235 (1984).
19. <http://www-cxro.lbl.gov/>, 1999.
20. C. Kunz, R. Haensel, and B. Sonntag, "Grazing Incidence Vacuum Ultraviolet Monochromator with Fixed Exit Slit for Use with Distant Sources," *J. Opt. Soc. Am.* **58**:1415 (1968).
21. W. R. Hunter, R. T. Williams, J. C. Rife, J. P. Kirkland, and M. N. Kaber, "A Grating/Crystal Monochromator for the Spectral Range 5 eV to 5 keV," *Nucl. Instr. Meth.* **195**:141–154 (1982).
22. R. Follath and F. Senf, "New Plane-Grating Monochromators for Third Generation Synchrotron Radiation Light Sources," *Nucl. Instrum. Meth.* **A390**:388–394 (1997).
23. D. Lepere, "Monochromators with Single Axis Rotation and Holographic Gratings on Toroidal Blanks for the Vacuum Ultraviolet," *Nouvelle Revue Optique* **6**:173 (1975).
24. R. P. Madden and D. L. Ederer, "Stigmatic Grazing Incidence Monochromator for Synchrotrons (abstract only)," *J. Opt. Soc. Am.* **62**:722 (1972).
25. C. T. Chen, "Concept and Design Procedure for Cylindrical Element Monochromators for Synchrotron Radiation," *Nucl. Instr. Meth.* **A256**:595–604 (1987).
26. T. Namioka, "Construction of a Grating Spectrometer," *Sci. Light* **3**:15–24 (1954).
27. M. Seya, "A New Mounting of Concave Grating Suitable for a Spectrometer," *Sci. Light* **2**:8–17 (1952).
28. T. Namioka, M. Seya, and H. Noda, "Design and Performance of Holographic Concave Gratings," *Jap. J. Appl. Phys.* **15**:1181–1197 (1976).
29. W. Eberhardt, G. Kalkoffen, and C. Kunz, "Grazing Incidence Monochromator FLIPPER," *Nucl. Instr. Meth.* **152**:81–4 (1978).
30. K. Miyake, P. R. Kato, and H. Yamashita, "A New Mounting of Soft X-Ray Monochromator for Synchrotron Orbital Radiation," *Sci. Light* **18**:39–56 (1969).
31. H. A. Padmore, "Optimization of Soft X-Ray Monochromators," *Rev. Sci. Instrum.* **60**:1608–1616 (1989).
32. F. C. Brown, R. Z. Bachrach, and N. Lien, "The SSRL Grazing Incidence Monochromator: Design Considerations and Operating Experience," *Nucl. Instrum. Meth.* **152**:73–80 (1978).
33. H. Petersen and H. Baumgartel, "BESSY SX/700: A Monochromator System Covering the Spectral Range 3 eV–700 eV," *Nucl. Instrum. Meth.* **172**:191–193 (1980).
34. W. Jark, "Soft X-Ray Monochromator Configurations for the ELETTRA Undulators: A Stigmatic SX700," *Rev. Sci. Instrum.* **63**:1241–1246 (1992).
35. H. A. Padmore, M. R. Howells, and W. R. McKinney, "Grazing Incidence Monochromators for Third-Generation Synchrotron Radiation Light Sources," in J. A. R. Samson and D. L. Ederer (eds.), *Vacuum Ultraviolet Spectroscopy*, vol. 31, Academic Press, San Diego, pp. 21–54, 1998.
36. S. Y. Rah, S. C. Irick, and M. R. Howells, "New Schemes in the Adjustment of Bendable Elliptical Mirrors Using a Long-Trace Profiler," in P. Z. Takacs and T. W. Tonnessen (eds.), *Materials Manufacturing and Measurement for Synchrotron-Radiation Mirrors*, Proc. SPIE, vol. 3152, SPIE, Bellingham, WA, 1997.
37. H. Petersen, "The Plane Grating and Elliptical Mirror: A New Optical Configuration for Monochromators," *Opt. Comm.* **40**:402–406 (1982).
38. M. Born and E. Wolf, *Principles of Optics*, Pergamon, Oxford, 1980.
39. T. Namioka and M. Koike, "Analytical Representation of Spot Diagrams and Its Application to the Design of Monochromators," *Nucl. Instrum. Meth.* **A319**:219–227 (1992).
40. W. T. Welford, *Aberrations of the Symmetrical Optical System*, Academic Press, London, 1974.
41. M. Nevriere, P. Vincent, and D. Maystre, "X-Ray Efficiencies of Gratings," *Appl. Opt.* **17**:843–845 (1978). (Nevriere can be reached at michel.nevriere@fresnel.fr.)

- 42. R. Petit (ed.), *Electromagnetic Theory of Gratings (Topics in Current Physics, Vol. 22)*, Springer Verlag, Berlin, 1980.
- 43. J. M. Bennett, "Laminar X-Ray Gratings," Ph.D. Thesis, London University, London, 1971.
- 44. A. Franks, K. Lindsay, J. M. Bennett, R. J. Speer, D. Turner, and D. J. Hunt, "The Theory, Manufacture, Structure and Performance of NPL X-Ray Gratings," *Phil. Trans. Roy. Soc.* **A277**:503–543 (1975).

CRYSTAL MONOCHROMATORS AND BENT CRYSTALS

Peter Siddons

*National Synchrotron Light Source
Brookhaven National Laboratory
Upton, New York*

39.1 CRYSTAL MONOCHROMATORS

For x-ray energies higher than 2 keV or so, gratings become extremely inefficient, and it becomes necessary to utilize the periodicity naturally occurring in a crystal to provide the dispersion. Since the periodicity in a crystal is 3-dimensional, the normal single grating equation must be replaced by the three grating equations, one for each dimension, called the Laue equations,¹ as follows:

$$\begin{aligned} \mathbf{a}_1 \cdot (\mathbf{k}_{H_1 H_2 H_3} - \mathbf{k}_0) &= H_1 \\ \mathbf{a}_2 \cdot (\mathbf{k}_{H_1 H_2 H_3} - \mathbf{k}_0) &= H_2 \\ \mathbf{a}_3 \cdot (\mathbf{k}_{H_1 H_2 H_3} - \mathbf{k}_0) &= H_3 \end{aligned} \tag{1}$$

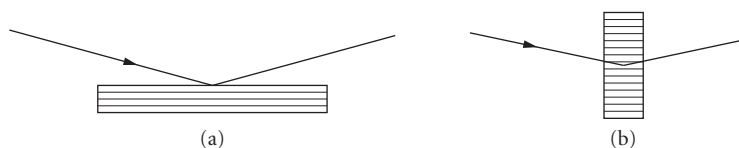
where the $\mathbf{a}$'s are the repeat vectors in the three dimensions, the $\mathbf{k}$'s are the wave vectors for the incident and scattered beams, and the H 's are integers denoting the diffraction order in the three dimensions. All of them must be simultaneously satisfied in order to have an interference maximum (commonly called a *Bragg reflection*). One can combine these equations into the well-known Bragg's law² for one component of the crystalline periodicity (usually referred to as a set of *Bragg planes*),

$$n\lambda = 2d \sin \theta \tag{2}$$

where n is an integer indicating the order of diffraction from planes of spacing d , and θ is the angle between the incident beam and the Bragg planes. This equation is the basis for using crystals as x-ray monochromators. By choosing one such set of Bragg planes and setting the crystal so that the incident x rays fall on these planes, the wavelength of the light reflected depends on the angle of incidence of the light. Table 1 shows some commonly used crystals and the spacings of some of their Bragg planes. The most common arrangement is the symmetric Bragg case (Fig. 1a), in which the useful surface of the crystal is machined so that it is parallel to the Bragg planes in use. Under these conditions, the incident and reflected angles are equal. The literature on crystal diffraction differs from that on grating instruments in that the incidence angle for x rays is called the *glancing angle* for

TABLE 1 Some Common Monochromator Crystals, Selected d-Spacings, and Reflection Widths at 1 Å (12.4 keV)

Crystal	Reflection	d-Spacing (nm)	Refl. Width (μrad)	Energy Resolution ($\Delta E/E$)
Silicon	(1 1 1)	0.31355	22.3	1.36×10^{-4}
	(2 2 0)	0.19201	15.8	5.37×10^{-5}
Germanium	(1 1 1)	0.32664	50.1	3.1×10^{-4}
	(2 2 0)	0.20002	37.4	1.37×10^{-4}
Diamond	(1 1 1)	0.20589	15.3	5.8×10^{-5}
	(4 0 0)	0.089153	5.2	7.4×10^{-6}
Graphite	(0 0 0 . 2)	0.3354	Sample-dependent	Sample-dependent

**FIGURE 1** The two most usual x-ray diffraction geometries used for monochromator applications: (a) the symmetric Bragg case and (b) the symmetric Laue case.

gratings. As the x-ray wavelength gets shorter and the angles get smaller, it can be difficult to obtain large enough crystals of good quality. In such cases it is possible to employ the Laue case (Fig. 1b), in which the surface is cut perpendicular to the Bragg planes and the x rays are reflected through the bulk of the crystal plate. Of course, the wavelength should be short enough or the crystal thin enough so that the x rays are not absorbed by the monochromator. This is true, for example, in silicon crystals around 1 mm thick above an x-ray energy of around 30 keV ($\lambda = 0.04$ nm).

The detailed calculation of the response of a crystal to x rays depends on the degree of crystalline perfection of the material in use, as well as its chemical composition. Two main theoretical treatments are commonly used: the kinematical theory of diffraction and the dynamical theory. The kinematical theory assumes single-scattering of the x rays by the crystal, and is appropriate for crystals with a high concentration of defects. Such crystals are called *mosaic crystals*, following C. G. Darwin.³ The dynamical theory, in contrast, explicitly treats the multiple scattering that arises in highly perfect crystals and is commonly used for the semiconductor monochromator materials that can be grown to a high degree of perfection. Of course, many crystals fall between these two idealized pictures and there exist approximations to both theories to account for some of their failures. We will not describe these theories in detail here, but will refer the reader to texts on the subject⁴⁻⁶ and content ourselves with providing some of the key formulas that result from them.

Both theories attempt to describe the variation in reflectivity of a given crystal as a function of the incidence angle of the x-ray beam near a Bragg reflection. The neglect or inclusion of multiple scattering changes the result quite dramatically. In the kinematical case, the width of the reflectivity profile is inversely related to the size of the coherently diffracting volume, and for an infinite perfect crystal it is a delta function. The integrated reflectivity increases linearly with the illuminated crystal volume, and is assumed to be small. In the dynamical case, the x-ray beam diffracted by one part of the crystal is exactly oriented to be diffracted by the same Bragg planes, but in the opposite sense. Thus, there coexist two waves in the crystal, one with its wave vector along the incident beam direction and the other with its vector along the diffracted beam direction. These waves are coherent and can interfere. It is these interferences that give rise to all the interesting phenomena that arise from this theory. The integrated reflectivity in this case initially increases with volume, as in the kinematical case, but eventually saturates to a constant value that depends on which of the geometries in Fig. 1 is taken.

For the kinematical theory, the reflectivity curve is approximated by⁷

$$r(\Delta) = \frac{a}{[1 + a + \sqrt{1 + 2a} \cdot \cot h(b\sqrt{1 + 2a})]} \quad (3)$$

where

$$a = \frac{w(\Delta)Q}{\mu}$$

$$b = \frac{\mu t}{\sin \theta_B}$$

$$Q = \left(\frac{e^2}{mc^2V} \right)^2 \cdot \frac{|F_H|^2 \lambda^3}{\sin 2\theta_B}$$

in which e is the electronic charge and m its mass, c is the velocity of light, θ_B is the Bragg angle for the planes whose structure factor is F_H at wavelength λ . $w(\Delta)$ represents the angular distribution of the mosaic blocks, which depends on the details of the sample preparation and/or growth history. The reflection width will primarily reflect the width of this parameter $w(\Delta)$, but often the curve will be further broadened by extinction.

The equivalent equation for the dynamical theory is sensitive to the exact geometry under consideration and so will not be given here. The crystal becomes birefringent near the Bragg angle, and this causes some interesting interference effects that are worthy of study in their own right. Reference 8 is a review of the theory with a physical approach that is very readable. However, the main results for the purposes of this section are summarized in Fig. 2; the key points are the following:

1. In the Bragg case (Fig. 1a) the peak reflectivity reaches nearly unity over a small range of angles near the Bragg angle.

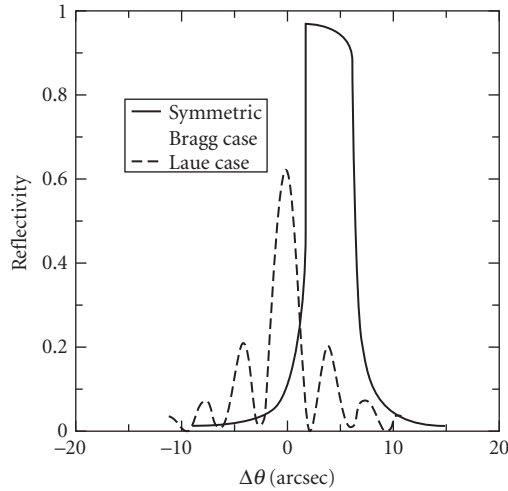


FIGURE 2 The perfect crystal reflectivity curves for the (111) reflection at 1 Å wavelength. The solid curve is for the thick Bragg case, and the dashed curve is for the Laue case for a 20 μm thick crystal

2. The range of angles over which this occurs is given by

$$\Omega = 2R\lambda^2 |C| \frac{\sqrt{|\gamma| F_h F_h}}{\pi V \sin 2\theta} \quad (4)$$

where R is the classical electron radius, 2.81794×10^{-15} m, λ is the x-ray wavelength, C is the polarization factor (1 for σ -polarized light and $\cos 2\theta$ for π -polarized light). γ is the asymmetry parameter (1 in the symmetric case), the F s are the structure factors for reflections (hkl) and ($-h-k-l$), V is the unit cell volume, and θ is the Bragg angle.

3. The effect of absorption in the Bragg case is to reduce the peak reflectivity and to make the reflectivity curve asymmetric.
4. In the Laue case, the reflectivity is oscillatory, with an average value of $1/2$, and the effect of absorption (including increasing thickness) is to damp out the oscillations and reduce the peak reflectivity.

Even small distortions can greatly perturb the behavior of perfect crystals, and there is still no general theory that can handle arbitrary distortions in the dynamical regime. There are approximate treatments that can handle some special cases of interest.^{4,5}

In general, the performance of a given monochromator system is determined by the range of Bragg angles experienced by the incident x-ray beam. This can be determined simply by the incident beam collimation or by the crystal reflectivity profile, or by a combination of both factors. From Bragg's law we have

$$\frac{\delta E}{E} = \frac{\delta \lambda}{\lambda} = \cot \theta \delta \theta + \delta \tau \quad (5)$$

where δE and E are the passband and center energy of the beam, and $\delta \tau$ is the intrinsic energy resolution of the crystal reflection as given in Table 1—essentially the dynamical reflection width as given previously, expressed in terms of energy. This implies that, even for a perfectly collimated incident beam, there is a limit to the resolution achievable, which depends on the monochromator material and the diffraction order chosen.

The classic Bragg reflection monochromator uses a single piece of crystal set to the correct angle to reflect the energy of interest. This is in fact a very inconvenient instrument, since its output beam direction changes with energy. For perfect-crystal devices, the reflectivity is sufficiently high that one can afford to use two of them in tandem, deviating in opposite senses in order to bring the useful beam back into the forward direction, independent of energy (Fig. 3). Such two-crystal monochromators have become the standard design, particularly for use at accelerator-based (synchrotron) radiation sources (as discussed in Chap. 55). Since these sources naturally generate an energy continuum, a monochromator is a common requirement for a beamline at such a facility. There is a wealth of literature arising from such applications.⁹

A particularly convenient form of this geometry was invented by Bonse and Hart,¹⁰ in which the two reflecting surfaces are machined from a single-crystal block of material (in their case germanium, but more often silicon in recent years).

For the highest resolution requirements, it is possible to take two crystals that deviate in the same direction, the so-called ++ geometry. In this case the first crystal acts as a collimator, generating a fan of beams with each angular component having a particular energy. The second one selects which angular (and hence energy) component of this fan to transmit. In this arrangement the deviation is doubled over the single-crystal device, and so the most successful arrangement for this so-called dispersive arrangement is to use two monolithic double-reflection devices like that in Fig. 2, with the ++ deviation taking place between the second and third reflections (Fig. 4).

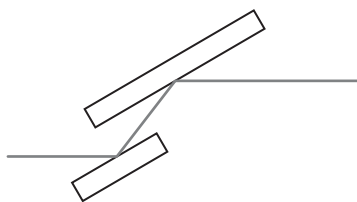


FIGURE 3 The $\pm$ double-crystal x-ray monochromator.

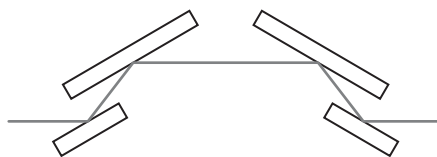


FIGURE 4 Two monolithic double reflectors arranged in a high-resolution configuration.

39.2 BENT CRYSTALS

There are two reasons to consider applying a uniform curvature to a crystal plate for a monochromator system. One is to provide some kind of beam concentration or focussing, and the other is to improve the energy resolution in circumstances where a divergent incident beam is unavoidable. The most common geometry is the Bragg-case one based on the Rowland circle principle, modified to account for the 3-dimensional nature of the crystal periodicity. This principle relies on the well-known property of a circle that an arc segment subtends a constant angle for any point on the circle. For a grating this is straightforwardly applied to a focussing spectrometer, but for crystals one has the added complication that the incidence angle for the local Bragg plane must also be constant. The result was shown by Johansson¹¹ (Fig. 5) to require the radius of curvature to be twice the radius of the crystal surface. This can be achieved by a combination of bending and machining the crystal. For applications in which the required optical aperture is small, the aberrations introduced by omitting the machining operation are small and quite acceptable.¹² Since this is a rather difficult operation, it is attractive to avoid it if possible.

Although Fig. 5 shows the symmetrical setting, it is possible to place the crystal anywhere on the circle and achieve a (de)magnification other than unity. In this case the surface must be cut at an angle to the Bragg planes to maintain the geometry. The Laue case can also be used in a similar arrangement, but the image becomes a virtual image (or source). For very short wavelengths (e.g., in a gamma-ray spectrometer) the Laue case can be preferable since the size of the crystal needed for a given optical aperture is much reduced. In both Laue and Bragg cases, there are changes in the reflection properties on bending. Depending on the source and its collimation geometry, and on the asymmetry angle of the crystal cut, the bending can improve or degrade the monochromator resolution. Each case must

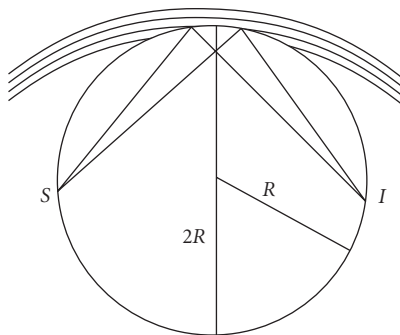


FIGURE 5 The Johansson bent/ground focussing monochromator.



FIGURE 6 The usual arrangement of a sagittally focussing monochromator, with the bent element as the second one of a two-crystal monochromator.

be considered in detail. Again, within the scope of this section we will content ourselves with some generalities:

1. If the angular aperture of the crystal as seen from the source location is large compared to the reflectivity profile of the crystal, then the Rowland circle geometry will improve things for the Bragg case and the symmetric Laue case.
2. When the absorption length of the incident radiation becomes large compared to the extinction distance, then the x rays can travel deep into the bent crystal even in the Bragg case, and the deformation means that the incidence angle changes with depth, leading to a broadening of the bandpass.
3. If the bending radius becomes very small, such that the Bragg angle changes more than the perfect-crystal reflection width within the extinction depth, then the peak reflectivity will fall and the reflection width will increase, and consequently the resolution will deteriorate.
4. In the asymmetric Laue case, the reflectivity profile width for perfect crystals also depends on the curvature,⁶ and so for strongly collimated beams such as synchrotron radiation sources, the bending may well degrade the resolution at the same time as it increases the intensity.

Arrangements of multiple consecutive curved crystals are unusual, but have found application in high-throughput synchrotron radiation (SR) monochromators, where the two-crystal device in Fig. 3 is modified by placing the first crystal on the Rowland circle and adjusting the convex curvature of the second to maximize its transmission of the convergent beam from the first.¹³ There are also examples of combinations of curved Laue and curved Bragg reflectors.¹⁴

Another geometry in which the technique of bending a monochromator crystal can be used is in the so-called sagittal-focussing geometry.¹⁵ This geometry, as its name indicates, has the bending radius perpendicular to that in the Rowland-circle geometry, and provides a focussing effect in the plane perpendicular to the diffraction plane. In the diffraction plane the crystal behaves essentially as though it were flat (Fig. 6). Antielastic bending of the bent crystal can greatly reduce its efficiency, since this curvature is in the meridional plane so that parts of the crystal move out of the Bragg-reflection range. Attempts to counteract this by using stiffening ribs machined into the plate¹⁶ or by making use of the elastic anisotropy of single crystals¹⁷ have been made, with some success.

39.3 REFERENCES

1. M. von Laue, *Münchener Sitzungsberichte* 363 (1912), and *Ann. der Phys.* **41**:989 (1913).
2. W. H. Bragg and W. L. Bragg, *Proc. Roy. Soc. London* **88**:428 (1913), and **89**:246 (1913).
3. C. G. Darwin, *Phil. Mag.* **27**:315, 657 (1914), and **43**:800 (1922).
4. D. Taupin, *Bull. Soc. Franc. Miner. Crist.* **87**:469–511 (1964); S. Takagi, *Acta Cryst.* **12**:1311 (1962).
5. Penning and Polder, *Philips Res. Rep.* **16**:419 (1961).
6. R. Caciuffo, S. Melone, F. Rustichelli, and A. Boeuf, *Phys. Rep.* **152**:1 (1987); P. Suortti and W. C. Thomlinson, *Nuclear Instrum. and Meth.* **A269**:639–648 (1988).

7. A. K. Freund, A. Munkholm, and S. Brennan, *Proc. SPIE* **2856**:68–79 (1996).
8. B. W. Batterman and H. Cole, *Rev. Mod. Phys.* **36**:681–717 (1964). (For a fuller treatment, see W. H. Zachariasen, *Theory of X-Ray Diffraction in Crystals*, Dover, New York, 1967.)
9. See for example, the *Handbook of Synchrotron Radiation* or browse the *Proceedings of the International Conference on Synchrotron Radiation Instrumentation*, published in *Nucl. Instr. and Meth.*, *Rev. Sci. Instrum.*, and *J. Synchr. Rad.* at various times.
10. U. Bonse and M. Hart, *Appl. Phys. Lett.* **7**:238–240 (1965).
11. T. Johansson, *Z. Phys.* **82**:507 (1933).
12. H. H. Johann, *Z. Phys.* **69**:185 (1931).
13. M. J. Van Der Hoek, W. Berne, and P. Van Zuylen, *Nucl. Instrum. and Meth.* **A246**:190–193 (1986).
14. U. Lienert, C. Schulze, V. Honkimäki, Th. Tschentschor, S. Garbe, O. Hignette, A. Horsewell, M. Lingham, H. F. Poulsen, W. B. Thomsen, and E. Ziegler, *J. Synchrot. Rad.* **5**:226–231 (1998).
15. L. Von Hamos, *J. Sci. Instr.* **15**:87 (1938).
16. C. J. Sparks, Jr., B. S. Borie, and J. B. Hastings, *Nucl. Instrum. and Meth.* **172**:237 (1980).
17. V. I. Kushnir, J. P. Quintana, and P. Georgopoulos, *Nucl. Instrum. and Meth.* **A328**:588 (1983).

This page intentionally left blank

ZONE PLATES

Alan Michette

*King's College London
United Kingdom*

40.1 INTRODUCTION

Some radiation incident on a linear transmission grating passes straight through (the *zero order*), some is diffracted to one side of the zero order (the *positive orders*), and some is diffracted to the other side (the *negative orders*). In the first order, the diffraction angle is $\beta = \sin^{-1}(\lambda/d) \approx \lambda/d$ in the small angle approximation; d is the grating period. Thus, for smaller periods, radiation is diffracted through larger angles. A *circular* grating with a constant period would therefore form an axial line focus of a point source (Fig. 1a), and the distance from a radial point r on the grating to a point on the axis is $z = r/\tan\beta \approx rd/\lambda$.

If the period is made to decrease as the radius increases (Fig. 1b) then the distance z can be made constant. The grating then acts as a lens in that radiation from a point source is brought to an axial focus (Fig. 1c). The positive diffraction orders are now defined as being on the opposite side to the source, with the negative orders on the same side.

This is the basis of zone plates, the focusing properties of which depend on

- The relationship between d and r
- The number of zones (For x-ray zone plates the usual convention is that the area between successive boundaries is a zone. Strictly speaking, and in keeping with the terminology used for diffraction gratings, this area should be called a *half-period zone* but zone is usually used.)
- The zone heights and profiles

40.2 GEOMETRY OF A ZONE PLATE

Referring to Fig. 1c, radiation from an object point A is brought to a focus, via the zone plate, to an image point B. To obtain constructive interference at B the optical path difference between successive zone boundaries must be $\pm m\lambda/2$, where m is the diffraction order. Thus, for the first order,

$$a_n + b_n = z_a + z_b + \frac{n\lambda}{2} \quad (1)$$

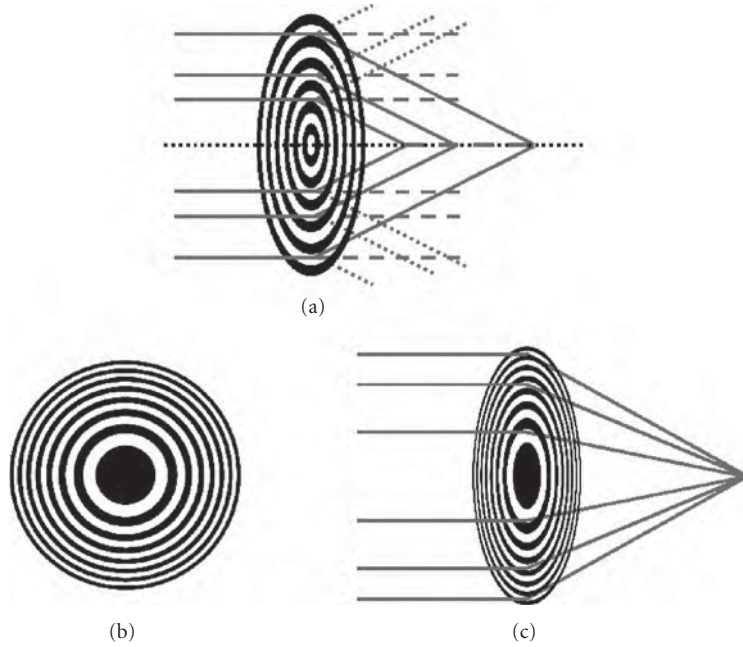


FIGURE 1 Diffraction by (a) a circular grating of constant period and (b, c) a zone plate. (See also color insert.)

where n is the zone number, counting outward from the center, and Δ is the optical path difference introduced by the central zone of radius r_0 . For a distant source ($a_n, z_a \rightarrow \infty$ with $a_n - z_a \rightarrow 0$) and with

$$b_n = \sqrt{z_b^2 + r_n^2} = \sqrt{f_1^2 + r_n^2} \quad (2)$$

where r_n is the radius of the n th zone and f_1 is the first-order focal length, squaring and simplifying leads to

$$r_n^2 = p\lambda f_1 + \left(\frac{p\lambda}{2}\right)^2 \quad (3)$$

where $p = n + 2\Delta/\lambda$.

For a finite source or object distance Eq. (3) still holds with the addition of higher-order terms in λ and if the term in λ^2 is multiplied by $(M^3+1)/(M+1)^3$, where M is the magnification. In most practical cases, terms in λ^2 and above are negligible and so, to a good approximation,

$$r_n^2 = n\lambda f_1 + 2\Delta f_1 = n\lambda f_1 + r_0^2 \quad (4)$$

since, for the central zone, $n = 0$ and $r_0^2 = 2\Delta f_1$. Equation (4) describes the *Fresnel zone plate* and, for $r_0 = 0$, the *Fresnel-Soret zone plate* (often referred to as the Fresnel zone plate). The latter is the most commonly used, with

$$r_n^2 = n\lambda f_1 = nr_1^2 \quad (5)$$

The higher-order terms ignored in deriving Eq. (5) result in aberrations. In particular, the term in λ^2 describes spherical aberration but only becomes comparable to the first term when $n \sim 4f_1/\lambda$, which

is rarely the case for x-ray zone plates since focal lengths are typically several orders of magnitude larger than the wavelength.

Equation (5) shows that the focal length is inversely proportional to the wavelength, so that monochromatic radiation with $\lambda/\Delta\lambda \sim N$, where N is the total number of zones, is needed to avoid chromatic aberration. The area of the n th zone is

$$\pi(r_n^2 - r_{n-1}^2) = \pi[n\lambda f_1 - (n-1)\lambda f_1] = \pi\lambda f_1 \quad (6)$$

which is constant, so that each zone contributes equally to the amplitude at the focus if the zone plate is evenly illuminated. The width d_n of the n th zone is

$$d_n = r_n - r_{n-1} = \sqrt{n\lambda f_1} - \sqrt{(n-1)\lambda f_1} = \sqrt{n\lambda f_1} \left[1 - \left(1 - \frac{1}{n} \right)^{1/2} \right] \approx \frac{r_n}{2n} \quad (7)$$

leading to an expression for the first-order focal length

$$f_1 = \frac{r_n^2}{n\lambda} \approx \frac{D_n d_n}{\lambda} \quad (8)$$

where D_n is the diameter of the n th zone. If D is the overall zone plate diameter and d is the outer zone width then

$$f_1 = \frac{Dd}{\lambda} \quad (9)$$

Since zone plates are diffractive optics they have many foci, corresponding to different diffraction orders. The m th-order focus can be described by m zones acting in tandem, so that the effective period is md and the focal lengths are given by

$$f_m = f_1/m \quad m=0, \pm 1, \pm 2, \pm 3, \dots \quad (10)$$

Positive values of m give real foci, while negative values give virtual foci and $m = 0$ corresponds to undiffracted, that is, unfocused radiation.

40.3 ZONE PLATES AS THIN LENSES

The sizes of the focal spots for a point object—the diffraction pattern at a focus—should be determined by successively adding (for an open zone) and subtracting (for a closed zone) the diffraction patterns of circular apertures of radii r_n .¹ However, when N is large enough (theoretically greater than ~ 100 , but in practice much less) a zone plate acts as a thin lens, so that the object, u , and image, v_m (in the m th order), distances are related by

$$\frac{1}{u} + \frac{1}{v_m} = \frac{1}{f_m} \quad (11)$$

and the diffraction pattern at a focus approximates to an Airy pattern.

For a lens of diameter D and focal length f the first zero of the Airy distribution, at a radius $f \tan(1.22\lambda/D)$, defines the lateral resolution ρ via the Rayleigh criterion. For a zone plate, using the expressions for the focal lengths and the small angle approximation, this gives the resolution in the m th order

$$\rho_m = 1.22 \frac{d}{m} \quad (12)$$

Equation (12) shows that, for high resolution, the outermost zone width must be small and that better resolutions can be obtained from higher diffraction orders. However, the lower diffraction efficiencies (see Sec. 40.4) in the higher orders can negate this advantage.

The depth of focus Δf_m is also determined using the thin lens analogy; for a thin lens $\Delta f = \pm 2(f/D)^2 \lambda$, which, for a zone plate, leads to

$$\Delta f_m = \pm \frac{f}{2mN} \quad (13)$$

40.4 DIFFRACTION EFFICIENCIES OF ZONE PLATES

The zone plate properties discussed so far depend solely on the relative placement of the zone boundaries; how much radiation can be focused into the various diffraction orders depends additionally on the zone heights and profiles.

Amplitude Zone Plates

A full analysis of the efficiency requires taking the Fourier transform of the zone distribution.² However, if the zone boundaries are in the correct positions, as discussed above, and for an amplitude zone plate in which alternate zones are totally absorbing or transmitting, a simpler discussion suffices. In this case half of the incident radiation is absorbed and half of the rest goes into the zeroth, undiffracted, order. The other even orders vanish since the amplitudes from adjacent zones cancel. The only orders which contribute are 0, ± 1 , ± 3 , ... and, from symmetry, it is clear that the $+m$ th and $-m$ th diffraction efficiencies are equal.

Thus 25 percent of the incident radiation remains to be distributed between the odd orders. The peak amplitudes in each diffraction order are equal, but Eq. (12) shows that the focal spot areas decrease as m^2 . Hence, if ϵ_m is the diffraction efficiency in the m th order,

$$0.25 = 2 \sum_{\substack{m=1 \\ m \text{ odd}}}^{\infty} \epsilon_m = 2\epsilon_1 \sum_{\substack{m=1 \\ m \text{ odd}}}^{\infty} \frac{1}{m^2} = 2\epsilon_1 \frac{\pi^2}{8} \quad (14)$$

so that

$$\epsilon_0 = 0.25; \quad \epsilon_m = \frac{1}{m^2 \pi^2} \quad m = \pm 1, \pm 3, \pm 5, \dots; \quad \epsilon_m = 0 \quad m = \pm 2, \pm 4, \dots \quad (15)$$

The first order therefore gives the highest focused intensity, but even so it is only ≈ 10 percent efficient. If the zone boundaries are displaced from the optimum positions then intensity is distributed into the even orders, at the expense of the odd, to a maximum of $1/m^2 \pi^2$ (Fig. 2). If the clear zones are not totally transmitting but have amplitude transmission A_1 —because of, for example, a supporting substrate—and the other zones have amplitude transmission A_2 , then the diffraction efficiencies are reduced by a factor $(A_1^2 - A_2^2)$.

The multiplicity of diffraction orders means that this type of zone plate must normally be used with an axial stop and a pinhole, the *order selecting aperture* (OSA), as shown in Fig. 3, to prevent loss of image contrast. The axial stop, typically with a diameter $\approx 0.4D$, reduces the focused intensity and the width of the central maximum of the diffraction pattern, while putting more intensity into the outer lobes. The pinhole also removes any other wavelengths present, so that zone plates can be used as linear monochromators.³

An alternative type of amplitude zone plate, the Gabor zone plate, has, instead of a square wave amplitude transmittance $T(r)$, an approximately sinusoidal one

$$T(r) = \frac{1}{2} \left[1 + \sin \frac{\pi r^2}{\lambda f_1} \right] \quad (16)$$

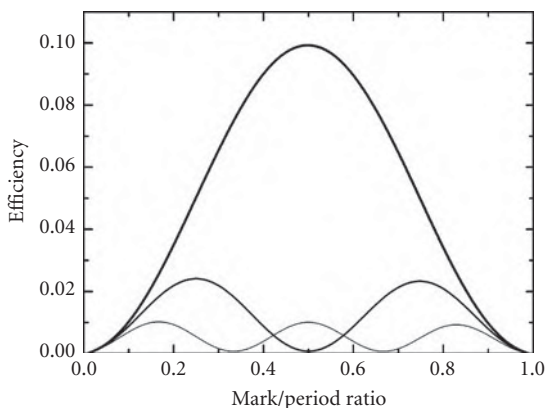


FIGURE 2 Amplitude zone plate diffraction efficiencies: heavy curve, first order; medium curve, second order; light curve, third order.

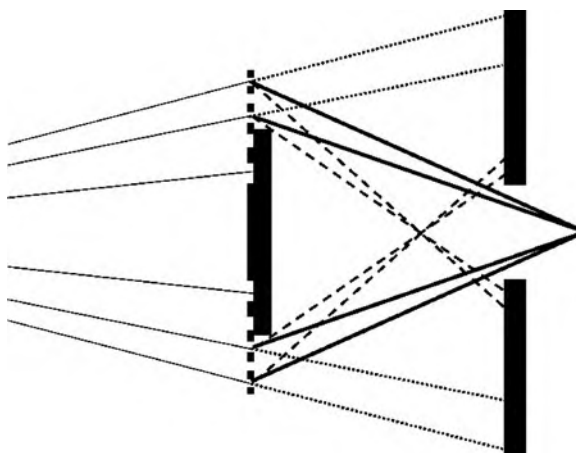


FIGURE 3 Removal of unwanted diffraction orders by use of an axial stop and a pinhole.

The diffraction efficiencies are then 0.25 in the zero order, 1/16 in the positive and negative first orders and zero in all other orders; the remaining 5/8 of the incoming intensity is absorbed. The OSA is no longer needed, but the central stop is, and the first-order diffraction efficiency is less than that for an ordinary amplitude zone plate. Gabor zone plates, with the correct profiles, are also more difficult to make.

Phase Zone Plates

If alternate zones can be made to change the phase of the radiation rather than (just) absorbing it, then the amplitude at a focus can be increased. In the absence of absorption, a phase change of π radians would double the focused amplitude so that the diffraction efficiency in the first order would be increased to ≈ 40 percent for rectangular zones. This is not possible for x rays since there

is always some absorption, but a significant improvement in diffraction efficiency can be made if zones of the correct thickness, determined as in the following analysis,⁴ are made.

Pairs of adjacent zones contribute equally to the overall amplitude in a given diffraction order, and so only one pair needs to be considered. The first zone of a pair is assumed to be open and the second has thickness t so that the amplitude is attenuated by a factor $\exp(-2\pi\beta t/\lambda)$ and the phase is retarded by $\Delta\phi = 2\pi\delta t/\lambda$, where δ and β are the optical constants defined by the complex refractive index

$$\tilde{n} = 1 - \delta - i\beta \quad (17)$$

The amplitude at the first-order focus from an open zone is

$$A_o = iC/\pi \quad (18)$$

where $C^2 = I_0$ is the intensity incident on the zone pair. From the phase-shifting zone,

$$A_p = -\frac{iC}{\pi} \exp(-i\Delta\phi) \exp\left(-\frac{2\pi\beta t}{\lambda}\right) \quad (19)$$

so that the contribution from a pair of zones to the intensity at the focus is

$$I_{f_1} = |A_o + A_p|^2 = \left(\frac{C}{\pi}\right)^2 [1 + \exp(-2\eta\Delta\phi) - 2\cos\Delta\phi \exp(-\eta\Delta\phi)] \quad (20)$$

where $\eta = \beta/\delta = 2\pi\beta t/\lambda\Delta\phi$. As for a square-wave amplitude zone plate, the focused intensities in the higher orders are $1/m^2$ in the first order, for odd positive and negative values of m . The maximum intensities are then determined by differentiating Eq. (20) with respect to $\Delta\phi$,

$$\frac{\partial I_{f_m}}{\partial(\Delta\phi)} = 0 = 2\left(\frac{C}{m\pi}\right)^2 [-\eta \exp(-2\eta\Delta\phi) + (\sin\Delta\phi + \eta \cos\Delta\phi) \exp(-\eta\Delta\phi)] \quad (21)$$

Equation (21) shows that the optimum phase shift $\Delta\phi_{\text{opt}}$ is given by the nontrivial solution of

$$\eta \exp(-\eta\Delta\phi_{\text{opt}}) = \sin\Delta\phi_{\text{opt}} + \eta \cos\Delta\phi_{\text{opt}} \quad (22)$$

with two limiting cases $\eta \rightarrow \infty$ for an amplitude zone plate and $\eta \rightarrow 0$ for a phase zone plate with no absorption. Substituting for $\eta \exp(-\eta\Delta\phi_{\text{opt}})$ in Eq. (21) and dividing by C^2 gives the m th-order diffraction efficiency for the optimum phase shift

$$\epsilon_m = \frac{1}{m^2\pi^2} \left(1 + \frac{1}{\eta^2}\right) \sin^2\Delta\phi_{\text{opt}} \quad (23)$$

The undiffracted amplitudes through the open and phase-shifting zones are

$$A_{o_u} = \frac{C}{2} \quad A_{p_u} = \frac{C}{2} \exp(-i\Delta\phi_{\text{opt}}) \exp(-\eta\Delta\phi_{\text{opt}}) \quad (24)$$

so that the zero-order intensity is

$$I_u = |A_{o_u} + A_{p_u}|^2 = \left(\frac{C}{2}\right)^2 [1 + \exp(-2\eta\Delta\phi_{\text{opt}}) + 2\cos\Delta\phi_{\text{opt}} \exp(-\eta\Delta\phi_{\text{opt}})] \quad (25)$$

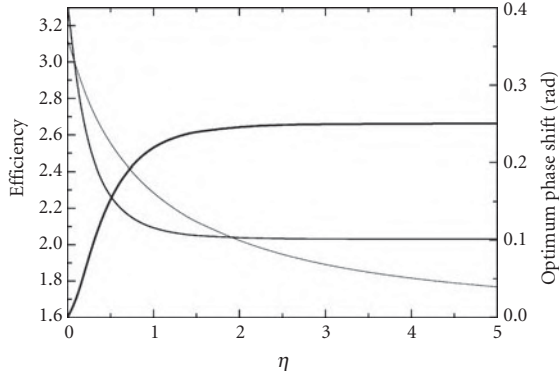


FIGURE 4 Zero (heavy curve) and first-order (medium curve) diffraction efficiencies of a phase zone plate at the optimum phase shift (light curve).

leading to the zero-order efficiency

$$\epsilon_0 = 0.25 \left[\sin^2 \Delta\phi_{\text{opt}} + \left(2 \cos \Delta\phi_{\text{opt}} + \frac{\sin \Delta\phi_{\text{opt}}}{\eta} \right)^2 \right] \quad (26)$$

Since $I_0 = C^2$ is the intensity incident on a zone pair, $I_0/2$ is transmitted by the open zone and $(I_0/2) \exp(-2\eta\Delta\phi_{\text{opt}})$ by the phase-shifting one, so that the total transmitted intensity is

$$I_t = \frac{C^2}{2} [1 + \exp(-2\eta\Delta\phi_{\text{opt}})] \quad (27)$$

leading to the total fractional transmitted intensity at the optimum phase shift

$$\epsilon_t = 0.5 \left[2 - \left(1 - \frac{1}{\eta^2} \right) \sin^2 \Delta\phi_{\text{opt}} + \frac{1}{\eta} \sin 2\Delta\phi_{\text{opt}} \right] \quad (28)$$

Figure 4 shows the variation of the zero and first-order diffraction efficiencies as functions of η , and Fig. 5 gives an example of the variation of the first-order efficiency with thickness, calculated using Eq. (28) for nickel at a wavelength of 3.37 nm. These figures demonstrate the significant enhancement in efficiency possible over that of an amplitude zone plate.

Applying a similar analysis to a Gabor zone plate gives a corresponding increase in the diffraction efficiency. Higher efficiencies could be obtained by using zone profiles in which the phase shift varies continuously across each zone.⁵ In the absence of absorption it is then possible, in principle, for any given diffraction order to contain 100 percent of the incident intensity. It is not yet possible to make such structures at high resolution, but stepped approximations to the profile have demonstrated efficiencies of ≈ 55 percent at an energy of 7 keV.⁶

Volume Effects

In order to achieve the optimum phase shift discussed in the section “Phase Zone Plates,” the zone thickness required is

$$t_{\text{opt}} = \frac{\Delta\phi_{\text{opt}} \lambda}{2\pi\delta} \quad (29)$$

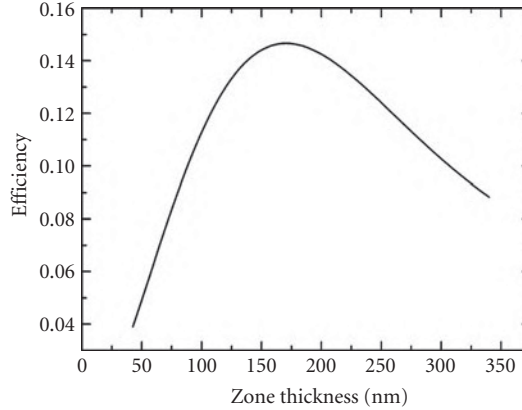


FIGURE 5 The first-order diffraction efficiency of a nickel zone plate at a wavelength of 3.37 nm.

Figure 5 shows that for a nickel zone plate at 3.37 nm, t_{opt} is around 170 nm. For high spatial resolution this means that the aspect ratio, t_{opt}/d , is large, and increases for shorter wavelengths. As well as the resulting technological problems, the previous discussion of spatial resolution in terms of the minimum zone width is no longer valid. The minimum zone width, introduced as the validity criterion of scalar diffraction theory,⁷ is

$$d_{\text{min}} = \sqrt{m\lambda t_{\text{opt}}} \quad (30)$$

this approximation being in good agreement with rigorous electromagnetic theory and with the theory of volume holograms.⁸ For zones with spacing less than d_{min} scalar diffraction theory is not valid due to multiple diffraction of radiation at the zone plate structure. Thus, for the nickel zone plate optimized for a wavelength of 3.37 nm in the first order, the spatial resolution can be no better than about 30 nm.

40.5 MANUFACTURE OF ZONE PLATES

Since the spatial resolution is determined primarily by the outer zone width, taking the discussion of the section “Volume Effects,” into account, zone plates must have small linewidths, along with large areas to provide large apertures and correct zone thicknesses to give optimum efficiencies. In addition, boundaries must be placed within about 1/3 of the outer zone width to maintain efficiencies and focusing properties.⁹ Electron-beam lithography (EBL), which routinely gives zone plates with diameters of around 200 μm and outer zone widths of 25 nm, is now the main method of manufacture but two other techniques—interference (holographic)¹⁰ and sputter and slice¹¹ methods—have historical significance.

In EBL the zone plate pattern is recorded in a polymer resist such as polymethyl methacrylate, followed by etching or electroplating to reproduce the pattern in, for example, nickel with a thickness of ~ 100 to 200 nm for the best efficiency at a few hundred electronvolts or gold or tungsten with thicknesses of ~ 0.5 to 1 μm for a few kiloelectronvolts.¹² Experimental efficiencies are lower than the theoretical optimum values due to manufacturing inaccuracies, primarily misplaced zone boundaries and profile errors.

40.6 BRAGG-FRESNEL LENSES

As discussed in the preceding sections of this chapter, x-ray zone plates work in transmission. However, like gratings they can also be used in reflection, and in this case the resolution-limiting effects of high aspect ratios can be alleviated.¹³ However, since near-normal incidence reflectivities are very small, to allow (near) circular symmetry to be maintained in-phase addition of many reflections is needed, as in crystals and multilayer mirrors. Optics which combine the Bragg reflection of crystals or multilayers with the Fresnel diffraction of gratings or zone plates are known as Bragg-Fresnel lenses.^{14,15} Their properties may be described by considering combinations of zone plates with multilayers or crystals; the generalisation to gratings is obvious.

Properties of Bragg-Fresnel Lenses

The diffraction pattern at a focus is determined as for an ordinary zone plate and the focused intensity is given by the diffraction efficiency combined with the Bragg reflectivity. Spherical waves from point sources S_1 and S_2 (Fig. 6) produce an elliptical interference pattern with S_1 and S_2 at the foci. A slice across the diffraction pattern, perpendicular to the line S_1S_2 , gives the structure of a circular transmission zone plate that focuses radiation emitted at S_1 to S_2 (Fig. 6a). If S_1 is moved to infinity then the interference pattern becomes parabolic and a standard zone plate is formed.

Taking the slice at an angle to the S_1S_2 axis produces an elliptical zone plate which forms a reflected image of S_1 at S_2 (Fig. 6b). The reflectivity is enhanced if the reflecting surface is a crystal or multilayer, with period d equal to the distance between the peaks of the interference pattern (Fig. 6c). Since the Bragg equation must be satisfied, the radiation is monochromatised with a bandpass $\Delta\lambda \sim \lambda/N_L$, where N_L is the number of layer pairs; the monochromaticity requirement of the zone plate, $\lambda/\Delta\lambda$ larger than the number of zones, must also be satisfied.

Defining the origin of the coordinate system to be at the center of lens, with the x and z axes parallel to the multilayer and the y axis perpendicular to the multilayers, the amplitude E of the reflected wave is

$$E(x, y) = r_M \sum_{l=1}^L \int_{Z_l} \exp\left[\frac{2\pi i}{\lambda}(R + r)\right] dr \quad (31)$$

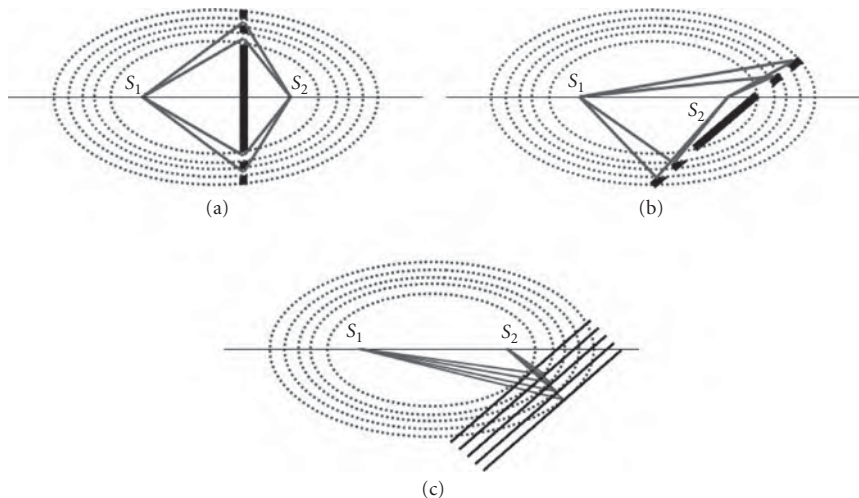


FIGURE 6 Structure of Bragg-Fresnel lenses. (See also color insert.)

where r_M is the peak amplitude reflectivity of the multilayer; the summation is over all layer pairs (l) and the integration is over the zone plate structure for each layer pair. If the source is far from the lens the distances R and r are given by

$$R = R_1 - x \frac{x_1}{R_1} - y \frac{y_1}{R_1} \quad r = r_2 - x \frac{x_2}{r_2} - y \frac{y_2}{r_2} + \frac{x^2}{2r_2} + \frac{y^2}{2r_2} \quad (32)$$

where $R_1 = (x_1^2 + y_1^2)^{1/2}$ is the distance from radiation source at $S_1(x_1, y_1)$ to the center of the lens, and $r_2 = (x_2^2 + y_2^2)^{1/2}$ is the distance from the center of the lens to the focal point $S_2(x_2, y_2)$. Since x varies along the multilayer surface and y varies into the multilayer, with $y = ld$ at the layer interfaces, x and y can be separated and the amplitude at the focal point is

$$E(S_2) = r_M \sum_{l=1}^L \exp \left\{ \frac{2\pi i}{\lambda} \left[-y \left(\frac{y_1}{R_1} + \frac{y_2}{r_2} \right) + \frac{y^2}{2r_2} \right] \right\} \times \int_{z_l} \exp \left\{ \frac{2\pi i}{\lambda} \left[-x \left(\frac{x_1}{R_1} + \frac{x_2}{r_2} \right) + \frac{x^2}{2r_2} \right] \right\} dx \quad (33)$$

The summation describes the wavelength-selecting properties of the multilayer and the integral describes the focusing property of the zone plate. With

$$P_l = \frac{2\pi}{\lambda} \left(\frac{l^2 d^2}{2r_2} - 2ld \sin \theta_0 \right) \quad (34)$$

where θ_0 is the incidence angle giving the maximum reflection at the center of the lens, the summation reduces to

$$G = \sum_{l=1}^L \exp \{iP_l\} \quad (35)$$

and the angular distribution of the reflected radiation is given by

$$\frac{1}{L^2} |G|^2 = \frac{1}{L^2} \left[\left(\sum_{l=1}^L \sin P_l \right)^2 + \left(\sum_{l=1}^L \cos P_l \right)^2 \right] \quad (36)$$

Manufacture of Bragg-Fresnel Lenses

Bragg-Fresnel lenses may be made by masking the surface of a multilayer mirror with an absorbing zone plate or by etching a zone plate pattern into the multilayer.¹⁶ Similar methods can be used for crystal based Bragg-Fresnel lenses.¹⁷ In order to obtain high efficiencies, phase-modulating effects can be used to enhance the efficiency of the zone plate part of the lens. This requires, for example, profiling the multilayer or depositing it on an anisotropically etched substrate.

40.7 REFERENCES

1. A. G. Michette, *Optical Systems for Soft X-Rays*, New York: Plenum, pp. 170–176 (1986).
2. A. G. Michette, *Optical Systems for Soft X-Rays*, New York: Plenum, pp. 178–179 (1986).
3. B. Niemann, D. Rudolph, and G. Schmahl, "Soft X-Ray Imaging Zone Plates with Large Zone Numbers for Microscopic and Spectroscopic Applications," *Opt. Comm.* **12**:160–163 (1974).
4. J. Kirz, "Phase Zone Plates for X-Rays and the Extreme UV," *J. Opt. Soc. Am.* **64**:301–309 (1974).

5. R. O. Tatchyn, "Optimum Zone Plate Theory and Design," in *X-Ray Microscopy, Springer Series in Optical Sciences*, Heidelberg: Springer, **43**:40–50 (1984).
6. E. Di Fabrizio, and M. Gentili, "X-Ray Multilevel Zone Plate Fabrication by Means of Electron-Beam Lithography: Toward High-Efficiency Performances," *J. Vac. Sci. Technol. B* **17**:3439–3443 (1999).
7. A. I. Erko, V. V. Aristov, and B. Vidal, *Diffraction X-Ray Optics*, Bristol: IOP Publishing, pp. 98–101 (1996).
8. R. J. Collier, Ch. B. Burckhardt, and L. H. Lin, *Optical Holography*, New York & London: Academic Press, (1971).
9. M. J. Simpson and A. G. Michette, "The Effects of Manufacturing Inaccuracies on the Imaging Properties of Fresnel Zone Plates," *Optica Acta* **30**:1455–1462 (1983).
10. P. Guttman, "Construction of a Micro Zone Plate and Evaluation of Imaging Properties," in *X-Ray Microscopy, Springer Series in Optical Sciences*, Heidelberg: Springer, **43**:75–90 (1984).
11. D. Rudolph, B. Niemann, and G. Schmahl, "High Resolution X-Ray Optics," *Proc. SPIE* **316**:103–105 (1982).
12. P. Charalambous, "Fabrication and Characterization of Tungsten Zone Plates for Multi KeV X-Rays," *AIP Conf. Proc.* **507**:625–630 (2000).
13. A. G. Michette, S. J. Pfauntsch, A. Erko, A. Firsov, and A. Svintsov, "Nanometer Focusing of X-Rays with Modified Reflection Zone Plates," *Opt. Commun.* **245**:249–253 (2005).
14. V. V. Aristov, A. I. Erko, and V. V. Martynov, "Principles of Bragg-Fresnel Multilayer Optics," *Rev. Phys. Appl.* **23**:1623–1630 (1988).
15. A. Erko, Y. Agafonov, La. Panchenko, A. Yakshin, P. Chevallier, P. Dhez, and F. Legrand, "Elliptical Multilayer Bragg-Fresnel Lenses with Submicron Spatial Resolution for X-Rays," *Opt. Commun.* **106**:146–150 (1994).
16. A. I. Erko, La. Panchenko, A. A. Firsov, and V. I. Zinenko, "Fabrication and Tests of Multilayer Bragg-Fresnel X-Ray Lenses," *Microelectron Eng.* **13**:335–338 (1991).
17. A. Firsov, A. Svintsov, A. Erko, W. Gudat, A. Asryan, M. Ferstl, S. Shapoval, and V. Aristov, "Crystal-Based Diffraction Focusing Elements for Third-Generation Synchrotron Radiation Sources," *Nucl. Instrum. Methods Phys. Res. A* **467–468**:366–369 (2001).

This page intentionally left blank

MULTILAYERS

Eberhard Spiller

*Spiller X-Ray Optics
Livermore, California*

41.1 GLOSSARY

$\tilde{n} = 1 - \delta + i\beta$	refractive index
θ_i	grazing angle of propagation in layer i
φ_i	phase at boundary i
λ	wavelength
q	x-ray wavevector perpendicular to the surface
d	layer thickness
Λ	multilayer period, $\Lambda = d_1 + d_2$
D	total thickness of the multilayer
$f, 1/f$	spatial frequency and spatial period along the surface or boundary
r_{nm}, t_{nm}	amplitude reflection and transmission coefficients
PSD	2-dimensional power spectral density
$a(f)$	roughness replication factor

41.2 INTRODUCTION

The reflectivity of all mirror materials is small beyond the critical grazing angle, and multilayer coatings are used to enhance this small reflectivity by adding the amplitudes reflected from many boundaries coherently, as shown in Fig. 1. Multilayers for the VUV and x-ray region can be seen as an extension of optical coatings toward shorter wavelengths or as artificial one-dimensional Bragg crystals with larger lattice spacings Λ than the d -spacings of natural crystals (see Chap. 39). In contrast to the visible region, no absorption-free materials are available for wavelengths $\lambda < 110$ nm. In addition, the refractive indices of all materials are very close to one, resulting in a small reflectance at each boundary and requiring a large number of boundaries to obtain substantial reflectivity. For absorption-free materials a reflectivity close to 100 percent can always be obtained, independent of the reflectivity r of an individual boundary by making the number of boundaries N sufficiently

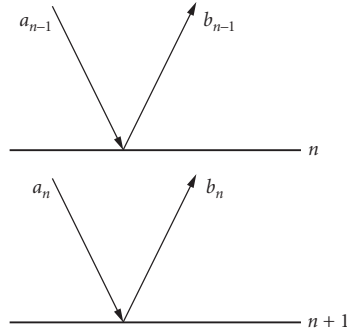


FIGURE 1 Two layers and their boundaries in a multilayer structure with the incoming wave amplitudes a_n and reflected amplitudes b_n in each layer.

large, $Nr \gg 1$. Absorption limits the number of periods that can contribute to the reflectivity in a multilayer to

$$N_{\max} = \frac{\sin^2 \theta}{2\pi\beta} \quad (1)$$

where β is the average absorption index of the coating materials. Multilayers became useful for $N_{\max} \gg 1$, and very high reflectivities can be obtained if $N_{\max}$ is much larger than $N_{\min} = 1/r$, the number required for a substantial reflectivity enhancement.

The absorption index β is in the order of 1 for wavelengths around 100 nm and decreases to very small values at shorter wavelengths in the x-ray range. Materials that satisfy the condition $N_{\max} > 1$ become available for $\lambda < 50$ nm. For $\lambda < 20$ nm, in wavelength regions not too close to absorption edges, β decreases rapidly with decreasing wavelength; $\beta \propto \lambda^3$. The reflected amplitude r from a single boundary also decreases with wavelength, albeit at a slower rate as $r \propto \lambda^2$, and one can compensate for this decrease by increasing the number of boundaries or periods N in a multilayer stack, by using $N \propto 1/\lambda^2$. With this method reflectivities close to 100 percent are theoretically possible at very short wavelengths in the hard x-ray range. A perfect crystal showing Bragg reflection can be seen as a multilayer that realizes this high reflectivity, with the atomic planes located at the nodes of the standing wave within the crystal.

The condition that all periods of a multilayer reflect in phase leads to the Bragg condition for the multilayer period Λ ,

$$m\lambda = 2\Lambda \sin \theta \Rightarrow \Lambda = \frac{m\lambda}{2\sin \theta} \approx \frac{m\lambda}{2\sin \theta_0 \sqrt{1 - 2\delta/\sin^2 \theta_0}} \quad (2)$$

where θ is the effective grazing angle of propagation within the multilayer material, and θ_0 the corresponding angle in vacuum. The refraction correction represented by the square root in Eq. (2) becomes large for small grazing angles, even for small values of δ . The path difference between the amplitude maxima from adjacent periods is $m\lambda$, and multilayers are most of the time used in order $m = 1$. The shortest period Λ for good multilayer performance is limited by the quality of the boundaries (the roughness σ should be smaller than $\Lambda/10$) and this quality is in turn limited by the size of the atoms for noncrystalline materials. Practical values for the shortest period are around $\Lambda = 1.5$ nm. Thus the high reflectivities that are theoretically possible for very hard x rays can only be realized with multilayers of periods $\Lambda > 1.5$ nm, which must be used at small grazing angles. By introducing the momentum transfer q at reflection,

$$q_i = \frac{4\pi}{\lambda} \tilde{n} \sin \theta_i \quad (3)$$

We can express the condition $\Lambda > 1.5 \text{ nm}$ as $|q| < 4 \text{ nm}^{-1}$. It is convenient to introduce the variable q in x-ray optics because, as long as the x-ray wavelength is not too close to an absorption edge, the performance of optical components is determined mainly by q and not by the specific values of λ and θ .

Attempts to observe x-ray interference from thin films started around 1920; by 1931 Kiessig¹ observed and analyzed the x-ray interference structure from Ni films. Multilayer structures produced around 1940² lost their x-ray reflectivity due to diffusion in the structure, and this fact became a tool to measure small diffusion coefficients.^{3–5} The usefulness of multilayers for normal incidence optics in the XUV and EUV was recognized in 1972.⁶ The deposition processes to produce these structures were developed by many groups in the next two decades. Today multilayer telescopes on the orbiting observatories SOHO and TRACE provide high-resolution EUV pictures of the solar corona (see <http://umbra.nascom.nasa.gov/eit/> and <http://vestige.lmsal.com/TRACE/>). Cameras with multilayer coated mirrors for the EUV are a main contender for the fabrication of the next generation of computer chips, and multilayer mirrors are found at the beamlines of all synchrotron radiation facilities and in x-ray diffraction equipment as beam deflectors, collimators, filters, monochromators, polarizers, and imaging optics. (See Ref. 7 and Chaps. 28, 43 to 46, and 54 in this volume for more details.)

41.3 CALCULATION OF MULTILAYER PROPERTIES

The theoretical treatment of the propagation of x rays in layered structures does not differ from that for other wavelength regions as discussed by Dobrowolski in Chap. 7 in Vol. IV of this *Handbook* and in many textbooks and review articles. At the boundary of two materials or at an atomic plane an incident wave is split into a transmitted and reflected part and the total field amplitude in the structure is the superposition of all these waves. Figure 1 sketches two layers as part of a multilayer and the amplitudes of the forward a_n and backward b_n running waves in each layer. The amplitudes in each layer are coupled to those of the adjacent layers by linear equations that contain the amplitude transmission $t_{n,m}$ and reflection coefficients r_{nm} ,

$$\begin{aligned} a_n &= a_{n-1} t_{n-1,n} e^{i\varphi_n} + b_n e^{2i\varphi_n} r_{n,n-1} \\ b_n &= a_n r_{n,n+1} + b_{n+1} e^{i\varphi_{n+1}} t_{n+1,n} \end{aligned} \quad (4)$$

The phase delay φ_n due to propagation through layer n of thickness d and angle θ_n from grazing incidence is given by

$$\varphi_n = \frac{2\pi}{\lambda} \tilde{n} d_n \sin \theta_n \quad (5)$$

and the transmitted and reflected amplitudes at each boundary are obtained from the Fresnel equations

$$r_{n,n+1} = \frac{q_n - q_{n+1}}{q_n + q_{n+1}} \quad (6)$$

$$t_{n,n+1} = \frac{2q_n}{q_n + q_{n+1}} \quad (7)$$

with the q values as defined in Eq. (3) for s-polarization and $q_i = (4\pi/\lambda \tilde{n}_i) \sin \theta_i$ for p-polarization.

Matrix methods^{8,9} are convenient to calculate multilayer performance using high-level software packages that contain complex matrix manipulation. The transfer of the amplitudes over a boundary and through a film is described by⁹

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} = \frac{1}{t_{i-1,i}} \begin{pmatrix} e^{i\varphi_i} & r_{i-1,i} e^{-i\varphi_i} \\ r_{i-1,i} e^{i\varphi_i} & e^{-i\varphi_i} \end{pmatrix} \begin{pmatrix} a_{i+1} \\ b_{i+1} \end{pmatrix} \quad (8)$$

and the transfer between the incident medium ($i = 0$) and the substrate ($i = n + 1$) by

$$\begin{pmatrix} a_0 \\ b_0 \end{pmatrix} = \frac{\prod_{i=1}^{i=n+1} M_i}{\prod_{i=1}^{i=n+1} t_{i,i+1}} \begin{pmatrix} a_{n+1} \\ b_{n+1} \end{pmatrix} \quad (9)$$

with the matrices M_i as defined in Eq. (8). For incident radiation from the top ($b_{n+1} = 0$) we can calculate the reflected and transmitted amplitudes from the elements of the product matrix m_{ij} as

$$\begin{aligned} r_{\text{ML}} &= b_0/a_0 = m_{21}/m_{11} \\ t_{\text{ML}} &= a_{n+1}/a_0 = (\prod t_{i,i+1})/m_{11} \end{aligned} \quad (10)$$

The reflected intensity is $R_{\text{ML}} = r_{\text{ML}} r_{\text{ML}}^*$ and the transmitted intensity $T_{\text{ML}} = t_{\text{ML}} t_{\text{ML}}^* (q_{n+1}/q_0)$ with the q values of Eq. (3).

In another convenient matrix formalism due to Abeles each matrix contains only the parameters of a single film;^{8,10} however, the transfer matrix method given earlier⁹ is more convenient when one wants to include the imperfections of the boundaries.

The achievable boundary quality is always of great concern in the fabrication of x-ray multilayers. The first effect of boundary roughness or diffusion is the reduction of the reflected amplitude due to dephasing of the contributions from different depth within the boundary layer. If one describes the reflected amplitude $r(z)$ as a function of depth by a Gaussian of width σ , one obtains a Debye-Waller factor for the reduction of boundary reflectivity,

$$r/r_o = \exp(-0.5 q_1 q_2 \sigma^2) \quad (11)$$

where r_o is the amplitude reflectivity for a perfect boundary and the q values are those of the films on either side of the boundary.^{7,11-14} Reducing the amplitude reflection coefficients in Eq. (8) by the Debye-Waller factor of Eq. (11) gives a good estimate of the influence of boundary roughness on multilayer performance. The roughness σ has by most authors been used as a fitting parameter that characterizes a coating. Usually the Debye-Waller factor is given for the intensity ratio [absolute value of Eq. (11) squared] with the vacuum q -values for both materials (refraction neglected). The phase shift in Eq. (11) due to the imaginary part of q produces a shift of the effective boundary toward the less absorbing material.

For a boundary without scattering (gradual transition of the optical constants or roughness at very high spatial frequencies) the reduction in reflectivity is connected with an increase in transmission according to $t_{12} t_{21} + r_{12}^2 = 1$. Even if the reflectivity is reduced by a substantial factor, the change in transmission can in most cases be neglected, because reflectivities of single boundaries are typically in the 10^{-4} range and transmissions close to one. Roughness that scatters a substantial amount of radiation away from the specular beam can decrease both the reflection and transmission coefficients, and the power spectral density (PSD) of the surface roughness over all relevant spatial frequencies has to be measured to quantify this scattering.

It is straightforward to translate Eqs. (4) to (11) into a computer program, and a personal computer gives typically a reflectivity curve of a 100 layer coating for 100 wavelengths or angles within a few seconds. Multilayer programs and the optical constants of all elements and of many compounds can be accessed on (<http://www-cxro.lbl.gov>) or downloaded¹⁵ (www.rxcollc.com/idl) (see Chap. 36). The first site also has links to other relevant sites, and the Windt programs can also calculate nonspecular scattering.

41.4 FABRICATION METHODS AND PERFORMANCE

Multilayer x-ray mirrors have been produced by practically any deposition method. Thickness errors and boundary roughness are the most important parameters; they have to be kept smaller than $\Lambda/10$ for good performance. Magnetron sputtering, pioneered by Barbee,¹⁶ is most widely used: sputtering systems are very stable, and one can obtain the required thickness control simply by timing. The same is true for Ion beam deposition systems.¹⁷ Thermal evaporation systems usually use an in situ soft x-ray

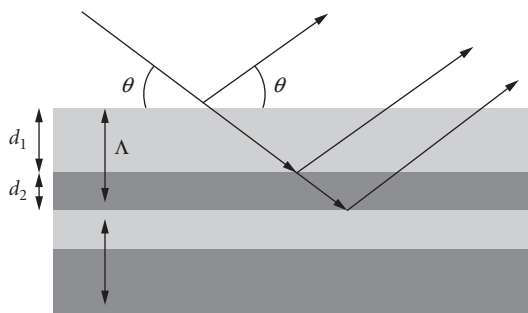


FIGURE 2 A quasiperiodic multilayer. The individual layer thicknesses change while the sum Λ remains constant.

reflectometer to control the thickness and an additional ion gun for smoothing of the boundaries.^{18–21} In the “quarter wave stack” of multilayer mirrors for the visible, all boundaries add their amplitudes in-phase to the total reflectivity, and each layer has the same optical thickness and extends between a node and an antinode of the standing wave generated by the incident and reflected waves. To mitigate the effect of absorption in the VUV and x-ray region one first selects a “spacer” material with the lowest available absorption and combines it with a “reflector” layer with good contrast with the optical constants of the spacer. The optimum design minimizes absorption by reducing the thickness of the reflector layer and by attempting to position it close to the nodes of the standing wave field, while for the spacer layer the thickness is increased and it is centered around the antinodes. The design accepts some dephasing of the contributions from adjacent boundaries and the optimum is a compromise between the effects of this dephasing and the reduction in absorption. The best values for $\gamma = d_h/\Lambda$ are between 0.3 and 0.4.^{7,22,23} Optimum multilayers for longer x-ray wavelengths ($\lambda > 20$ nm) are quasiperiodic, as shown in Fig. 2, with decreasing γ from the bottom to the top of a multilayer stack. One always attempts to locate the stronger absorber close to the nodes of the standing wavefield within the coating to reduce absorption, and the absorption reduction is greater near the top of the stack where the standing wave between incident and reflected radiation has more contrast.⁶

The paper by Rosenbluth²³ gives a compilation of the best multilayer materials for each photon energy in the range from 100 to 2000 eV. Absorption is the main performance limit for multilayers of larger period $\Lambda > 80$ nm used at longer wavelength near normal incidence. In this region it is important to select the material with the smallest absorption as the first component of a structure (usually a light element at the long wavelength side of an absorption edge) and the absorption of this material becomes the main limit for the multilayer performance. A list of materials of low absorption at selected wavelengths with their values for β and $N_{\max}$ is given in Table 1.

Boundary Quality

The quality of the boundary is the main limitation for the performance of a multilayer with short periods. Roughness of a boundary scatters radiation away from the specular beam and reduces both reflectivity and transmission at each boundary. Diffusion of the two materials at a boundary (or roughness at very high spatial frequencies that scatter into evanescent waves) also reduces the reflectivity but can increase the transmission. Deposition at high energy of the incident materials enhances relaxation of atoms of the growing surface and reduces roughness but also increases diffusion at the boundary. For each set of coating materials the deposition energy for the best compromise has to be found. A good solution is to separate the two problems, depositing each film at low energy to minimize diffusion and then polishing the top of each layer after the deposition. Examples are thermal deposition of each layer and ion polishing after deposition,²⁴ or low-energy deposition at the start of each layer and higher energy near the top.²⁵ Polishing of thin films by an ion beam has also been used to remove defects from the substrates of masks in EUV lithography. In a Mo/Si multilayer one can either deposit a thicker Si film than

TABLE 1 Absorption Index β and $N_{\max}$ for the Largest q Values (Normal Incidence for $\lambda > 3.15$ nm and $\sin \theta = \lambda/\pi$ for $\lambda < 3.14$ nm) of Good Spacer Materials near Their Absorption Edges and Absorption of Some Materials at $\lambda = 0.154$ nm

	λ_{edge} (nm)	β	$N_{\max}$
Mg-L	25.1	7.5×10^{-3}	21
Al-L	17.1	4.2×10^{-3}	38
Si-L	12.3	1.6×10^{-3}	99
Be-K	11.1	1.0×10^{-3}	155
Y-M	8.0	3.5×10^{-3}	45
B-K	6.6	4.1×10^{-4}	390
C-K	4.37	1.9×10^{-4}	850
Ti-L	3.14	4.9×10^{-4}	327
N-K	3.1	4.4×10^{-5}	3,580
Sc-L	3.19	2.9×10^{-4}	557
V-L	2.43	3.4×10^{-4}	280
O-K	2.33	2.2×10^{-5}	3,980
Mg-K	0.99	6.6×10^{-6}	2,395
Al-K	0.795	6.5×10^{-6}	1,568
Si-K	0.674	4.2×10^{-6}	1,744
SiC	0.674	6.2×10^{-6}	1,182
TiN	3.15	4.9×10^{-4}	327
Mg ₂ Si	25.1	7.4×10^{-3}	21
Mg ₂ Si	0.99	6.8×10^{-6}	2,324
Be	0.154	2.0×10^{-9}	189,000
B	0.154	5.7×10^{-9}	67,450
C	0.154	1.2×10^{-8}	32,970
Si	0.154	1.7×10^{-7}	5,239
Ni	0.154	5.1×10^{-7}	750
W	0.154	3.9×10^{-6}	98

needed and etch the excess thickness away with the ion beam or use a more aggressive deposition/etch process on the Si substrate before the multilayer reflecting coating is applied.²⁶

Most good multilayer systems can be described by a simple growth model: particles or atoms arrive randomly and can relax sideways to find locations of lower energy. The random deposition produces a flat power spectrum at low frequencies for the surface roughness with most of the roughness at very high spatial frequencies. The relaxation then reduces roughness at the highest spatial frequencies.

Roughness is characterized by a power spectral density (PSD) such that the intensity of the scattered light is proportional to the PSD. Consistent values for the PSD can be obtained by atomic force microscopy and from scatter measurements.^{27–29} The roughness height σ from the spatial frequency range from f_1 to f_2 is related to the PSD by

$$\sigma^2 = 2\pi \int_{f_1}^{f_2} \text{PSD}(f) f df \quad (12)$$

We assume that the surface is isotropic in f and the roughness in Eq. (12) is obtained by integrating over rings with area $2\pi f df$. During the development phase of multilayer x-ray mirrors it was fortuitous that practically perfect smooth substrates were available as Si wafers, float glass, and mirrors fabricated for laser gyros. The 2-dimensional power spectral density (see Chap. 8 by Eugene L. Church and Peter Z. Takacs in Vol. I and Chap. 44 of this volume) of a film on a perfectly smooth substrate is given by^{7,30–32}

$$\text{PSD}(q_s, d) = \Omega \frac{1 - \exp(-2l_r^{n-1} dq_s^n)}{2l_r^{n-1} q_s^n} \quad (13)$$

TABLE 2 Growth Parameters of Multilayer Systems*

System	Λ (nm)	N	Ω (nm ³)	l_r (nm)	n	d (nm)	σ (nm)
Co/C	3.2	150	0.016	1.44	4	480	0.14
Co/C	2.9	144	0.016	1.71	4	423	0.12
Co/C	2.4	144	0.016	1.14	4	342	0.15
Co/C	3.2	85	0.016	1.71	4	274	0.19
Mo/Si	7.2	24	0.035	1.71	4	172	0.14
Ni/C (a)	5.0	30	0.035	1.71	4	150	0.14
W/B ₄ C (b)	1.22	350	0.01	1.22	4	427	0.10
W/B ₄ C (b)	1.78	255	0.018	1.22	3	454	0.12
Mo/Si (c)	6.8	40	0.035	1.36	4	280	0.19
Mo/Si (d)	6.8	40	0.055	1.20	2	280	0.12
Si in Mo/Si (c)	6.8	40	0.02	1.36	4	280	0.14
Mo in Mo/Si (c)	6.8	40	0.5	1.36	4	280	0.23

*Multilayer period Λ , number of periods N , growth parameters Ω , l_r , and n , total thickness d , and roughness σ of the multilayer film calculated for a perfectly smooth substrate within the spatial frequency range $f = 0.001 - 0.25 \text{ nm}^{-1}$.

Coating (a) produced by dual ion beam sputtering courtesy of J. Pedulla (NIST), (b) of Y. Platonov (Osmic). Data for Mo/Si produced by magnetron sputtering (c) from Ref. 36, and ion beam sputtering (d) from P. Kearney and D. Stearns (LLNL). Parameters are the average value for both coating materials, for sputtered Mo/Si we give also the parameters for Mo and Si separately.

where Ω is the particle volume, l_r is a relaxation length, d is the total thickness of the coating, and f a spatial frequency on the surface. The parameter n characterizes the relaxation process; $n = 1$ indicates viscous flow, 2 condensation and re-evaporation, 3 bulk diffusion, and 4 surface diffusion. Equation (13) yields the flat PSD of the random deposition with roughness $\sigma \propto \sqrt{d}$ for small spatial frequencies and a power law with exponent n that is independent of thickness at high spatial frequencies. Roughness is replicated throughout a stack with a replication factor

$$a(f) = \exp - (l_r^{n-1} d q_s^n) \quad (14)$$

so that the PSD of a film on a substrate has a power spectral density PSD^{tot}

$$\text{PSD}^{\text{tot}} = \text{PSD}^{\text{film}} + a^2 \text{PSD}^{\text{sub}} \quad (15)$$

The growth parameters of some multilayer structures are given in Table 2.³³ Note that the roughness values in the last column do not include contributions from spatial frequencies $f > 0.25 \text{ nm}^{-1}$ and from diffuse transition layers. These values have to be added in the form $\sigma_{\text{tot}}^2 = \sigma_1^2 + \sigma_2^2$ in Eq. (11) for the calculation of the reflectivity.

High-Reflectivity Mirrors

A compilation of the reflectivities of multilayers obtained by different groups can be found at www-cxro.lbl.gov/multilayer/survey.html. The highest normal incidence reflectivities, around 70 percent, have been obtained with Mo/Be and Mo/Si wavelengths around $\lambda = 11.3$ and 13 nm, near the absorption edges of Be and Si.³⁴ The peak reflectivity drops at longer wavelengths due to the increased absorption. At shorter wavelengths, roughness becomes important and reduces the reflectivity of normal incidence mirrors. One can, however, still obtain very high reflectivity at short wavelengths by keeping the multilayer period above 2 nm and using grazing angles of incidence to reach short wavelengths.

For hard x rays, where N_{max} is very much larger than the number, $N_{\text{min}} \sim 1/r$ needed to obtain good reflectivity, one can position multilayers with different periods on top of each other (depth-graded multilayers, as shown in Fig. 3) to produce a coating with a large spectral or angular bandwidth. Such “supermirrors” are being proposed to extend the range of grazing incidence telescopes up to the 100 keV.^{35,36} “Supermirrors” are common for cold neutrons where absorption-free materials are

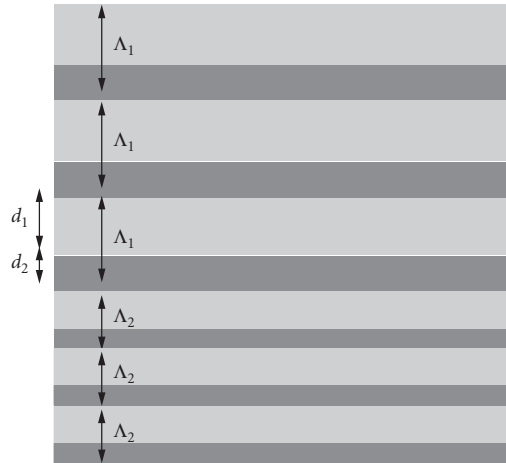


FIGURE 3 A depth-graded multilayer. The repeat distance Λ changes with depth. (The change can be abrupt, as shown, or more gradual.)

available.³⁷ The low absorption for hard x rays makes it also possible to produce coatings with very narrow bandwidth. The spectral width is determined by the effective number of periods contributing to the reflectivity, $\lambda/\Delta\lambda = N_{\text{eff}}$. Reducing the reflectivity from a single period, for example, by using a very small value of $\gamma = d_H/\Lambda$, allows radiation to penetrate deeper into the stack, allowing more layers to contribute, thus reducing the bandwidth.⁷

Multilayer Coated Optics

Multilayers for imaging optics usually require a lateral grading of the multilayer period Λ across the face of the optic to adapt to the varying angle of incidence according to Eq. (2), as shown in Fig. 4. Many methods have been used to produce the proper grading of the multilayer period during deposition, among them shadow masks in front of the rotating optics, substrate tilt, speed control of the platter that moves the optics over the magnetrons, and computer-controlled shutters.^{38–44} The reproducibility that has been achieved in EUV lithography from run to run is better than 0.1 percent. Multilayer mirrors

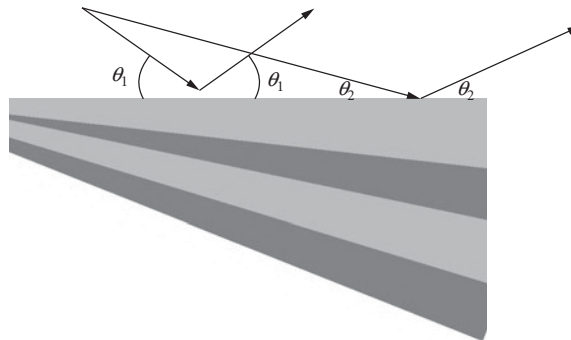


FIGURE 4 Laterally graded multilayer.

with a linear grading of the thickness on parabolically bent Si wafers are commercially available (e.g. www.rigaku.com/optics/index.html) and are being used as collimators in x-ray diffraction experiments at $\lambda = 0.154$ nm. At this wavelength the reflectivity is over 75 percent and the flux to the diffractometer can be increased by about an order of magnitude.⁴⁶

It is still a challenge to produce figured substrates that have both good figure and finish in the 0.1-nm range.⁴⁷ Multilayer mirrors can remove high spatial roughness with spatial periods of less than 50 nm and graded coatings can be used to correct low-order figure errors; however, mirrors that do not meet specifications for high resolution imaging usually have considerable roughness at spatial frequencies between these limits that cannot be modified by thin films.³³

Diffraction-limited performance of multilayer coated mirrors has been achieved in the cameras for EUV lithography. The mirrors of these cameras usually require lateral grading of the layer thickness due to the change of the angle of incidence over the surface of a mirror. These graded coatings modify the shape of the mirror by two components: the total thickness of the multilayer and the shift in the phase of the reflected wave. The two effects have opposite sign in their sensitivity to thickness errors, and for the mirrors used in EUV lithography around $\lambda = 13.5$ nm the total error in the figure is around 75 percent of the error produced by an error in the multilayer thickness alone. After subtracting the changes that can be compensated by alignment the remaining figure error added by the coating is well below 0.1 nm.^{48,49}

Polarizers and Phase Retarders

The Brewster angle in the VUV and x-ray region occurs close to 45° for all materials, so all reflectors used near 45° are effective polarizers. Multilayer coatings do not change the ratio of the reflectivities for s- and p-polarization, but enhance the reflectivity for s-polarization to useful values.^{50,51} The reflectivity for p-polarization at the Brewster angle is zero for absorption-free materials but increases with absorption. Therefore the achievable degree of polarization is higher at shorter wavelengths where absorption is lower. Typical values for the reflectivity ratio R_s/R_p are 10 around $\lambda = 30$ nm and over 1000 around $\lambda = 5$ nm.

It is not possible to design effective 90° phase retarding multilayer reflectors with high reflectivity for both polarizations. The narrower bandwidth of the reflectivity curve for p-polarization allows one to produce a phase delay at incidence angles or wavelengths that are within the high reflectivity band for s- but not for p-polarization; however, because of the greatly reduced p-reflectivity of such a design, one cannot use them to transform linear polarized into circular polarized radiation.⁵² Multilayers used in transmission offer a better solution. Near the Brewster angle p-polarized radiation is transmitted through a multilayer structure without being reflected by the internal boundaries and has a transmission determined mainly by the absorption in the multilayer stack. A high transmission for s-polarization is also obtained from a multilayer stack that is used off-resonance near a reflectivity minimum or transmission maximum. However, the transmitted radiation is delayed due to the internal reflection within the multilayer stack. Calculated designs for wavelength of 13.4 nm^{53–56} produce a phase retardation of 80° at a grazing angle of 50° . One can tune such a phase retarder in wavelength by producing a graded coating or by using different incidence angles.⁵⁷ A 90° phase retarder is possible with high-quality multilayers that can obtain peak reflectivities above 70 percent. Boundary roughness reduces the phase retardation because it reduces the effective number of bounces within the structure. The maximum phase retardation measured at x-ray wavelengths $\lambda < 5$ nm are in the 10° range.⁵⁸

41.5 MULTILAYERS FOR DIFFRACTIVE IMAGING

The development of free electron lasers (FELs) for x rays promises imaging at high resolution of general specimens; the specimens do not have to be crystallized. An image of the specimen is reconstructed from diffraction patterns produced by powerful, short coherent pulses. The specimen is destroyed

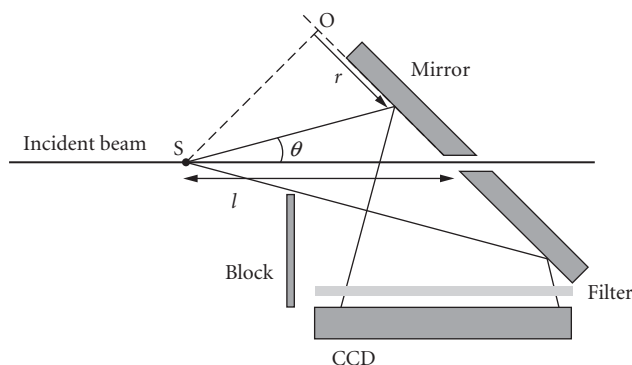


FIGURE 5 Camera used for diffraction imaging. The FEL pulse illuminates the sample S. Radiation diffracted by the sample S is reflected by the multilayer mirror to the CCD detector while the direct beam passes a hole in the mirror. (From Ref. 60.)

by the radiation, but the diffraction pattern recorded during the short pulse represents the specimen before it explodes. Thousands of such diffraction patterns from identical specimens at different rotation angles are needed for a high-resolution 3-D reconstruction of the specimen. Multilayer mirrors have been used to safely direct the diffracted radiation to the detector, transmitting the direct beam through a hole in the center and suppressing the light produced by the exploding specimen (Fig. 5). The thickness of the layers in the multilayer mirror is graded in such a way as to reflect only the wavelength of the incident beam at the correct angle to the detector. First experiments by Chapman et al.^{59,60} have demonstrated the principle using EUV radiation, but considerable challenges remain to transfer the technique to the x-ray region and to interesting 3-D specimen. (See Chap. 27.)

41.6 REFERENCES

1. H. Kiessig, "Interferenz von Röntgenstrahlen an dünnen Schichten," *Ann. der Physik* 5. Folge **10**:769–788 (1931).
2. J. DuMond and J. P. Joutz, "An X-Ray Method of Determining Rates of Diffusion in the Solid State," *J. Appl. Phys.* **11**:357–365 (1940).
3. H. E. Cook and J. E. Hilliard, "Effect of Gradient Energy on Diffusion in Gold–Silver Alloys," *J. Appl. Phys.* **40**:2191–2198 (1969).
4. J. B. Dinklage, "X-Ray Diffraction by Multilayered Thin-Film Structures and Their Diffusion," *J. Appl. Phys.* **38**:3781–3785 (1967).
5. A. L. Greer and F. Spaepen, "Diffusion," in *Modulated Structures*, edited by L. Chang and B. C. Giess (Academic Press, New York, 1985), p. 419.
6. E. Spiller, "Low-Loss Reflection Coatings Using Absorbing Materials," *Appl. Phys. Lett.* **20**:365–367 (1972).
7. E. Spiller, *Soft X-Ray Optics* (SPIE Optical Engineering Press, Bellingham, WA, 1994).
8. F. Abelès, "Recherches sur la propagation des ondes électromagnétique inusoidales dans les milieux stratifiés. Application aux couches minces," *Ann. de Physique* **5**:596–639 (1950).
9. O. S. Heavens, *Optical Properties of Thin Solid Films* (Dover, New York, 1966).
10. M. Born and E. Wolf, *Principles of Optics*, 5th ed. (Pergamon Press, Oxford, 1975).
11. P. Croce, L. Nénot, and B. Pardo, "Contribution à l'étude des couches mince par réflexion spéculaire de rayon X," *Nouv. Rev. d'Optique appliquée* **3**:37–50 (1972).

12. P. Croce and L. Nénot, "Étude des couches mince et des surfaces par réflexion rasante, spéculaire ou diffuse, de rayon X," *J. De Physique Appliquée* **11**:113–125 (1976).
13. L. Nénot and P. Croce, "Caractérisation des surfaces par réflexion rasante de rayon X, Application à étude du polissage de quelques verres silicates," *Revue Phys. Appl.* **15**:761–779 (1980).
14. F. Stanglmeier, B. Lengeler, and W. Weber, "Determination of the Dispersive Correction $f'(E)$ to the Atomic Form Factor from X-Ray Reflection," *Acta Cryst.* **A48**:626–639 (1992).
15. D. Windt, "IMD—Software for Modeling the Optical Properties of Multilayer Films," *Computers in Physics* **12** (4):360–370 (1998).
16. T. W. Barbee, Jr., "Multilayers for X-Ray Optics," *Opt. Eng.* **25**:893–915 (1986).
17. E. Spiller, S. Baker, P. Mirkarimi, et al., "High Performance Mo/Si Multilayer Coatings for EUV Lithography Using Ion Beam Deposition," *Appl. Opt.* **42**:4049–4058 (2003).
18. E. Spiller, A. Segmüller, J. Rife, et al., "Controlled Fabrication of Multilayer Soft X-Ray Mirrors," *Appl. Phys. Lett.* **37**:1048–1050 (1980).
19. E. Spiller, "Enhancement of the Reflectivity of Multilayer X-Ray Mirrors by Ion Polishing," *Opt. Eng.* **29**:609–613 (1990).
20. E. J. Puik, M. J. van der Wiel, H. Zeijlemaker, et al., "Ion Bombardment of X-Ray Multilayer Coatings: Comparison of Ion Etching and Ion Assisted Deposition," *Appl. Surface Science* **47**:251–260 (1991).
21. E. J. Puik, M. J. van der Wiel, H. Zeijlemaker, et al., "Ion Etching of Thin W Layers: Enhanced Reflectivity of W-C Multilayer Coatings," *Appl. Surface Science* **47**:63–76 (1991).
22. A. V. Vinogradov and B. Ya. Zel'dovich, "X-Ray and Far UV Multilayer Mirrors; Principles and Possibilities," *Appl. Optics* **16**:89–93 (1977).
23. A. E. Rosenbluth, "Computer Search for Layer Materials that Maximize the Reflectivity of X-Ray Multilayers," *Revue Phys. Appl.* **23**:1599–1621 (1988).
24. E. Spiller, "Smoothing of Multilayer X-Ray Mirrors by Ion Polishing," *Appl. Phys. Lett.* **54**:2293–2295 (1989).
25. Fredrik Eriksson, Goeran A. Johansson, Hans M. Hertz, et al., "Enhanced Soft X-Ray Reflectivity of Cr/Sc Multilayers by Ion-Assisted Sputter Deposition," *Opt. Eng.* **41** (11):2903–2909 (2002).
26. P. B. Mirkarimi, E. Spiller, S. L. Baker, et al., "A Silicon-Based, Sequential Coat- and- Etch Process to Fabricate Nearly Perfect Substrate Surfaces," *J. Nanoscience and Nanotechnology* **6**:28–35 (2006).
27. E. M. Gullikson, D. G. Stearns, D. P. Gaines, et al., "Non-Specular Scattering from Multilayer Mirrors at Normal Incidence," presented at the *Grazing Incidence and Multilayer X-Ray Optical Systems*, 1997 (unpublished).
28. E. M. Gullikson, "Scattering from Normal Incidence EUV Optics," *Proc. SPIE* **3331**:72–80 (1998).
29. V. Holý, U. Pietsch, and T. Baumbach, *High Resolution X-Ray Scattering from Thin Films and Multilayers* (Springer, Berlin, 1999).
30. D. G. Stearns, D. P. Gaines, D. W. Sweeney, et al., "Nonspecular X-Ray Scattering in a Multilayer-Coated Imaging System," *J. Appl. Phys.* **84** (2):1003–1028 (1998).
31. E. Spiller, D. G. Stearns, and M. Krumrey, "Multilayer X-Ray Mirrors: Interfacial Roughness, Scattering, and Image Quality," *J. Appl. Phys.* **74**:107–118 (1993).
32. W. M. Tong and R. S. Williams, "Kinetics of Surface Growth: Phenomenology, Scaling, and Mechanisms of Smoothing and Roughening," *Annu. Rev. Phys. Chem.* **45**:401–438 (1994).
33. E. Spiller, S. Baker, E. Parra, et al., "Smoothing of Mirror Substrates by Thin Film Deposition," *Proc. SPIE* **3767**:143–153 (1999).
34. C. Montcalm, R. F. Grabner, R. M. Hudyma, et al., "Multilayer Coated Optics for an Alpha-Class Extreme-Ultraviolet Lithography System," *Proc. SPIE* **3767** (1999).
35. P. Hoghoj, E. Ziegler, J. Susini, et al., "Focusing of Hard X-Rays with a W/Si Supermirror," *Nucl Instrum Meth Phys Res B* **132** (3):528–533 (1997).
36. K. D. Joensen, P. Voutov, A. Szentgyorgyi, et al., "Design of Grazing-Incidence Multilayer Supermirrors for Hard-X-Ray Reflectors," *Appl Opt* **34** (34):7935–7944 (1995).
37. F. Mezei, "Multilayer Neutron Optical Devices," in *Physics, Fabrication, and Applications of Multilayered Structures*, edited by P. Dhez and C. Weisbuch (Plenum Press, New York, 1988), pp. 311–333.
38. D. J. Nagel, J. V. Gilfrich, and T. W. Barbee, Jr., "Bragg Diffractors with Graded-Thickness Multilayers," *Nucl. Instrum. Methods* **195**:63–65 (1982).

39. D. G. Stearns, R. S. Rosen, and S. P. Vernon, "Multilayer Mirror Technology for Soft-X-Ray Projection Lithography," *Appl. Opt.* **32** (34):6952–6960 (1993).
40. S. P. Vernon, M. J. Carey, D. P. Gaines, et al., "Multilayer Coatings for the EUV Lithography Test Bed," presented at the POS Proc. on Extreme Ultraviolet Lithography, Monterey, Calif., 1994 (unpublished).
41. E. Spiller and L. Golub, "Fabrication and Testing of Large Area Multilayer Coated X-Ray Optics," *Appl. Opt.* **28**:2969–2974 (1989).
42. E. Spiller, J. Wilczynski, L. Golub, et al., "The Normal Incidence Soft X-Ray, $\lambda = 63.5$ Å Telescope of 1991," *Proc. SPIE* **1546**:168–174 (1991).
43. D. L. Windt and W. K. Waskiewicz, "Multilayer Facilities Required for Extreme-Ultraviolet Lithography," *J. Vac. Sci. Technol. B* **12** (6):3826–3832 (1994).
44. M. P. Bruijn, P. Chakraborty, H. W. van Essen, et al., "Automatic Electron Beam Deposition of Multilayer Soft X-Ray Coatings with Laterally Graded d Spacing," *Opt. Eng.* **25**:916–921 (1986).
45. D. W. Sweeney, R. M. Hudyma, H. N. Chapman, et al., "EUV Optical Design for a 100-nm CD Imaging System," *Proc. SPIE* **3331**:2–10 (1998).
46. M. Schuster and H. Göbel, "Parallel-Beam Coupling into Channel-Cut Monochromators Using Curved Graded Multilayers," *J. Phys. D: Appl. Phys.* **28**:A270–A275 (1995).
47. J. S. Taylor, G. E. Sommargren, D. W. Sweeney, et al., "The Fabrication and Testing of Optics for EUV Projection Lithography," *Proc. SPIE* **3331**:580–590 (1998).
48. R. Soufli, E. Spiller, M. A. Schmidt, et al., "Multilayer Optics for an Extreme Ultraviolet Lithography Tool with 70 nm Resolution," *Proc. SPIE* **4343**:51–59 (2001).
49. R. Soufli, R. M. Hudyma, E. Spiller, et al., "Sub-Diffraction-Limited Multilayer Coatings for the 0.3 Numerical Aperture Micro-Exposure Tool for Extreme Ultraviolet Lithography," *Appl. Opt.* **46** (18):3736–3746 (2007).
50. E. Spiller, "Multilayer Interference Coatings for the Vacuum Ultraviolet," in *Space Optics*, edited by B. J. Thompson and R. R. Shannon (National Academy of Sciences, Washington, D.C., 1974), pp. 581–597.
51. A. Khandar and P. Dhez, "Multilayer X-Ray Polarizers," *Proc. SPIE* **563**:158–163 (1985).
52. E. Spiller, "The Design of Multilayer Coatings for Soft X Rays and Their Application for Imaging and Spectroscopy," in *New Techniques in X-ray and XUV Optics*, edited by B. Y. Kent and B. E. Patchett (Rutherford Appleton Lab., Chilton, U.K., 1982), pp. 50–69.
53. J. B. Kortright and J. H. Underwood, "Multilayer Optical Elements for Generation and Analysis of Circularly Polarized X Rays," *Nucl. Instrum. Meth.* **A291**:272–277 (1990).
54. J. B. Kortright, H. Kimura, V. Nikitin, et al., "Soft X-Ray (97-eV) Phase Retardation Using Transmission Multilayers," *Appl. Phys. Lett.* **60**:2963–2965 (1992).
55. J. B. Kortright, M. Rice, S. K. Kim, et al., "Optics for Element-Resolved Soft X-Ray Magneto-Optical Studies," *J. Magnetism and Magnetic Materials* **191**:79–89 (1999).
56. S. Di Fonzo, B. R. Muller, W. Jark, et al., "Multilayer Transmission Phase Shifters for the Carbon K Edge and the Water Window," *Rev. Sci. Instrum.* **66** (2):1513–1516 (1995).
57. J. B. Kortright, M. Rice, and K. D. Franck, "Tunable Multilayer EUV Soft-X-Ray Polarimeter," *Rev. Sci. Instrum.* **66** (2):1567–1569 (1995).
58. F. Schäfers, H. C. Mertins, A. Gaupp, et al., "Soft-X-Ray Polarimeter with Multilayer Optics: Complete Analysis of the Polarization State of Light," *Appl. Opt.* **38**:4074–4088 (1999).
59. N. N. Chapman, A. Barty, M. J. Bogan, et al., "Femtosecond Diffraction Imaging with a Soft-X-Ray Free-Electron Laser," *Nat. Phys.* **2**:839–843 (2006).
60. S. Bajt, H. N. Chapman, E. Spiller, et al., "A Camera for Coherent Diffractive Imaging and Holography with a Soft-X-Ray Free Electron Laser," *Appl. Opt.* **47**:1673 (2008).

NANOFOCUSING OF HARD X-RAYS WITH MULTILAYER LAUE LENSES

Albert T. Macrander,¹ Hanfei Yan,^{2,3} Hyon Chol Kang,^{4,5}
Jörg Maser,^{1,2} Chian Liu,¹ Ray Conley,^{*1,3} and
G. Brian Stephenson^{2,4}

¹*X-Ray Science Division
Argonne National Laboratory
Argonne, Illinois*

²*Center for Nanoscale Materials
Argonne National Laboratory
Argonne, Illinois*

³*National Synchrotron Light Source II
Brookhaven National Laboratory
Upton, New York*

⁴*Materials Science Division
Argonne National Laboratory
Argonne, Illinois*

⁵*Advanced Materials Engineering Department
Chosun University
Gwangju, Republic of Korea*

ABSTRACT

Multilayer Laue lenses (MLLs) have the potential to provide hard x-ray beams focused to unprecedented dimensions that approach the atomic scale. A focus of 5 nm or below is on the horizon. We review the diffraction theory as well as the experimental results that support this vision, and we present reasons to prefer harder x rays in attempting to achieve this goal.

*Ray Conley is now at National Synchrotron Light Source II, Brookhaven National Laboratory.

42.1 INTRODUCTION

Soon after he discovered x rays and explored their strong penetrating power, William Roentgen also found that they were only very weakly deflected. That is, the new rays traveled almost straight through the materials Roentgen put in front of them.¹ Small angular deflections via refraction arise from a value of the index refraction close to that of the vacuum. In the case of x rays, the index of refraction of any material is only slightly smaller than unity.² Just as for other electromagnetic radiation, focusing of x rays inherently involves deflecting the rays, that is, deflecting the Poynting vector of wavefronts. The consequence of this fundamental property of matter, namely, that the index of refraction for x rays is almost unity, implies that x rays are very difficult to focus to dimensions approaching the x-ray wavelength. (Here we consider a typical x-ray wavelength as 1.24 nm, corresponding to an x-ray energy of 1 keV, which we take as the border between soft and hard x rays.) The numerical aperture (NA) can be increased by placing many refractive lenses in series, and the net effect can approach the sum of the NAs of the individual lenses.^{3,4} However, we concentrate in this review on another means of achieving high NA by employing Bragg diffraction. For Bragg diffraction from crystals, an angle of 45° is not unusual. If a lens can be made that employs diffraction to approach this Bragg angle, it would come close to a NA of unity. A multilayer Laue lens (MLLs), shown schematically in Fig. 1, is an optic that can, in principle, achieve high diffraction angles, and correspondingly large NA, efficiently. The diffracted beam is transmitted through the lens. When crystal diffraction occurs in transmission the diffraction geometry is known as a Laue case,⁵ and, analogously, the name of Max von Laue was used to name this type of lens.

The Rayleigh criterion is well known in classical optics and determines the diffraction-limited focus of a lens. In the focal plane this limit corresponds closely to the distance d of the first minimum of the Fraunhofer diffraction pattern of the lens aperture from the optical axis,

$$d = \alpha [\lambda/\text{NA}] \quad (1)$$

where λ is the wavelength and α is a constant on the order of unity. For two-dimensional focusing by a round lens α equals 0.61, whereas for a linear (or rectangular) lens, α equals 0.5.⁶ Since hard

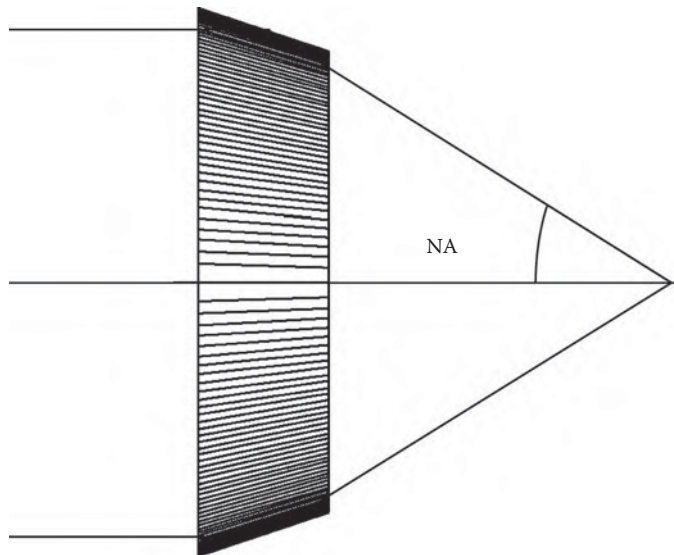


FIGURE 1 Multilayer Laue lens. A wedged version is shown. A local Bragg angle is made between incoming x rays and multilayer interfaces. The numerical aperture (NA) is shown.

x rays have wavelengths at or near atomic dimensions, focus sizes approaching atomic dimensions are thereby, in principle, feasible for values of the NA near unity.

If one can fabricate the layers in an MLL with control over the layers that is of atomic dimensions, one should be able to achieve a focus approaching atomic dimensions.^{7–9} Such control is available with several thin-film techniques, such as magnetron sputtering, atomic layer deposition,¹⁰ and many types of epitaxy.¹¹ However, there are other important factors to be considered in choosing the deposition technique, as discussed below. To date, only magnetron sputtering has been shown to allow one to usefully deposit a very large number of total zones as required for a useful linear Fresnel zone plate.

We note that phase-reversal zone plates that focus x rays have been designed since 1974,¹² and that these are also diffractive optics (see Chap. 40). In principle, zone plates made by the traditional photolithographic steps are also capable of very large NAs. However, in practice, the photolithographic process is limited to maximum aspect ratios of ~ 20 . That is, for an outermost zone width of 1 nm, the depth of the zone plate cannot be larger than ~ 20 nm. (See Fig. 2 for the definition of the dimension of a zone plate or MLL that we presently refer to as the depth.) This situation limits the efficiency very severely for hard x rays.

The efficiency of phase zone plates depends on the phase shift difference of waves transmitted through adjacent zones. If absorption losses can be ignored, the x-ray waves propagating through adjacent zones should be perfectly out of phase in order to achieve maximum efficiency.¹² The phase-shift difference increases both with an increasing index of refraction contrast between adjacent zones as well as with increasing zone plate depth. The index of refraction contrast is a function of the wavelength of the x rays and, aside from absorption edges, decreases with decreasing x-ray wavelength, that is, with increasing energy. Consequently, zone plates designed for optimum efficiency at high energies must have an increased depth compared to ones designed for optimum efficiency at low energies in order to compensate for the reduced index of refraction contrast at high energies. This is the fundamental reason that efficient focusing of hard x rays requires zone plates with very large aspect ratios. Although zone plates have been stacked on one another with some success,¹³ the MLL technology provides essentially limitless aspect ratios since practically any lens thickness can be chosen. We note that an aspect ratio of 2000 has been demonstrated recently.¹⁴

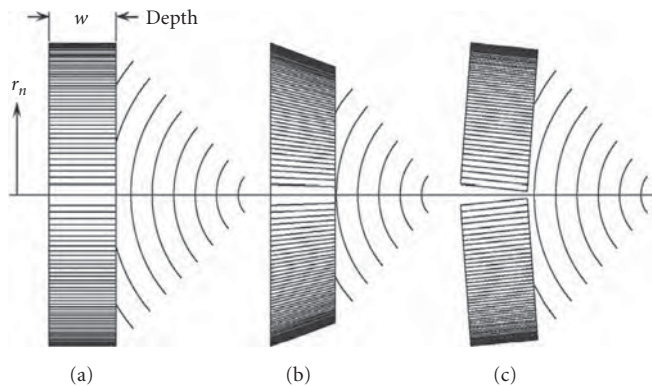


FIGURE 2 Three different types of MLLs. (a) Flat, equivalent to a linear Fresnel zone plate, in which the interfaces are parallel to an the optical axis. Here a Bragg condition is not satisfied. (b) Ideal (also called wedged), as in Fig. 1, in which each interface is angled so as to meet its own local Bragg condition. (c) Tilted, in which a flat case lens is split into two halves, each of which is tilted to an angle. This angle is set to meet a Bragg condition for one of the multilayer pair spacings within the lens. The optical depth dimension w and the “radius” to the n th zone r_n are shown. (Reprinted from Ref. 9. Copyright 2006, with permission from the American Physical Society. <http://link.aps.org/abstract/PRL/v96/e127401>.)

As first explored by Maser and Schmahl¹⁵ and as discussed below, MLLs can also be used in a novel diffractive mode by fulfilling the Bragg condition for Laue-case diffraction from layers. This innovative idea is implemented by tilting the layers to the Bragg angle, and has the consequence that efficiencies are greatly enhanced and wavefront aberrations minimized. Unlike the case for “thin” zone plates,¹² a regime of diffraction known as volume diffraction, akin to x-ray dynamical diffraction in crystals, is needed to model the performance of MLLs.¹⁵ A key result of this theoretical description is that efficiencies in the range of 60 to 70 percent can be achieved. Furthermore, wavefront aberrations can be kept to a level where the Rayleigh resolution can be achieved. For hard x rays, lens thicknesses of tens of micrometers are needed for optimum efficiency. This implies that aspect ratios of 10,000 or larger are needed for a lens capable of focusing x rays to 1 nm. MLLs offer the promise that such a focus can actually be achieved with excellent diffraction efficiency.

42.2 MLL CONCEPT AND VOLUME DIFFRACTION CALCULATIONS

MLLs are made by deposition of bilayers as shown schematically in Fig. 2. As shown in Fig. 2a, a “flat” MLL can be viewed as a linear Fresnel zone plate made by deposition of bilayers. The zone positions r_n must follow the well-known zone plate law given by¹²

$$r_n^2 = n \lambda f + n^2 \lambda^2 / 4 \quad (2)$$

Here λ is the x-ray wavelength and f is the focal length. The second term on the right can be omitted when $n\lambda \ll 4f$, which is still a good approximation for MLLs designed for angstrom wavelengths and millimeter focal lengths, which are the primary subject of this review. The bilayer pair must have significant contrast in charge density resulting in a significant phase and/or intensity difference upon x-ray wave transmission, as in the case of a thin lens,¹² as well as a significant structure factor for local diffraction as a result of satisfying the Bragg condition given by

$$\lambda = 2 (2\Delta r_n) \sin(\theta_n) \quad (3)$$

where $\Delta r_n = (r_n - r_{n-1})$ is the spacing between two successive interfaces. (We note that this equation is for first-order Bragg diffraction. For the sake of simplicity we have not denoted higher orders of diffraction.)

Arranging the diffraction geometry so that $2\Delta r_n$ acts as the Bragg spacing is an essential innovation relative to normal zone-plate usage and theory. Such a diffraction geometry is shown in Fig. 2c, which is called *tilted*. It is achieved by oppositely tilting the two halves to some angle. Normal x-ray zone plates are illuminated parallel to the axis of the zone plate. Their properties are well described by a theory that treats the waves at the exit surface, after accounting for a phase shift and absorption upon transmission through the bilayers, as sources of wavelets in a Huygens construction. These wavelets then propagate to form the focus.¹² This is known as the thin-lens case. The thin-lens treatment fails when multiple scattering of x rays takes place inside the MLL. If that occurs, dynamical diffraction approaches have to be used, and the deviations from the Bragg condition need to be considered. One must then apply a diffraction theory that can also account for multiple internal reflections in a manner akin to dynamical diffraction in crystals.⁵ This more sophisticated theory is known as volume diffraction and was first applied to MLLs by Maser and Schmahl using a coupled-wave approach.¹⁵ Results for this theory are shown in Fig. 3 for an MLL tilted so that the Bragg condition is satisfied for an intermediate zone.⁸ Here the local diffraction efficiency is plotted versus increasing zone number for MLLs having outermost zone widths of 2 and 10 nm. For the central zones of low order, the results are as expected for thin lenses and yield a local diffraction efficiency of 26 percent. However, for the zone where dynamic diffraction occurs (at 2 μm radius and above), the local diffraction efficiency is increased significantly, otherwise the diffraction efficiency becomes very small, similar to x-ray diffraction from a crystal rocked through a diffraction peak. That is, a transition in diffraction properties occurs as one approaches the dynamical diffraction regime, at which point diffraction properties become very sensitive to the Bragg condition.

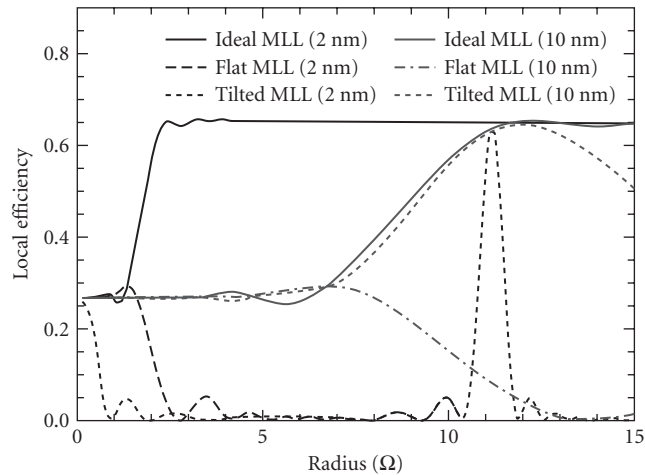


FIGURE 3 Calculated diffraction efficiency at 0.064 nm (19.5 keV) for outermost zone widths of both 10 nm (gray) and 2 nm (black) as a function of radius for ideal (wedged), flat, and tilted MLLs. For flat MLLs efficiencies do not exceed 26 percent and only very low-order zones diffract in the 2-nm case. For an outermost zone width of 10 nm, the ideal and tilted cases have almost the same performance, but for a 2-nm outermost zone, the ideal case is far superior. For the 2-nm tilted case, only a sharp Bragg peak is seen at the radius for which a Bragg condition is satisfied. This figure shows that meeting the Bragg condition everywhere becomes increasingly important for outermost zones less than 10 nm in order to ensure a high efficiency throughout most of the MLL. (See also color insert.) (Reprinted from Ref. 2. Copyright 2006, with permission from the American Physical Society. <http://link.aps.org/abstract/PRL/v96/e127401>.)

As also shown in Fig. 3, the number of zones having a high diffraction efficiency narrows as one considers smaller outermost zones. This reduces the overall efficiency. This would seem to rule out the usefulness of MLLs to structures with outermost zones of 2 nm or less. However, as depicted in Fig. 2b, if, instead of depositing an MLL for which all interfaces are parallel, one can build a local Bragg angle into the structure itself, a wide ranging local efficiency of 65 percent can be achieved. As discussed below, this structure, at first called “ideal” and later “wedged” (since even more ideal MLLs with curved interfaces can be considered as discussed below), has recently been demonstrated and now appears feasible.

42.3 MAGNETRON-SPUTTERED MLLS

Preliminary Study on Periodic Multilayers

It has been known for many years that magnetron sputtering can be used to make periodic multilayers useful as diffraction optics for hard x rays.^{16–18} It has only recently been discovered that layer placement can be well controlled over thousands of layers to form multilayers that are many micrometers thick.¹⁴ The ability to deposit such thick multilayers in a reasonable length of time with the requisite perfection is what sets magnetron sputtering apart from other thin-film deposition technologies such as molecular beam epitaxy.

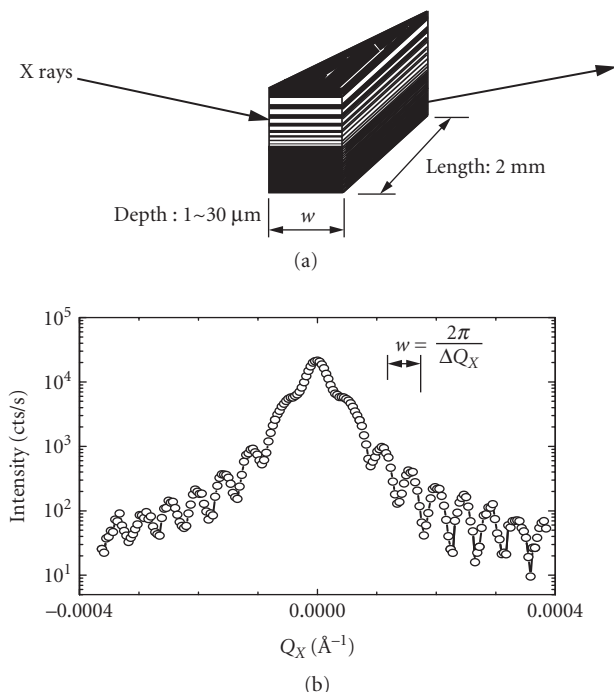


FIGURE 4 (a) Schematic illustration of an MLL made to have a tapered shape so that the optical depth w varies in a direction transverse to the optical axis. This local tapered width can be measured from fringe spacings observed in a transverse (rocking) scan, as shown in (b). (Reprinted from Ref. 25. Copyright 2007, with permission from the American Institute of Physics.)

The resultant structure must be formed into an actual lens, and this requires processing steps such as dicing, thinning, and polishing. The multilayer must be robust enough to withstand these steps without delamination, and this imposes constraints on the internal strain. That sufficiently thick multilayers could be produced and were robust enough to withstand the needed processing was first demonstrated in 2005 for W/Si as well as for Mo/Si periodic multilayers.¹⁹

Another important observation was made in the preliminary studies on thick periodic multilayers. Just as for Laue-case diffraction from thin crystals, thickness fringes occurred that permitted an accurate determination of the thickness in transmission. Far-field diffraction data showing the fringes is shown in Fig. 4.¹⁹ The diffraction efficiency depends strongly on this thickness, and its calibration is important for comparisons to theory.

WSi₂/Si Bilayers and Metrology Based on Scanning Electron Microscope Data

If thick MLLs are to have well-defined zones throughout, then interface roughness should ideally not build up progressively. Kinetic roughening of a growing surface is a complex phenomenon.²⁰ Furthermore, diffusion across the interface driven by a chemical potential gradient should be obviated as much as possible. Evidence for interdiffusion for W/Si multilayers has been reported by Windt et al.¹⁸ For these reasons, WSi₂/Si bilayers were chosen for MLLs. The choice has proven to be

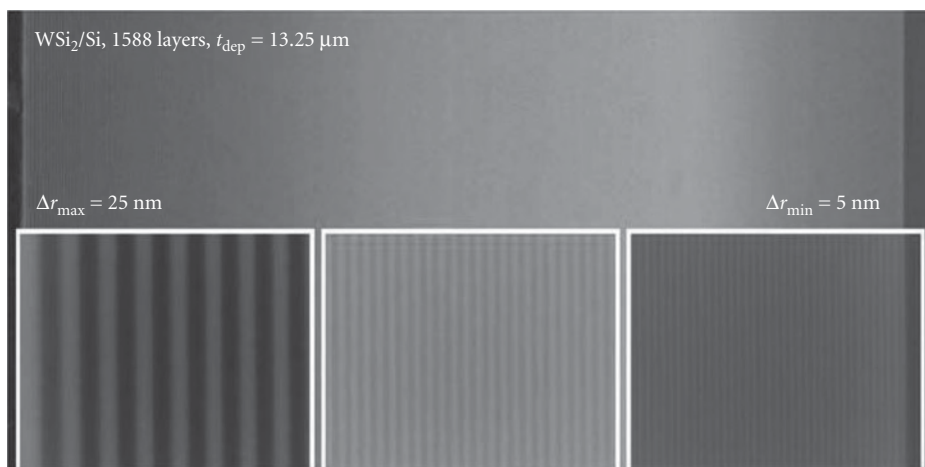


FIGURE 5 A scanning electron microscope image of a partial MLL having a 5-nm-thick outermost zone. The structure was used to obtain a focus of 16 nm for x rays of 0.064-nm (19.5-keV) wavelength. (See also color insert.) (Reprinted from Ref. 14. Copyright 2008, with permission from American Institute of Physics.)

propitious. X-ray reflectivity measurements made in situ, as a multilayer was built up, revealed that the growth of WSi_2 smoothened an underlying Si surface, and that there was not a net build up of roughness, albeit for only five periods.²¹

We note that MoSi_2/Si MLLs are also achievable. A focus of 28.5 nm has recently been reported.²² Such MLLs may be preferable at photon energies below ~ 12 keV due to the L absorption edges of W.

SEM images of cross sections of MLL wafers have proven to be very useful for analyses of the sputtering procedures. Such an SEM image is shown in Fig. 5.¹⁴

Not only must the interface roughness of each interface be well controlled, but also each layer must be grown at the Fresnel zone position with good accuracy. Ignoring the second term in Eq. (2), one finds that the width of the n th zone is given by

$$(\Delta r_n)^{-1} = (r_n - r_{n-1})^{-1} = 2r_n / \lambda f \quad (4)$$

This very useful result has been used to evaluate MLL wafers prior to making lenses and to provide feedback information for the crystal growth. Images taken by scanning electron microscopes on cross-sectioned MLL wafers were used to produce plots of Δr_n versus r_n , such as shown in Fig. 6.^{23,24} Ideally these plots should be linear, and deviations from linearity provided clues on how to adjust the sputtering process.²⁴

Lens Processing

Lenses can be successfully made starting from a deposited multilayer wafer.²⁵ The process is shown illustratively in Fig. 7. The first step is to form a “sandwich structure” by gluing a covering wafer on the multilayer surface with an epoxy. The entire structure is then diced on a high-speed saw, and the resultant bar-shaped sections are blocked with protective end caps. The bars are then bonded to a polishing fixture with the plane of the multilayers perpendicular to the plane of the fixture. The mounted bar is then thinned and polished on one side. The bar is then released, flipped over, and polished on the other side. These steps are needed because sawing damage to the multilayer cannot be avoided and damage on both sides needs to be removed. After mechanical polishing, ion beam milling is needed on each side to remove possible damage introduced by the mechanical polishing. After fine manual polishing, the final MLL is then mounted on a Mo holder and is then ready for beamline use.

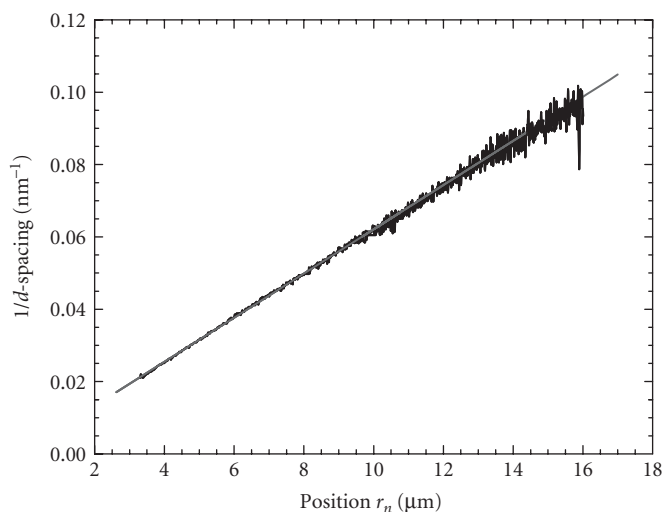


FIGURE 6 Results of analyses of electron microscope (SEM) images for layer thicknesses. The inverse of layer thickness when plotted as a function of MLL radius is shown in as the jagged curve. The straight line shows the desired linear functional dependence for Fresnel zones.

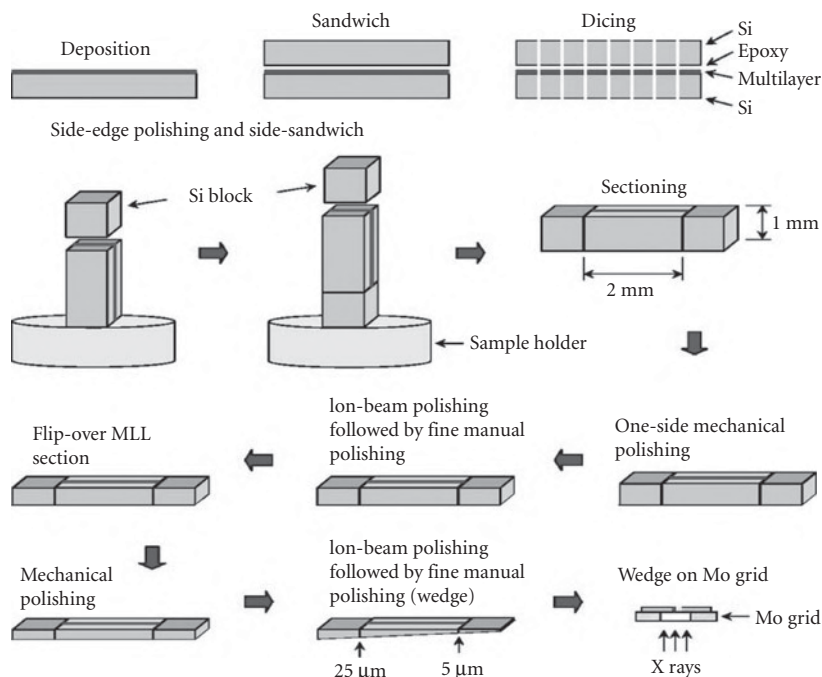


FIGURE 7 Cartoon illustrating the sequence of steps used to process a lens suitable for use with x rays, starting from an as-sputtered MLL wafer. (See also color insert.) (Reprinted from Ref. 25. Copyright 2007, with permission from the American Institute of Physics.)

42.4 INSTRUMENTAL BEAMLINE ARRANGEMENT AND MEASUREMENTS

Measurements of the focus have been made at beamlines 26 ID, 12 BM, and 8 ID at the Advanced Photon Source.^{8,14} The setup used for the measurements is shown in Fig. 8. For all the measurements made to date only a partial MLL was used. That is, an incomplete linear zone plate structure was studied. These were incomplete in two ways. First, not all zones were deposited. That is, sputter depositions were done by first sputtering an outermost zone, but the sputtering was halted before growing the inner most zones. Second, only one side above the center line in Fig. 2c was studied so that only a single tilt angle was employed. For the measurements, a horizontal diffraction plane was used. The MLL interfaces were positioned parallel to a vertical plane, and tilting was done by rotating MLLs about a vertical axis. An upstream beamline slit to aperture the horizontal source size was used to increase the horizontal coherence length, and reduction of this slit aperture proved to be needed to achieve the smallest focuses. X rays were monochromatized with a standard Si(111) double bounce arrangement. The energy was left fixed during all measurements. Data were obtained both at 19.5 and 29.5 keV.

Focus-size measurements were made by driving a specially made “analyzer” horizontally through the focused beam as shown in Fig. 8. The analyzer was a sputtered Pt thin-film processed in a manner similar to that of the MLLs themselves. In this way, a Pt thin-film segment was illuminated edge-on

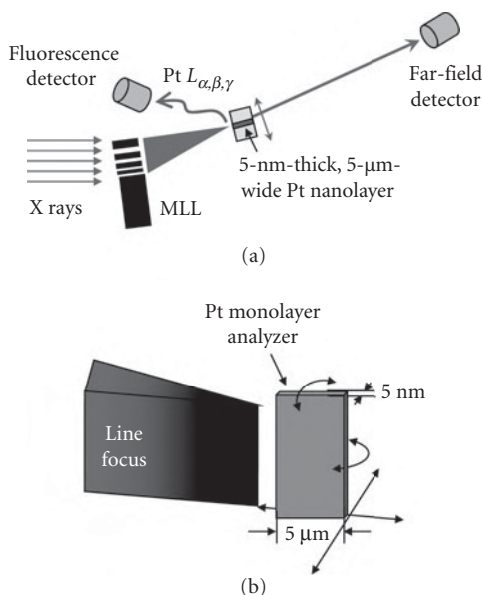


FIGURE 8 (a) Experimental layout of MLL focusing measurements. For the measurements, a specially made analyzer made from a Pt thin film deposited by sputtering was scanned through the focal plane. One detector was positioned to count Pt fluorescence photons and another was positioned to count diffracted photons in the far field. (b) Schematic illustration of the radiation pattern near the focal plane and the Pt analyzer. The two must be well aligned for correct measurement of the width of the line focus.

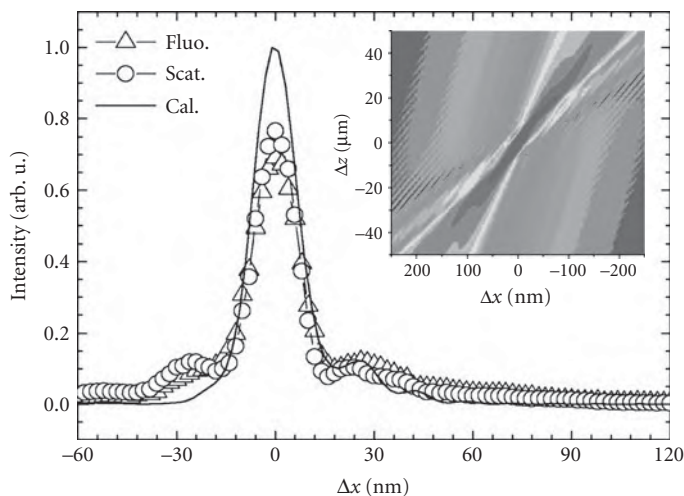


FIGURE 9 Measured and calculated intensity profiles for the focus of the MLL shown in Fig. 5. The FWHM values are 17.6 nm from fluorescence data and 15.6 nm from far-field scattering data. These should be compared to a calculated value of 15.0 nm. The calculated intensities were scaled according to measured and calculated efficiencies of 30 and 32 percent, respectively. The calculated results do not include the effects of vibrations, finite analyzer width, and finite source size. The inset shows the calculated isophote pattern near the focal plane. (See also color insert.) (Reprinted from Ref. 14. Copyright 2008, with permission from the American Institute of Physics.)

(or nearly so) by the line focus of an MLL. Angular alignment of the Pt edge to that of the MLL focal line was critical in obtaining the smallest attainable focus. As shown schematically in Fig. 8, two signals were recorded to measure the profile of a focused beam as the analyzer was scanned. A p-i-n detector was positioned at roughly 90° to record Pt fluorescence while a scintillation counter collected the signal in the far field. The far-field signal recorded scattering from the analyzer including, in particular, reflectivity from the Pt thin film. Recent results for the measurement of the focus at 19.5 keV are shown in Fig. 9.¹⁴ The far-field detector was also used to measure the efficiency.

Efficiencies of MLLs were measured by two methods. First, as shown in Fig. 10, without an analyzer present and with the detector set to observe the direct beam that passes through an MLL, one observes a dip in the transmitted intensity as the MLL is scanned through the incident beam. With the MLL tilted far from a Bragg condition, there is a dip due to absorption. But with the MLL tilted to a Bragg angle, a significantly larger dip was found. The difference between the dips is akin to extinction in crystal diffraction, since intensity is removed from the direct beam by Laue-case diffraction. Unlike for thin Fresnel lenses (or gratings), the Bragg condition depletes diffraction intensity from other orders, and, as a consequence, the extinction measurement is a reasonable measure of the overall efficiency. The second method to obtain the diffraction efficiency is based on the observation that a $\theta - 2\theta$ diffraction scan, where θ is the tilting angle, yields a measure of the local diffraction efficiency. This follows since the zone spacings satisfy the Bragg condition progressively as θ is increased so that the local efficiency is measured. An example far-field scan of the local diffraction efficiency is shown in Fig. 11. Calculated values are also shown in Fig. 11. The overall diffraction efficiency can be obtained by integration of the diffraction scan data and by normalizing to the incident beam intensity. The two methods have been found to agree to within a few percent. At 19.5 keV the efficiency for the MLL shown in Fig. 5 was found to be 31 percent.⁸

Results were also obtained at 29.25 keV for the MLL shown in Fig. 5. Fluorescence data for scans through the focus revealed a FWHM of 16 nm, as shown in Fig. 12. The measured efficiency in this case was 17 percent.

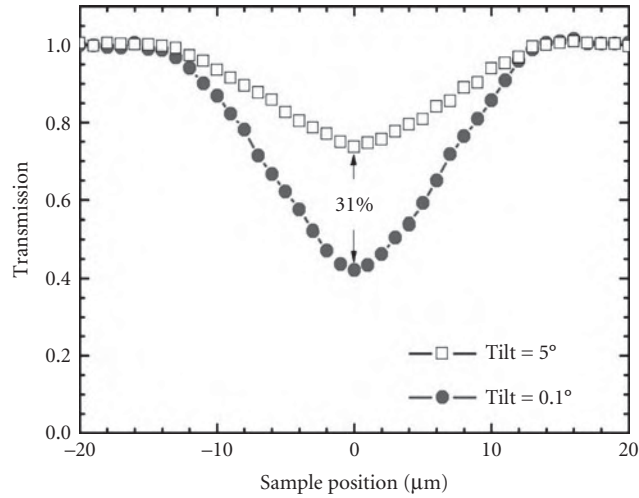


FIGURE 10 The fraction of the photons transmitted through the MLL of Fig. 5 at the Bragg condition (circles) and away from the Bragg condition (squares) as the MLL is scanned through the incident beam. The difference between the two is due to x rays that are directed to the focus only when the Bragg condition is satisfied, a phenomenon known as extinction.

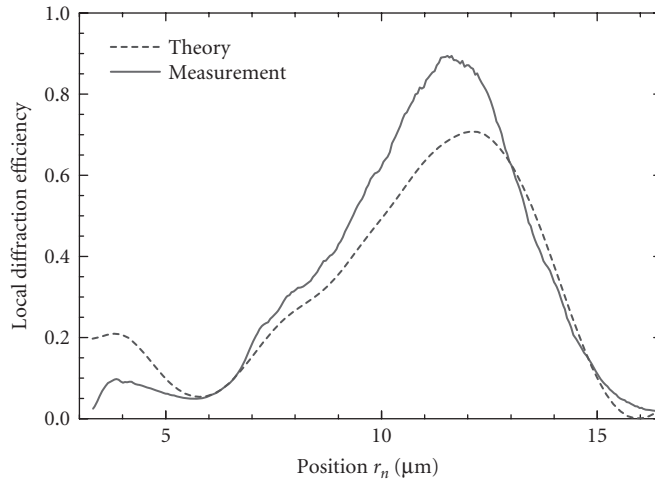


FIGURE 11 Measured and calculated focusing efficiency as a function of radius for an MLL with 5-nm-thick outermost zone width. The integrated efficiencies are 33 percent measured and 30 percent calculated. The calculations were made with coupled wave theory and show the transition from kinematic to dynamic properties at d-spacings of ~ 10 nm. The larger diffraction efficiency is most likely due to measurement uncertainties.

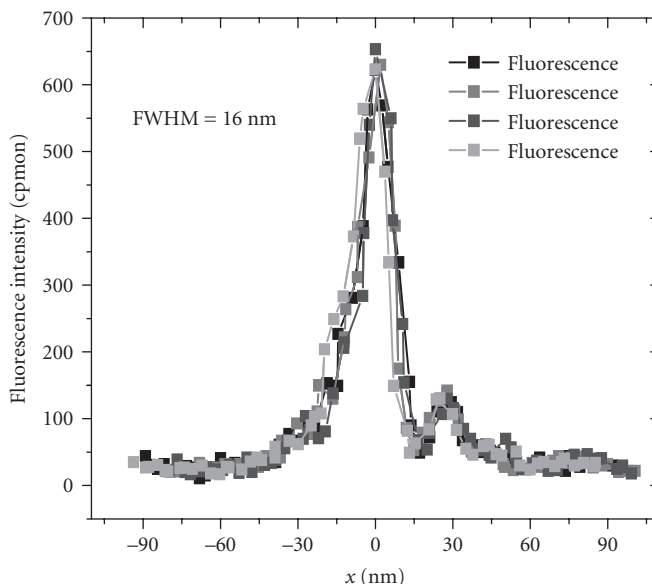


FIGURE 12 Measured line focuses of the MLL in Fig. 5 with 0.042-nm (29.25-keV) x rays. Results from four scans of the fluorescence intensity are shown. The efficiency was measured to be 17 percent. A FWHM value of 16 nm applies. (See also color insert.)

42.5 TAKAGI-TAUPIN CALCULATIONS

There does not appear to be a fundamental reason why x-ray focus sizes should not reach down to atomic dimensions. The coupled-wave-theory (CWT) used by Maser and Schmahl approximates an MLL as consisting of regions that are locally periodic and solves for local grating solutions for the diffracted waves. The orders of diffraction that result from the local gratings are then spliced to solve for the dominant diffracted waves. For values of the NA approaching unity, this approximate procedure breaks down. However, other theoretical approaches have been applied to explore a possible smallest focus on fundamental grounds. A focus size of 0.83 nm was calculated by Schroer for a wedged MLL,²⁶ and Pfeiffer et al.²⁷ report finding no lower limit for a flat MLL in a 1-to-1 imaging condition. In addition, a new diffraction theory has been developed⁹ based on the Takagi-Taupin equations for crystal diffraction,²⁸ which starts from the MLL layer structure, a priori. This new diffraction treatment proceeds by making a coordinate transformation that maps the zone plate spacings to a pseudo-Fourier series for the charge density. The electric field waves are then solved as in the case of a bulk crystal. Possible focuses below 1 nm are also predicted by the Takagi-Taupin approach.⁹

42.6 WEDGED MLLS

The new Takagi-Taupin-based theory has been applied to the case of an MLL in which each interface is tilted to a separate angle so that all zones may satisfy the Bragg condition for a divergent fan of radiation arising from a point source (see Fig. 1). When viewed internally the zones have wedge-like shapes, and this more or less “ideal” MLL structure has been called a wedged MLL (wMLL).²⁹ It was recently demonstrated that such a structure can be deposited by sputtering past a mask. The effect of the mask is to create a lateral gradient in the deposited film thickness, and, since for each

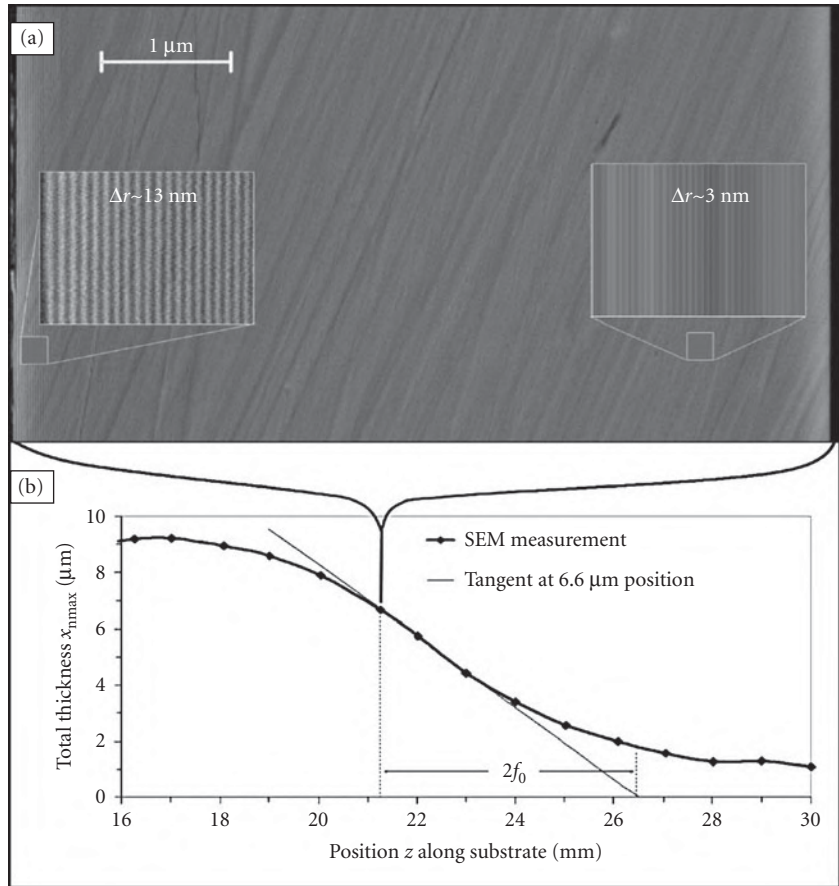


FIGURE 13 Ideal (wedged) MLL. (a) SEM image for one wafer location of a multilayer grown by sputtering past a mask such that a lateral gradient in layer thicknesses is produced over the surface of the wafer. The linear mask is designed to produce a gradient that creates a set of ideal MLL structures of varying focal lengths. The thickness of the total multilayer structure is laterally also changed accordingly, and, as shown in (b), demonstrates the lateral gradient. (Reprinted from Ref. 29. Copyright 2008, with permission from the American Institute of Physics.)

sputtered layer the ratio of its thickness at the entrance surface to its thickness at the exit surface is the same, an entire wMLL can be built up by sputtering past a single mask. A proof-of-principle result is shown in Fig. 13. SEM data at one location are shown in Fig. 13a, and the total thickness as a function of position is shown in Fig. 13b. These data reveal that the total deposited thickness varies laterally as required. The lateral variation required is different for different focal lengths, and one such sputtered wafer can, in principle, provide MLLs for a series of focal lengths. Results of zone thickness analyses of the SEM micrograph shown in Fig. 13a are shown in Fig. 14a.

The expected results as calculated by the Takagi-Taupin approach for a wMLL having a 3-nm outermost zone width are shown in Fig. 14b. The width in the focal plane is only 2.4 nm. We note that sputtered layer thicknesses as small as 0.75 nm have been demonstrated,³⁰ and this value can be used to estimate the smallest feasible focus presently achievable with sputtering. For x rays with an energy of 19.5 keV, a wMLL having an outermost zone width of 0.75 nm is predicted to achieve a focus of 0.7 nm, a value of atomic dimensions. The efficiency of such an MLL is calculated to be 50 percent.

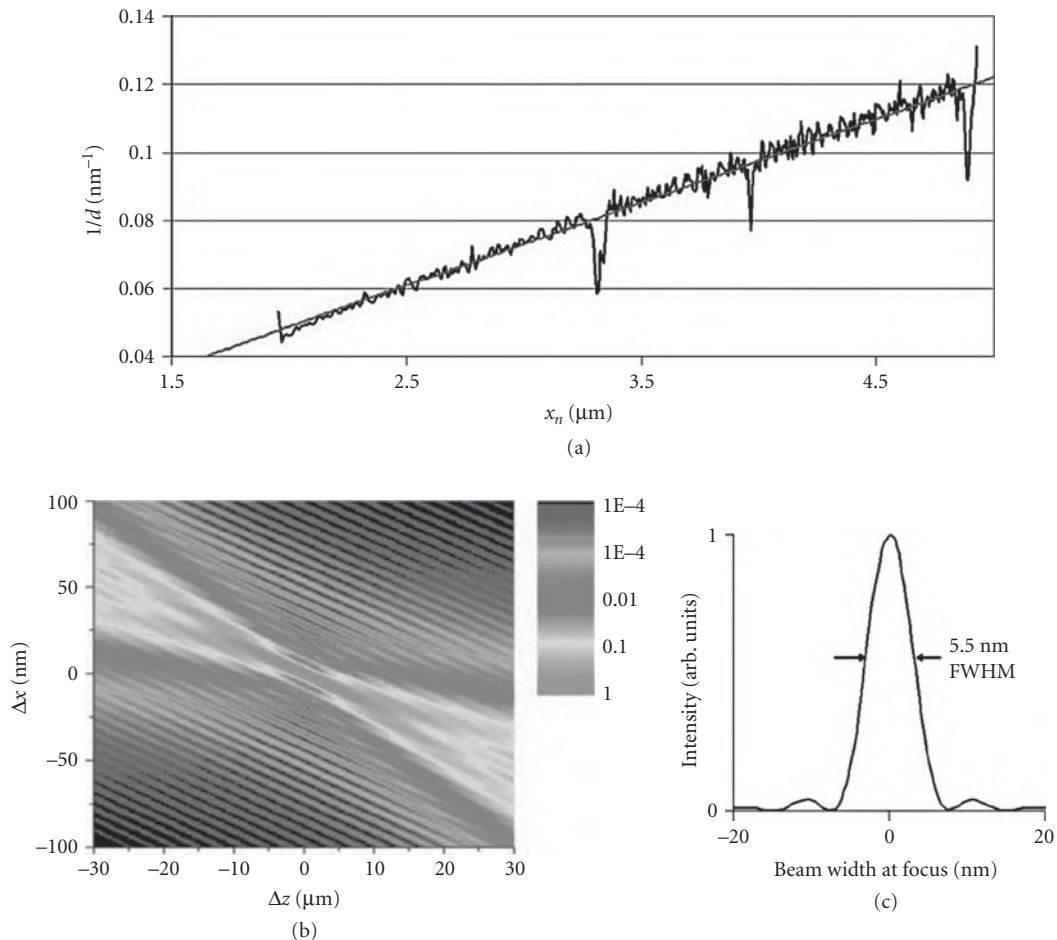


FIGURE 14 Analyses for ideal (wedged) MLL structure of Fig. 13a. The inverse layer thickness vs. radius is shown in (a). The calculated isophote pattern and intensity in the focal plane are shown in (b) and (c), respectively. (See also color insert.) (Reprinted from Ref. 29. Copyright 2008, with permission from the American Institute of Physics.)

42.7 MMLs WITH CURVED INTERFACES

The Takagi-Taupin-based calculations can be extended to curved interfaces. That curved interfaces might be more ideal is suggested by the observation that the Bragg condition cannot be met both at the entrance and at the exit surfaces of a wedged MLL. That is, the angle of a ray from a point source to any given interface at the entrance side is not the same as at the exit side. This simple observation is born out by a rigorous treatment, with the result that the surfaces must, in general, lie on concentric ellipses. For a parallel incident beam, these are concentric parabolas. If thin enough zones can somehow be deposited, the calculations predict that focuses of a few angstroms should be possible. That is, diffraction theory does not rule out true atomic dimensions for the focus of an MML. The calculated result is shown in Fig. 15. We stress that achieving such a focus will require the actual deposition of such a structure, and at the present time this appears to be beyond the limits of magnetron sputtering.

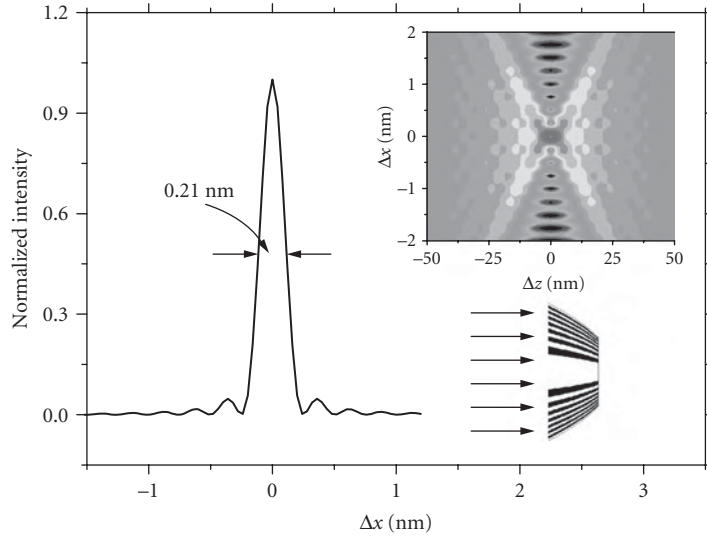


FIGURE 15 Calculated intensity in the focal plane for an MLL having parabolically shaped interfaces and an outermost zone width of 0.25 nm. The lower inset shows a cartoon of the MLL, and the upper inset shows an isophote pattern around the focus. (See also color insert.) (Reprinted from Ref. 9. Copyright 2007, with permission from the American Physical Society. <http://link.aps.org/abstract/PRB/v76/e115438>.)

42.8 MLL PROSPECTS

We consider now several practical limits that apply to the useful application of MLLs.³¹ First, useful lenses should have reasonably large focal lengths to provide space for samples and sample environments. Somewhat arbitrarily, we examine the consequences for limiting the focal length to 1 mm.

By combining the first term in Eq. (2) with Eq. (4), we obtain

$$r_N = \lambda f / (2\Delta r_N) \quad (5)$$

Here N is the outermost zone number given by

$$N = \lambda f / (2\Delta r_N)^2 \quad (6)$$

It is convenient to write Eq. (5) in terms of the total deposited thickness needed for a full linear zone plate, $D_N = 2r_N$, and the x-ray photon energy, $E = hc/\lambda$. Here $hc = 1.24$ keV·nm is the product of Planck's constant and the speed of light. The net result is

$$D_N = f hc / (2\Delta r_N E) \quad (7)$$

This equation is plotted in Fig. 16 for $f = 1$ mm and for $\Delta r_N = 1, 3, 10$, and 30 nm. As shown in Fig. 16, for a 1-nm focus, a total of 100 μm must be deposited for an MLL that functions at 15 keV. A very important observation is that higher photon energies relax this requirement to smaller thicknesses.

The related plot of N as a function of E is shown in Fig. 17. The significance of this figure relates to chromatic blurring which occurs for an incident bandpass, $\Delta E/E$, greater than $1/N$.³² So, for example, to avoid chromatic blurring for a 1-nm focus at 15 keV, a monochromator upstream of the MLL is needed with a specification $E/\Delta E = 2 \times 10^4$.

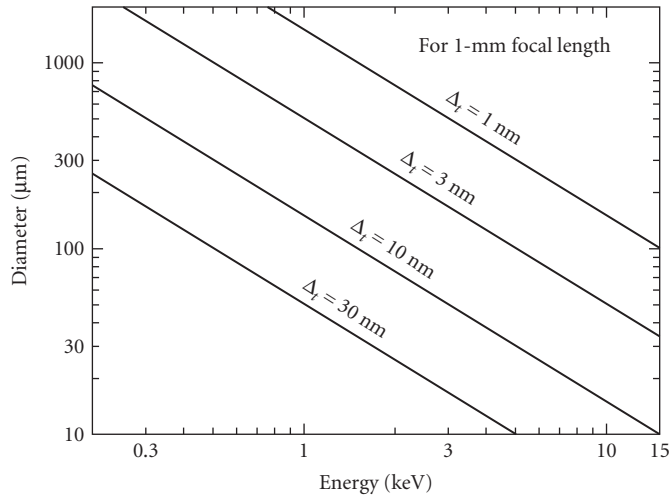


FIGURE 16 Required MLL diameter as a function of x-ray photon energy with a fixed focal length of 1 mm. Diameters are shown for focuses of 1, 3, 10, and 30 nm. For a completed MLL of a certain diameter, the focus size decreases with increasing energy.

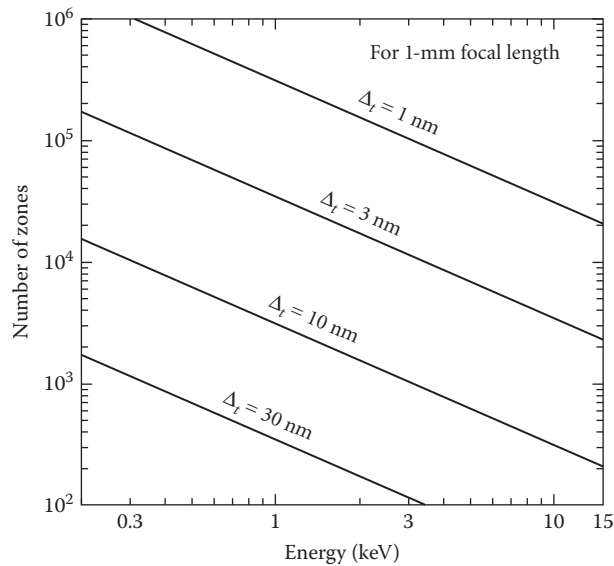


FIGURE 17 Required number for Fresnel zones as a function of x-ray photon energy with a fixed focal length of 1 mm. Diameters are shown for focuses of 1, 3, 10, and 30 nm. For a completed MLL with a certain number of zones, the focus size decreases with increasing energy.

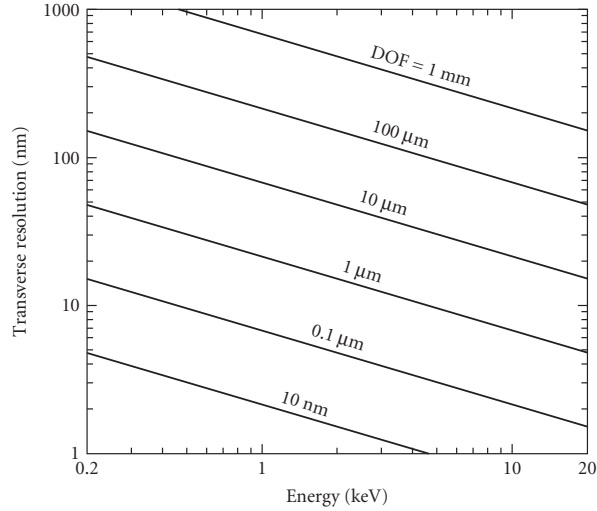


FIGURE 18 The transverse resolution (the width of the outermost zone) plotted as a function of photon energy for several depths of focus (DOF). Here the focal length has been fixed at 1 mm. For a given resolution, the DOF increases with photon energy.

Finally, we consider the depth of focus (DOF). This is also a very important consideration, since it determines the thickness of a sample that can be studied. The DOF is given by

$$\text{DOF} = \lambda / \text{NA}^2 \quad (8)$$

Here the NA is $D_N/2f$. The lateral resolution Δr_N is plotted in Fig. 18 as a function of photon energy for DOF values of 1 mm, 100 μm , 10 μm , 1 μm , 0.1 μm , and 10 nm. For a 1-nm resolution the DOF is 48 nm at 15 keV.

As shown in Figs. 16, 17, and 18, MLLs are increasingly achievable and propitious as one considers ever harder x rays. Smaller total deposited thickness, relaxed monochromaticity requirements, and larger depth of focus all apply for hard x rays compared to soft x rays.

42.9 SUMMARY

Multilayer Laue lenses have the potential to provide hard x-ray beams focused to unprecedented dimensions that approach the atomic scale. A focus of 5 nm is on the horizon. We have reviewed the diffraction theory as well as the experimental results that support this vision. MLLs are especially promising for hard x rays with energies above 20 keV, since fewer layers are required for very small focuses. A primary task yet to be accomplished is to refine the sputtering to increase the number of total layers.

42.10 ACKNOWLEDGMENTS

We are grateful for the support of the Scientific User Facilities/Advanced Photon Source (APS), Materials Science Division (MSD), and Center for Nanoscale Materials at Argonne National Laboratory. We are also grateful to the Electron Microscopy Center of MSD for the use of their facilities. Work at the APS is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06 CH 11357.

42.11 REFERENCES

1. A. Stanton, "Wilhelm Conrad Röntgen on a New Kind of Rays: Translation of a Paper Read before the Würzburg Physical and Medical Society, 1895," *Nature* **53**:274–276, doi:10.1038/053274b0.
2. B. L. Henke, E. M. Gullikson, and J. C. Davis, *At. Data Nucl. Data Tables* **54**:181 (1993).
3. A. Snigirev, V. Kohn, I. Snigireva, B. Lengeler, *Nature* **384**:49 (1996).
4. C. G. Schroer and B. Lengeler, *Phys. Rev. Lett.* **94**:054802 (2005).
5. W. H. Zachariasen, *Theory of X-Ray Diffraction in Crystals*, Constable and Co., London, 1945; A. Authier, *Dynamical Theory of X-Ray Diffraction*, Oxford University Press, 2001.
6. M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon, New York, 1980.
7. J. Maser, G. B. Stephenson, S. Vogt, W. Yun, A. Macrander, H. C. Kang, C. Liu, R. Conley, *Proc. SPIE* **5539**:185 (2004).
8. H. C. Kang, J. Maser, G. B. Stephenson, C. Liu, R. Conley, A. T. Macrander, and S. Vogt, *Phys. Rev. Lett.* **96**:127401 (2006).
9. H. Yan, J. Maser, A. T. Macrander, Q. Shen, S. Vogt, B. Stephenson, and H. C. Kang, *Phys. Rev. B* **76**:115438 (2007).
10. M. Ritala, K. Kukli, A. Rahtu, P. I. Räsänen, M. Leskela, T. Sajavaara, and J. Keinonen, *Science* **288**:319 (2000).
11. V. Swaminathan and A. T. Macrander, *Materials Aspects of GaAs and InP Based Structures*, Prentice Hall, Englewood Cliffs, N.J., 1991.
12. J. Kirz, *J. Opt. Soc. Am.* **64**:301 (1974).
13. J. Maser, B. Lai, W. Yun, S. D. Shastri, Z. Cai, W. Rodrigues, S. Xu, and E. Trakhtenberg, *Proc. SPIE* **4783**:74 (2002).
14. H. C. Kang, H. Yan, R. P. Winarski, M. V. Holt, J. Maser, C. Liu, R. Conley, S. Vogt, A. T. Macrander, and G. B. Stephenson, *Appl. Phys. Lett.* **92**:221114 (2008).
15. J. Maser and G. Schmahl, *Opt. Commun.* **89**:355 (1992).
16. T. W. Barbee, *Opt. Eng.* **25**:898 (1986).
17. E. Spiller, *Soft X-Ray Optics*, SPIE, Bellingham, W. A., 1994.
18. D. L. Windt, F. E. Christensen, W. W. Craig, C. Hailey, F. A. Harrison, M. Jimenez-Garate, R. Kalyanaraman, and P. H. Mao, *J. Appl. Phys.* **88**:460 (2000).
19. H. C. Kang, G. B. Stephenson, C. Liu, R. Conley, A. T. Macrander, J. Maser, S. Bajt, and H. N. Chapman, *Appl. Phys. Lett.* **86**:151109 (2005).
20. A.-L. Barabási and H. E. Stanley, *Fractal Concepts in Surface Growth*, Cambridge University Press, Cambridge, England, 1995.
21. Y.-P. Wang, H. Zhou, L. Zhou, R. L. Headrick, A. T. Macrander, and A. S. Özcan, *J. Appl. Phys.* **101**:023503 (2007).
22. T. Koyama, S. Ichimaru, T. Tsuji, H. Takano, Y. Kagoshima, T. Ohchi, H. Takenaka, *Appl. Phys. Express* **1**:117003 (2008).
23. C. Liu, R. Conley, A. T. Macrander, J. Maser, H. C. Kang, M. A. Zurbuchen, and G. B. Stephenson, *J. Appl. Phys.* **98**:113519 (2005).
24. C. Liu, R. Conley, A. T. Macrander, J. Maser, H. C. Kang, and G. B. Stephenson, *Thin Solid Films* **515**:654 (2006).
25. H. C. Kang, G. B. Stephenson, C. Liu, R. Conley, R. Khachatryan, M. Wiczorek, A. T. Macrander, H. Yan, J. Maser, J. Hiller, and R. Koratala, *Rev. Sci. Instrum.* **78**:046103 (2007).
26. C. G. Schroer, *Phys. Rev. B* **74**:033405 (2006).
27. F. Pfeiffer, C. David, J. F. van der Veen, C. Bergemann, *Phys. Rev. B* **73**:245331 (2006).
28. S. Takagi and H. H. Wills, *Acta Crystallogr.* **15**:1311 (1962); D. Taupin, *Bull. Soc. Fr. Mineral. Cristallogr.* **87**:469 (1964).
29. R. Conley, C. Liu, J. Qian, C. Kewish, A. T. Macrander, H. Yan, H. C. Kang, J. Maser, and G. B. Stephenson, *Rev. Sci. Instrum.* **79**:053104 (2008).
30. Y. Chu, C. Liu, D. Mancini, F. DeCarlo, A. T. Macrander, B. Lai, and D. Shu, *Rev. Sci. Instrum.* **73**:1485 (2002).
31. C. Jacobsen, private communication.
32. Center for X-Ray Optics and Advanced Light Source, *X-Ray Data Booklet*, <http://xdb.lbl.gov/>.

POLARIZING CRYSTAL OPTICS

Qun Shen

*National Synchrotron Light Source II
Brookhaven National Laboratory
Upton, New York*

43.1 INTRODUCTION

Being able to produce and analyze a general polarization of an electromagnetic wave has long benefited scientists and researchers in the field of visible light optics, as well as those engaged in studying the optical properties of condensed matter.^{1–3} In the x-ray regime, however, such abilities have been very limited because of the weak interaction of x rays with materials, especially for production and analysis of circularly polarized x-ray beams. The situation has changed significantly in recent years. The growing interest in studying magnetic and anisotropic electronic materials by x-ray scattering and spectroscopic techniques has initiated many new developments in both the production and the analysis of specially polarized x rays. Routinely available, high-brightness synchrotron radiation sources can now provide naturally collimated x rays that can be easily manipulated by special x-ray optics to generate energy-tunable as well as polarization-tunable x-ray beams. The recent developments in x-ray phase retarders and multiple-Bragg-beam interference have allowed complete analyses of general elliptical polarization of x rays from special optics and special insertion devices. In this article, we will review these recent advances, especially in the area of the production and detection of circular polarization.

As for any electromagnetic wave,^{1–3} a general x-ray beam defined by

$$\mathbf{E}(\mathbf{r}, t) = (E_{\sigma}\hat{\sigma} + E_{\pi}e^{i\varepsilon}\hat{\pi})e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \quad (1)$$

can be linearly polarized if $\varepsilon = 0$ or 180° , circularly polarized if $\varepsilon = \pm 90^\circ$ and $E_{\sigma} = E_{\pi}$, and elliptically polarized for other values of ε , E_{σ} , and E_{π} . If $\varepsilon = \text{random values}$, then the x-ray beam is unpolarized or

has an unpolarized component. The polarization is in general characterized by three Stokes-Poincaré parameters (P_1, P_2, P_3):

$$\begin{aligned} P_1 &= \frac{I_{0^\circ} - I_{90^\circ}}{I_{0^\circ} + I_{90^\circ}} = \frac{E_\sigma^2 - E_\pi^2}{E_\sigma^2 + E_\pi^2} \\ P_2 &= \frac{I_{45^\circ} - I_{-45^\circ}}{I_{45^\circ} + I_{-45^\circ}} = \frac{2E_\sigma E_\pi \cos \epsilon}{E_\sigma^2 + E_\pi^2} \\ P_3 &= \frac{I_+ - I_-}{I_+ + I_-} = \frac{2E_\sigma E_\pi \sin \epsilon}{E_\sigma^2 + E_\pi^2} \end{aligned} \quad (2)$$

which represent the degrees of the 0° and 90° [σ (perpendicular) and π (parallel)] linear polarization, the $\pm 45^\circ$ -tilted linear polarization, and the left- and right-handed circular polarization, respectively.¹⁻³ The unpolarized portion in the beam is characterized by its fraction P_0 of the total intensity, given by $P_0 = 1 - (P_1^2 + P_2^2 + P_3^2)^{1/2}$. The unpolarized component is generally related to the incoherency in the x-ray beam, where the phase between the σ and the π wavefields is not well-defined, and can exist only in partially coherent radiation.¹

43.2 LINEAR POLARIZERS

Today's high-brightness synchrotron sources provide natural, linearly polarized x rays on the orbital plane of the storage ring. The degree of linear polarization is usually better than from 90 to 95 percent. The actual degree of linear polarization is not important for most x-ray experiments, as long as it is known and unchanged during the course of an experiment. Certain types of synchrotron experiments, such as linear dichroism spectroscopy, magnetic scattering, and atomic and nuclear resonant scattering, do require or prefer a higher degree of linear polarization. One of the most demanding experiments is an optical activity measurement⁴ in which the plane of polarization is defined and its rotation is measured to a high precision.

Using a $2\theta = 90^\circ$ Bragg reflection (Fig. 1a) from a perfect crystal such as silicon, the natural linear polarization from a synchrotron source can be further enhanced. In general, the linear polarization P'_1 after a Bragg reflection in the vertical diffraction plane can be improved from the incoming polarization P_1 by a factor given by

$$P'_1 = \frac{(1 + P_1) - (1 - P_1)p}{(1 + P_1) + (1 - P_1)p} \quad (3)$$

where p is the polarization factor of the reflection: $p = \cos^2 2\theta$ for a kinematic and $p = |\cos 2\theta|$ for a dynamical Bragg reflection. Obviously, $P_1 \leq P'_1 \leq 1$, depending on the actual 2θ angle used. The efficiency of a 90° Bragg reflection polarizer is very high, only limited by the effective reflectivity, which is usually better than 50 percent for a kinematic crystal and at least 90 percent for a perfect dynamical crystal.

Another type of linear polarizer for x rays is based on the principle of Bormann transmission (Fig. 1b) in dynamical diffraction.⁵ When a strong Bragg reflection [e.g., Si (220)], is excited in the Laue transmission mode, one of the σ -wave field branches has a smaller-than-usual absorption coefficient and is thus preferentially transmitted through the crystal in the forward direction. This anomalously transmitted beam is therefore highly σ -polarized, and the Laue-diffracting crystal (typically with $\mu t \sim 10$) can be used as a linear polarizer. This method has the advantage that the insertion of the polarizer does not alter the beam path, which is a common and desirable feature in visible light optics. However, because of its extremely narrow angular acceptance range (a fraction of the reflection

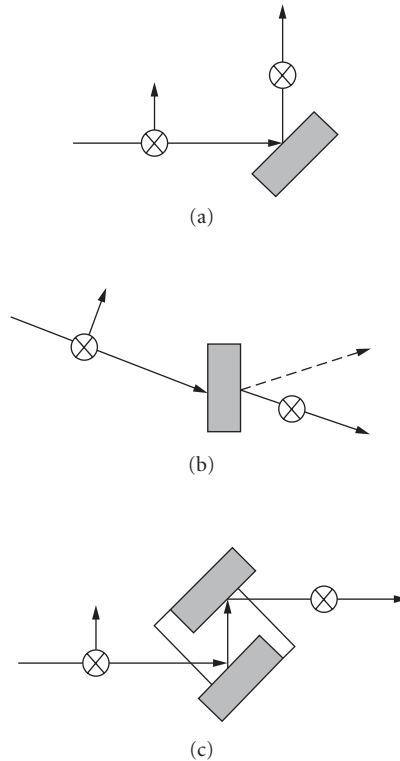


FIGURE 1 Possible linear polarizers for x rays: (a) 90° Bragg reflection; (b) Bormann transmission; and (c) multiple-bounce Bragg reflection with possible detuning.

angular width) and its intensity loss due to small but nonzero absorption, the Bormann transmission method is seldom used in practice as a linear polarizer.

The 90° Bragg reflection polarizer can be used in series to increase the purity of linear polarization even further. This can be easily achieved in practice using a channel-cut single crystal (Fig. 1c), in which the incident x-ray beam is diffracted multiple times between the two parallel crystal slabs. The highest degree of linear polarization (~ 99.99 percent) can be provided by a channel-cut crystal with a weak link between the two slabs.^{6,7} By slightly detuning the two crystal slabs, the intensity ratio of the π -polarized to the σ -polarized x rays can be suppressed by another factor of from 10 to 100. It has another advantage that the requirement of $2\theta = 90^\circ$ can be further relaxed and the polarizer is effectively tunable in a wider energy range (e.g., ± 20 percent) at 9 keV using Si (440).

Synchrotron experiments designed to study linear dichroism in materials sometimes require fast switching between two perpendicular linear polarization states in an incident x-ray beam. Although the switching can be done by a simple 90° rotation of a linear polarizer around the incident beam, a rapid switching can be achieved by a couple of methods. One is to use two simultaneously excited Bragg reflections whose scattering planes are perpendicular (or close to perpendicular) to each other. The other is to insert an x-ray half-wave phase retarder based on the effect of diffractive birefringence in dynamical diffraction from perfect crystals.⁸ In the latter method, if one uses a transmission-type phase retarder, the incident beam direction and position are not affected by the polarization switching, which is a significant advantage.

43.3 LINEAR POLARIZATION ANALYZERS

In general, any linear polarizer can be used as an analyzer for linear polarizations P_1 and P_2 . The simplest form of linear analyzer is to measure the 90° elastic scattering from an amorphous target⁹ or a powder sample.¹⁰ For better precisions, Bragg reflections with $2\theta = 90^\circ$ from single crystals are usually used. Again, one can achieve very high precision by means of a multiply bounced Bragg reflection from a channel-cut perfect crystal. The requirement for $2\theta = 90^\circ$ can be relaxed if one takes into account a scale factor involving the actual 2θ angle used in the measurement, as given in the following.

All these methods are based on measuring the intensity variation I_b as a function of the rotation angle χ of the scattering plane around the incident beam (Fig. 2) with respect to a reference polarization direction (e.g., the σ direction):

$$I_b(\chi) = \frac{1}{2} [(1+p) + (P_1 \cos 2\chi + P_2 \sin 2\chi)(1-p)] \quad (4)$$

where p is again the polarization factor of the scattering process as defined in Eq. (3). For $2\theta = 90^\circ$, $p = 0$.

It is straightforward to show that for an arbitrary $2\theta \neq 0$, the incident linear polarizations P_1 and P_2 can be obtained by measuring two difference-over-sum ratios,

$$\begin{aligned} P_1(2\theta) &= \frac{I_b(0^\circ) - I_b(90^\circ)}{I_b(0^\circ) + I_b(90^\circ)} \\ P_2(2\theta) &= \frac{I_b(45^\circ) - I_b(-45^\circ)}{I_b(45^\circ) + I_b(-45^\circ)} \end{aligned} \quad (5)$$

and using a proper scale factor involving the actual 2θ ,

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \frac{1+p}{1-p} \begin{bmatrix} P_1(2\theta) \\ P_2(2\theta) \end{bmatrix} \quad (6)$$

Equations (5) and (6) are useful when a single Bragg reflection is used for linear polarization analyses over a wide energy range.^{11,12}

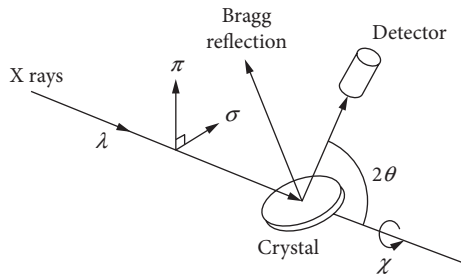


FIGURE 2 Bragg reflection from a crystal serves as a linear polarization analyzer.

43.4 PHASE PLATES FOR CIRCULAR POLARIZATION

In principle, circular polarization can be produced from linearly polarized radiation by one of the following three effects (Fig. 3): circular absorption dichroism, linear birefringence or double refraction, and reflection birefringence analogous to Fresnel rhomb. The last two effects only work with linearly polarized radiation, while the first effect can be used to convert both linear and unpolarized radiation.

Circular absorption dichroism is an effect of differential absorption between the left- and right-handed circular polarizations. For x rays, the natural circular dichroism in materials is a very weak effect. To date the only x-ray circular dichroism that has been studied extensively is the magnetic

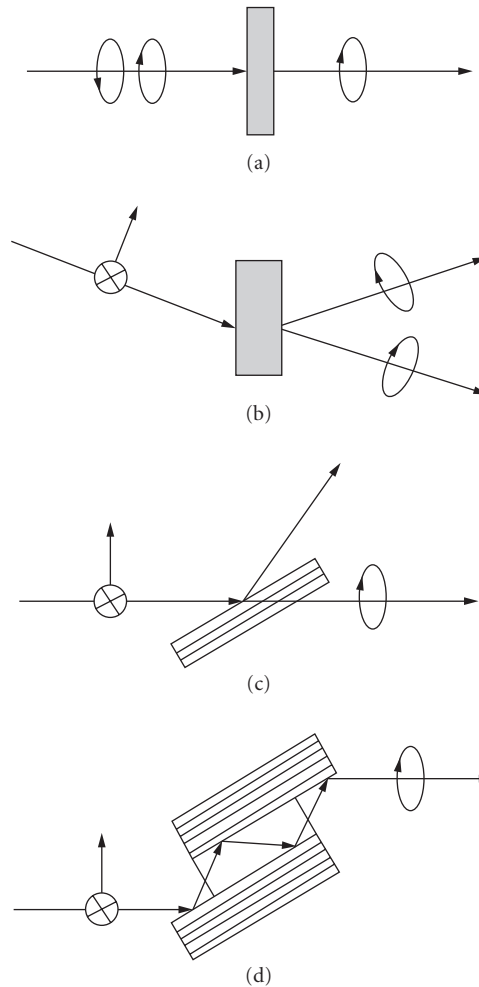


FIGURE 3 Circular phase plates for x rays based on: (a) circular absorption dichroism; (b) linear birefringence in Laue geometry; (c) linear birefringence in Bragg transmission geometry; and (d) Fresnel rhomb in Bragg reflection geometry.

circular dichroism in magnetic materials. However, the largest difference experimentally observed in absorption coefficients between the left- and the right-handed circular polarization is on the order of 1 percent, which is too small for any practical use as an x-ray circular polarizer.¹³

Fresnel rhomb is based on the principle that the reflected waves from the surface of a material can have different phase shifts between two orthogonal linear states, σ and π . The x-ray analog of this occurs at a Bragg reflection from a crystal and arises from the difference in the intrinsic angular widths (Darwin widths) of the σ and the π diffracted waves.^{14,15} The width for the π polarization is smaller than that for the σ polarization. Because of a continuous phase shift of 180° from one side of the reflection width to the other, a difference in the phase shift can be obtained between the σ and the π polarizations if the angular position is selected to be near the edge of the angular width. Experimentally, a multiply bounced Bragg reflection is needed to make a $\pm 90^\circ$ phase shift. The phase shift is independent of the crystal thickness, but this method requires a highly collimated incident beam, about 1/10 of the reflection width, which is usually the limiting factor for its throughput.¹⁶

The linear birefringence or double refraction effect relies on the difference in the magnitudes of the wavevectors of two orthogonal linear polarization states, σ and π , when a plane wave travels through a crystalline material. Because of this difference, a phase shift between the σ and the π traveling waves can be accumulated through the thickness of the birefringent material:¹⁷

$$\Delta = 2\pi(K_\sigma - K_\pi)t \quad (7)$$

where t is the thickness, and K_σ and K_π are the magnitudes of the wavevectors inside the crystal for the σ and π wavefields, respectively. When the phase shift Δ reaches $\pm 90^\circ$, circularly polarized radiation is generated, and such a device is usually termed a *quarter-wave phase plate* or a *quarter-wave phase retarder*.

For x rays, large birefringence effects exist near strong Bragg reflections in relatively perfect crystals. Depending on the diffraction geometry, one can have three types of transmission birefringence phase retarders: Laue transmission, Laue reflection, and Bragg transmission, as illustrated in Fig. 3b and c. The Laue reflection-type^{9,18} works at full excitation of a Bragg reflection, while the Laue and the Bragg transmission types^{19–22} work at the tails of a Bragg reflection, which has the advantage of a relaxed angular acceptance. In the past few years, it has been demonstrated that the Bragg transmission-type phase retarders are very practical x-ray circular polarizers. With good-quality diamond single crystals, such circular phase-retarders can tolerate a larger angular divergence and their throughputs can be as high as 0.25, with a degree of circular polarization in the range of from 95 to 99 percent. The handedness of the circular polarization can be switched easily by setting the diffracting crystal to either side of the Bragg reflection rocking curve. There have been some excellent review articles^{20–22} in this area and the reader is referred to them for more details.

43.5 CIRCULAR POLARIZATION ANALYZERS

Circular polarization P_3 of an x-ray beam can be qualitatively detected by magnetic Compton scattering¹⁸ and by magnetic circular dichroism. In general, these techniques are not suitable for quantitative polarization determination because these effects are relatively new and because of the uncertainties in the materials themselves and in the theories describing the effects.

Two methods have been developed in the past few years for quantitative measurements of circular polarization in the x-ray regime. One is to use a quarter-wave phase plate to turn the circular polarization into linear polarization which can then be measured using a linear polarization analyzer (Fig. 4a), as described in the previous sections. This method is entirely analogous to the similar techniques in visible optics and has been used by a couple of groups to characterize special insertion devices and quarter-wave phase plates.^{20,23}

Multiple-beam Bragg diffraction (MBD) is the other way to measure the degree of circular polarization of x rays (Fig. 4b). This technique makes use of the phase shift δ between the σ and the π wave

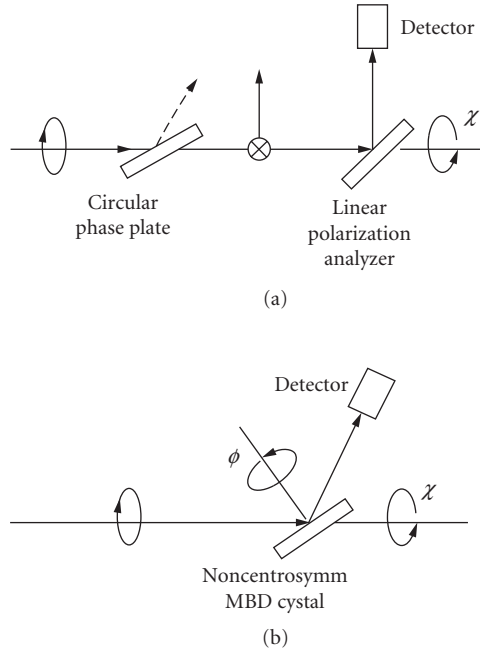


FIGURE 4 Two types of optics for analyzing circularly polarized x rays: (a) circular phase plate plus a linear polarization analyzer and (b) multiple-beam Bragg diffraction from a noncentrosymmetric crystal.

fields that arises from the phase-sensitive interference²⁴ and possible polarization mixing in an MBD process in a single crystal.²⁵ The phase shift δ is insensitive to crystal perfection and thickness as well as x-ray wavelength, since it is strictly determined by the crystal structure factors. Thus the MBD method has a broad applicable energy range and a good tolerance in the angular divergence of an x-ray beam.

Multiple-beam Bragg diffraction is also called simultaneous Bragg reflection, or *Umweganregung* (detour), as first noted by Renninger.²⁶ It occurs when two or more sets of atomic planes satisfy the Bragg's law simultaneously inside a crystal. A convenient way (Fig. 5) of realizing a multiple-beam

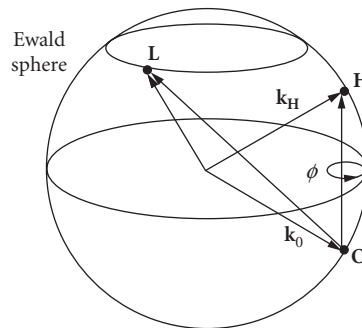


FIGURE 5 Ewald sphere construction showing a multiple-beam diffraction geometry.

reflection condition is exciting first one Bragg reflection, say $\mathbf{H}$, then rotating the diffracting crystal around $\mathbf{H}$ to bring another reciprocal node, say $\mathbf{L}$, onto the Ewald sphere.²⁷ The rotation around the $\mathbf{H}$ is represented by the azimuthal angle ϕ . It has been known^{28–30} that the simultaneously excited Bragg waves interfere with each other and give rise to a diffracted intensity that is sensitive to the *structural phase triplet* $\delta = \arg(F_{\mathbf{L}}F_{\mathbf{H-L}}/F_{\mathbf{H}})$, where $F_{\mathbf{H}}$ is the structure factor of the main reflection $\mathbf{H}$ and $F_{\mathbf{L}}$ and $F_{\mathbf{H-L}}$ are the structure factors of the detoured reflections $\mathbf{L}$ and $\mathbf{H-L}$. Another effect that exists in MBD is polarization mixing due to the intrinsic double Thomson scattering process,^{31,32} which causes a part of the π wave field amplitude to be scattered into the σ channel and vice versa. The combination of these two effects makes the MBD intensity sensitive to the phase difference between the σ and the π wave fields and thus to the circular polarization in the incident x-ray beam.

For maximum sensitivity to circular polarizations, a noncentrosymmetric crystal structure can be chosen as the analyzer and an MBD reflection with $\delta = \pm 90^\circ$ can be used. A complete determination of all Stokes-Poincaré parameters can be obtained either by measuring *three* independent MBD profiles^{25,33} and solving for (P_1, P_2, P_3) , or by separate measurements^{11,12,34} of P_1 and P_2 . This latter method provides a very convenient way for complete determination of Stokes-Poincaré parameters. Once the analyzer crystal is oriented and the proper multiple reflection is found, a complete determination of (P_1, P_2, P_3) , requires four rocking curve measurements of the integrated 2-beam intensities, $I_b(0^\circ)$, $I_b(90^\circ)$, $I_b(-45^\circ)$, and $I_b(+45^\circ)$, at $\chi = 0^\circ, 90^\circ, -45^\circ$, and 45° , respectively, and two rocking curve measurements of the MBD intensities, $I(+\Delta\phi)$ and $I(-\Delta\phi)$, at each side of the multiple reflection peak. The (P_1, P_2, P_3) are then obtained by taking the three pairs of difference-over-sum ratios,¹² with a scale factor for P_3 involving the multiple-beam structure factors.

We would like to point out that for an MBD reflection to have maximum sensitivity to circular polarization, $\delta = \pm 90^\circ$ is desired. This condition can be conveniently satisfied by choosing a noncentrosymmetric crystal such as GaAs. However, as we have previously shown,³² because of the *dynamical* phase shift *within* the full excitation of a multiple reflection, an MBD in a centrosymmetric crystal such as Si or Ge can also be sensitive to circular polarization. The only drawback is that it requires an extremely high angular collimation. With the arrival of low-emittance undulator-type sources, one should keep in mind that the use of Si or Ge crystals for MBD circular polarimetry may be feasible for well-collimated x-ray beams.

43.6 ACKNOWLEDGMENTS

The author is grateful to B. W. Batterman, K. D. Finkelstein, S. Shastri, and D. Walko for many useful discussions and collaborations. This work is supported by the National Science Foundation through CHES under Grant No. DMR-97-13424.

43.7 REFERENCES

1. J. D. Jackson, *Classical Electrodynamics*, 2nd ed., John Wiley and Sons, New York, 1975.
2. D. J. Griffiths, *Introduction to Electrodynamics*, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
3. M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon, New York, 1983.
4. M. Hart and A. R. D. Rodrigues, "Optical Activity and the Faraday Effect at X-Ray Frequencies," *Phil. Mag.* **B43**:321 (1981).
5. B. W. Batterman and H. Cole, "Dynamical Diffraction of X Rays by Perfect Crystals," *Rev. Mod. Phys.* **36**:681 (1964).
6. M. Hart, "X-Ray Polarization Phenomena," *Phil. Mag.* **B38**:41 (1978).
7. D. P. Siddons, M. Hart, Y. Amemiya, and J. B. Hastings, "X-Ray Optical Activity and the Faraday Effect in Cobalt and Its Compounds," *Phys. Rev. Lett.* **64**:1967 (1990).
8. C. Giles, C. Malgange, J. Goulon, C. Vettier, F. de Bergivin, A. Freund, P. Elleaume, E. Dartyge, A. Fontaine, C. Giorgetti, and S. Pizzini, "X-Ray Phase Plate for Energy Dispersive and Monochromatic Experiments," *SPIE Proceedings* **2010**:136 (1993).

9. J. A. Golovchenko, B. M. Kincaid, R. A. Levesque, A. E. Meixner, and D. R. Kaplan, "Polarization Pendellosung and the Generation of Circularly Polarized X Rays with a Quarter-Wave Plate," *Phys. Rev. Lett.* **57**:202 (1986).
10. G. Materlik and P. Suortti, "Measurements of the Polarization of X Rays from a Synchrotron Source," *J. Appl. Cryst.* **17**:7 (1984).
11. Q. Shen and K. D. Finkelstein, "A Complete Characterization of X-Ray Polarization State by Combination of Single and Multiple Bragg Reflections," *Rev. Sci. Instrum.* **64**:3451 (1993).
12. Q. Shen, S. Shastri, and K. D. Finkelstein, "Stokes Polarimetry for X Rays Using Multiple-Beam Diffraction," *Rev. Sci. Instrum.* **66**:1610 (1995).
13. K. D. Finkelstein and Q. Shen, "Feasibility of Using Magnetic Circular Dichroism to Produce Circularly Polarized X Rays," unpublished.
14. O. Brummer, Ch. Eischmidt, and H. Hoche, "Polarization Phenomena of X Rays in the Bragg Case," *Acta Crystallogr.* **A40**:394 (1984).
15. B. W. Batterman, "X-Ray Phase Plate," *Phys. Rev.* **B45**:12677 (1992).
16. S. Shastri, K. D. Finkelstein, Q. Shen, B. W. Batterman, and D. Walko, "Undulator Test of a Bragg Reflection Elliptical Polarizer at ~ 7.1 keV," *Rev. Sci. Instrum.* **66**:1581 (1994).
17. V. Belyakov and V. Dmitrienko, "Polarization Phenomena in X-Ray Optics," *Sov. Phys. Uspekhi.* **32**:697 (1989).
18. D. M. Mills, "Phase-Plate Performance for the Production of Circularly Polarized X Rays," *Nucl. Instrum. Meth.* **A266**:531 (1988).
19. K. Hirano, K. Izumi, T. Ishikawa, S. Annaka, and S. Kikuta, "An X-Ray Phase Plate Using Bragg Case Diffraction," *Jpn. J. Appl. Phys.* **30**:L407 (1991).
20. C. Giles, C. Malgrange, J. Goulon, F. de Bergivin, C. Vettier, A. Fontaine, E. Dartyge, S. Pizzini, F. Baudelet, and A. Freund, "Perfect Crystal and Mosaic Crystal Quarter-Wave Plates for Circular Magnetic X-Ray Dichroism Experiments," *Rev. Sci. Instrum.* **66**:1549 (1995).
21. K. Hirano, T. Ishikawa, and S. Kikuta, "Development and Application of X-Ray Phase Retarders," *Rev. Sci. Instrum.* **66**:1604 (1995).
22. Y. Hasegawa, Y. Ueji, K. Okitsu, J. M. Ablett, D. P. Siddons, and Y. Amemiya, "Tunable X-Ray Polarization Reflector with Perfect Crystals," *Acta Cryst.* **A55**:955–962 (1999).
23. T. Ishikawa, K. Hirano, and S. Kikuta, "Complete Determination of Polarization State in the Hard X-Ray Region," *J. Appl. Cryst.* **24**:982 (1991).
24. Q. Shen and K. D. Finkelstein, "Solving the Phase Problem with Multiple-Beam Diffraction and Elliptically Polarized X Rays," *Phys. Rev. Lett.* **45**:5075 (1990).
25. Q. Shen and K. D. Finkelstein, "Complete Determination of X-Ray Polarization Using Multiple-Beam Bragg Diffraction," *Phys. Rev.* **B45**:5075 (1992).
26. M. Renninger, "Umweganregung, eine bisher unbeachtete Wechselwirkungserscheinung bei Raumgitterinterferenzen," *Z. Phys.* **106**:141 (1937).
27. H. Cole, F. W. Chambers, and H. M. Dunn, "Simultaneous Diffraction: Indexing Umweganregung Peaks in Simple Cases," *Acta Crystallogr.* **15**:138 (1962).
28. M. Hart and A. R. Lang, *Phys. Rev. Lett.* **7**:120–121 (1961).
29. Q. Shen, "A New Approach to Multi-Beam X-Ray Diffraction Using Perturbation Theory of Scattering," *Acta Crystallogr.* **A42**:525 (1986).
30. Q. Shen, "Solving the Phase Problem Using Reference-Beam X-Ray Diffraction," *Phys. Rev. Lett.* **80**:3268–3271 (1998).
31. Q. Shen, "Polarization State Mixing in Multiple-Beam Diffraction and Its Application to Solving the Phase Problem," *SPIE Proceedings* **1550**:27 (1991).
32. Q. Shen, "Effects of a General X-Ray Polarization in Multiple-Beam Bragg Diffraction," *Acta Crystallogr.* **A49**:605 (1993).
33. K. Hirano, T. Mori, A. Iida, R. Colella, S. Sasaki, and Q. Shen, "Determination of the Stokes-Poincaré Parameters for a Synchrotron X-Ray Beam by Multiple Bragg Scattering," *Jpn. J. Appl. Phys.* **35**:5550–5552 (1996).
34. J. C. Lang and G. Srajer, "Bragg Transmission Phase Plates for the Production of Circularly Polarized X Rays," *Rev. Sci. Instrum.* **66**:1540–1542 (1995).

This page intentionally left blank

SUBPART

5.3

REFLECTIVE OPTICS

This page intentionally left blank

IMAGE FORMATION WITH GRAZING INCIDENCE OPTICS

James E. Harvey

*CREOL, The College of Optics and Photonics
University of Central Florida
Orlando, Florida*

44.1 GLOSSARY

ACV	autocovariance function
AUDPSF	average unregistered detected point spread function
AXAF	Advanced X-Ray Astrophysical Facility
DPSF	<i>detected</i> point spread function
ESA	European Space Agency
EUV	extreme ultraviolet
EUVE	extreme ultraviolet explorer
GOES	Geostationary Orbiting Environmental Satellite
GPSF	geometrical point spread function
HH	hyperboloid-hyperboloid
HPR	half power radius
MTF	modulation transfer function
OFOV	operational field-of-view
PSF	point spread function
RDPSF	registered detected point spread function
ROSAT	Rontgensatellit
SAX	Italian x-ray astronomy satellite
SSPSF	surface scatter point spread function
SXI	solar x-ray imager
WS	Wolter-Schwarzschild

44.2 INTRODUCTION TO X-RAY MIRRORS

Conventional mirrors have traditionally exhibited useful reflectances at x-ray wavelengths only for grazing incidence angles. The first two-dimensional image produced by deflecting x rays in a controlled manner was obtained by Kirkpatrick and Baez in 1948 with two grazing incidence mirrors as

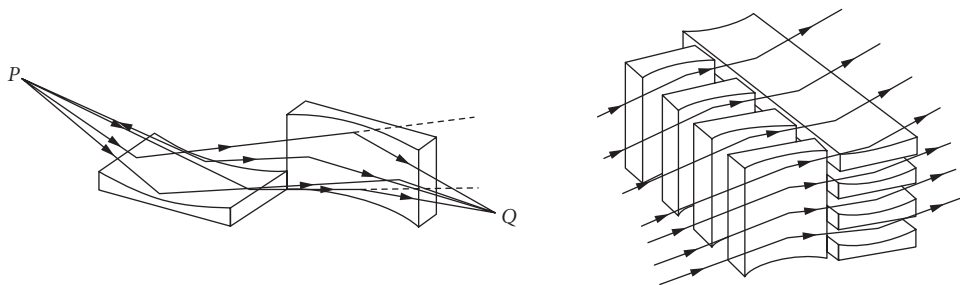


FIGURE 1 (a) The Kirkpatrick-Baez telescope consists of two orthogonal grazing incidence parabolic sheet mirrors and (b) a multiple stack of several mirrors can be used to substantially increase the collecting area.

illustrated in Fig. 1a.¹ The extremely small collecting area of such an imaging system can be alleviated by constructing the multiplate Kirkpatrick-Baez mirror of Fig. 1b.

In 1952 Hans Wolter published a paper in which he discussed several rotationally symmetric grazing incidence x-ray telescope systems.² The Wolter Type I telescope consists of a coaxial paraboloid (primary mirror) and hyperboloid (secondary mirror) as illustrated in Fig. 2a. The focus of the paraboloid is coincident with the rear focus of the hyperboloid and the reflection occurs on the inside of both mirrors. The Wolter Type II telescope also consists of a coaxial paraboloid and hyperboloid. However, the focus of the paraboloid is coincident with the front focus of the hyperboloid and the reflection occurs on the inside of the paraboloid and the outside of the hyperboloid. This system, illustrated in Fig. 2b, is the grazing incidence analog to the classical Cassegrain telescope. The Wolter Type III telescope consisting of a grazing incidence paraboloid and ellipsoid is illustrated in Fig. 2c. The focus of the paraboloid is coincident with the front focus of the ellipsoid, and the reflection occurs on the outside of the paraboloid and the inside of the ellipsoid.

The Wolter Type I telescope typically has a grazing angle of less than a degree and is used for hard x rays (greater than 1 keV). The Wolter Type II telescope typically has a grazing angle of approximately 10 degrees and is used for soft x rays and the extreme ultraviolet (EUV). The Wolter Type III optical design is not practical for astronomical telescopes; however, a finite-conjugate version of it is frequently used for x-ray microscopes.³ (See Chap. 51.)

The grazing incidence optical configurations are all free of spherical aberration; however, they exhibit severe field curvature, coma, and astigmatism. They are also cumbersome and difficult to fabricate and align, and scattering effects from imperfectly polished surfaces severely degrade image quality for these very short wavelengths. Aschenbach has written a nice review of these scattering effects in x-ray telescopes.⁴

Primarily due to much improved optical surface metrology capabilities, the conventional optical fabrication techniques of grinding and polishing glass substrates are resulting in major advances in the resolution of grazing incidence x-ray telescopes.^{5,6} The European ROSAT (Rontgensatellit) telescope⁷ consisting of four nested Wolter Type I grazing incidence telescopes provided substantial improvement in both effective collecting area and resolution over the Einstein Observatory which was launched in 1978,^{8,9} and NASA's Chandra Observatory, called the Advanced X-Ray Astrophysical Facility (AXAF) during the years it was being designed and fabricated, has demonstrated that the technology exists to produce large Wolter Type I grazing incidence x-ray telescopes with subarcsecond resolution. This progress in grazing incidence x-ray optics performance is illustrated in Fig. 3.¹⁰

The very smooth surfaces required of high-resolution x-ray optics have been achieved by tedious and time-consuming optical polishing efforts of skilled opticians. AXAF and ROSAT were thus very expensive to produce.

When such high resolution is not required, other optical materials and fabrication techniques may be applicable. The extreme ultraviolet explorer (EUV) was a NASA-funded astronomy mission intended to perform an all-sky survey in the 70 to 760 Å spectral region. The deep survey and spectroscopic portion of the mission utilized a Wolter Type II grazing incidence telescope built at the Space Sciences Laboratory

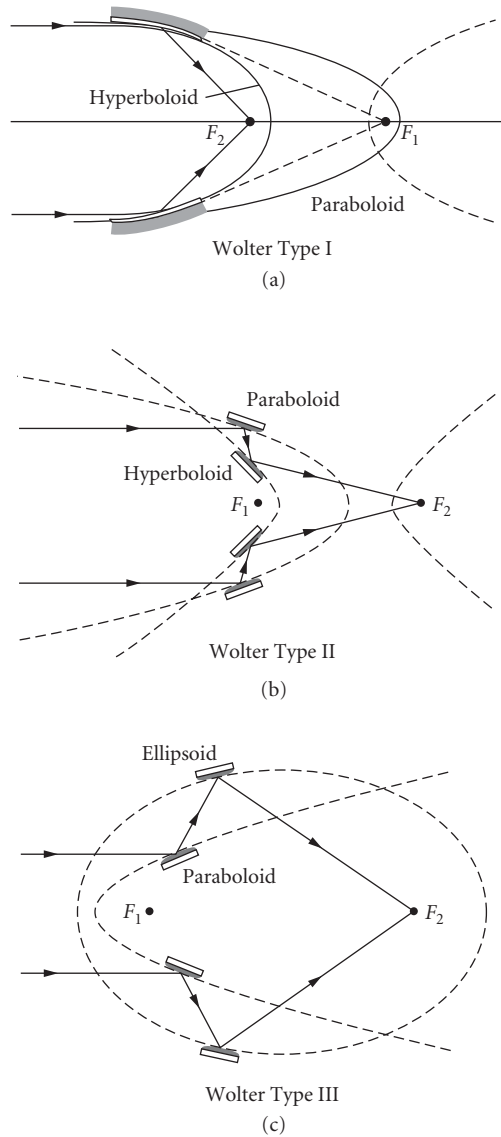


FIGURE 2 Hans Wolter's three classical, confocal, cylindrically symmetric grazing incidence x-ray telescope designs.

at the University of California, Berkeley. Its mirrors were fabricated from aluminum substrates by forging, rough machining, diamond turning to the desired figure, nickel plating, polishing, and coating with gold. An image half-power-width of approximately 1.5 arcsec was achieved.¹¹

Smooth x-ray mirror surfaces have been achieved without any labor-intensive polishing by merely dipping diamond-turned metal substrates in lacquer, then depositing tungsten or gold coatings to yield the desired high reflectance.^{12,13}

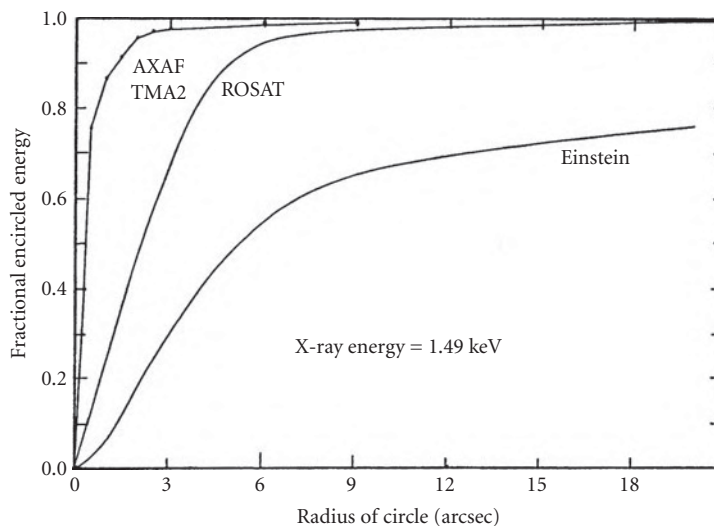


FIGURE 3 Fractional encircled energy plots at 1.49 keV for the Einstein Observatory, the European ROSAT telescope, and the AXAF technology mirror assembly (TMA2) are compared.

The apparent smoothness of the lacquer-coated surfaces and a desire for light weight and a high throughput or filling factor led Petre and Serlemitsos to develop the concept of tightly nested conical foil x-ray telescopes.^{14,15} These conical foil x-ray imaging mirrors are discussed in more detail in Chap. 48.

Still other novel optical fabrication concepts for grazing incidence x-ray optical surfaces include a variety of replication techniques.¹⁶ The Italian x-ray astronomy satellite (SAX) consisted of 30 nested coaxial mirrors electroformed over conical mandrels to a thickness ranging from 0.2 to 0.4 mm.¹⁷ And the European Space Agency (ESA) provided a dramatic increase in collecting area with its high throughput x-ray spectroscopy XMM mission featuring several modules of 58 tightly nested conical Wolter Type I telescopes fabricated with a metal/epoxy replication technique.¹⁸

Another attempt to avoid the problems of grazing incidence telescope configurations has resulted in the practice of depositing enhanced reflectance multiple high and low atomic number thin films onto more conventional Schwarzschild configurations (or other normal incidence designs). This rapidly emerging technology¹⁹⁻²¹ is currently effective for wavelengths longer than about 40 Å and then only for very narrow bandwidths.

These advances in the fabrication of x-ray optics, along with new technologies for the production of x rays such as synchrotron sources, free electron lasers, and laser generated x-ray plasmas are stimulating renewed efforts in the areas of x-ray/EUV astronomy, soft x-ray microscopy, and x-ray/EUV microlithography.

44.3 OPTICAL DESIGN AND RESIDUAL ABERRATIONS OF GRAZING INCIDENCE TELESCOPES

The two-mirror grazing incidence telescopes described above (Wolter Types I, II, and III) were axially symmetric, confocal, and were designed by following the principles of on-axis stigmatic imaging developed over 300 years ago by Newton, Gregory, and Cassegrain. Wolter did not

attempt to optimize his first designs for off-axis or finite conjugate imaging, but he did write a second paper later that same year formulating a completely aplanatic (corrected for both spherical aberration and coma) version of his designs; that is, the Wolter-Schwarzschild designs.²² Van Speybroeck and Chase used computerized ray tracing algorithms to empirically determine the parametric effects of varying designs on the imaging performance of the Wolter²³ and Wolter-Schwarzschild²⁴ Type-I telescopes. Their findings, published in 1972 and 1973, were extremely useful but lacked the identification and interpretation of conventional aberrations (i.e., defocus, spherical aberration, coma, astigmatism, field curvature, distortion, etc.). In 1977, Werner²⁵ attempted the computational optimization of a Wolter Type I telescope by relaxing the surface shape constraint to that of a generalized axial polynomial. This resulted in almost flat imaging response across the field of view but significantly compromised the on-axis performance. Also in 1977, Winkler and Korsch²⁶ published an apparently decisive and thorough formulation of two-mirror grazing incidence aberration theory. The results showed, however due to their limited precision, that any classical Wolter type telescope was *already* aplanatic. Even Wolter himself would not agree with this as evidenced by his second paper.²² In 1979, a paper by Cash et al.²⁷ concluding that standard, near normal incidence aberration theory could be *exactly* applied to grazing incidence optical elements. However, Korsch²⁸ showed (even with his low precision) that there exists a first order coma term *not* present in normal aberration theory for a single mirror. Furthermore, Nariai²⁹ stated in 1987 that “*it is not possible to use ordinary aberration theory because the expansion of the aberration in series of powers on the height of the object and on the radius of the pupil does not converge, . . . etc.*” Nariai³⁰ later showed analytically that all aberrations in his own expansion must be integrated over the entire annular pupil, apparently, aberration *coefficients* in grazing incidence systems are themselves a function of pupil coordinates. Saha^{31–34} published several extensive theoretical and numerical analyses of the aberrations of generalized Wolter Type II as well as some Wolter Type I configurations. These papers provided much-needed insight and understanding concerning the aberrations of grazing incidence x-ray optical systems. The aberration theory of Wolter type x-ray telescopes is briefly reviewed in the Chap. 45.

A Wolter Type I grazing incidence x-ray telescope made up of a paraboloid and hyperboloid is illustrated in Fig. 4. The equation for a paraboloid with its vertex at z_p is given by

$$r_p^2 = 2r_p(z - z_p) \quad (1)$$

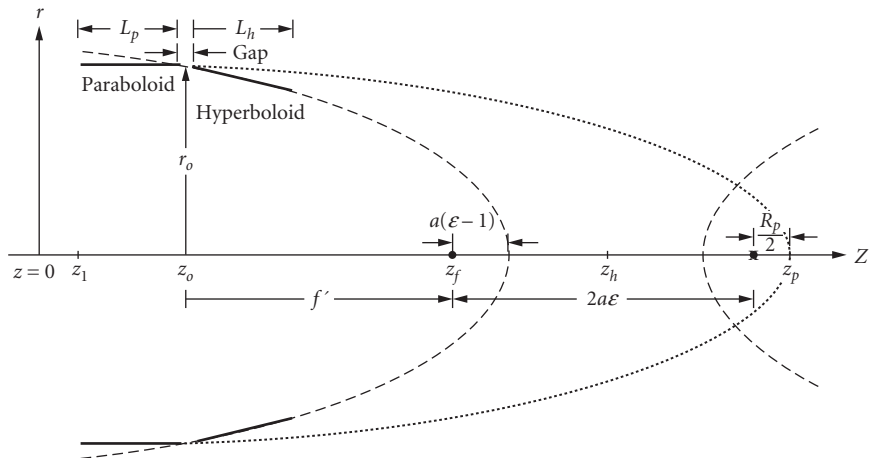


FIGURE 4 Wolter Type I grazing incidence telescope configuration.

where R_p is the paraboloid vertex radius of curvature and r_p is the radius of the paraboloid at the axial position z . The equation for a hyperboloid centered at z_h is given by

$$\frac{(z-z_h)^2}{a^2} - \frac{r_h^2}{b^2} = 1 \quad (2)$$

where a and b are the semimajor and semiminor axes of the hyperboloid. The eccentricity of the hyperboloid is determined by a and b

$$\varepsilon = \sqrt{\frac{a^2}{b^2} + 1} \quad (3)$$

The separation of the two hyperboloid foci is given by $2a\varepsilon$. If we superpose the rear hyperboloid focus with the paraboloid focus, the front hyperboloid focus becomes the system focus and $f' = z_f - z_o$ becomes the nominal focal length of the telescope. This is accomplished by positioning the origin of our coordinate system an arbitrary distance z_1 in front of the front edge of the paraboloid mirror and setting

$$z_p = z_1 + L_p + gap/2 + f' + 2a\varepsilon + R_p/2 \quad \text{and} \quad z_h = z_1 + L_p + gap/2 + f' + a\varepsilon \quad (4)$$

where L_p is the length of the paraboloid mirror and gap is the width of the gap between the paraboloid and the hyperboloid.

The optical prescription of a classical Wolter Type I x-ray telescope can thus be completely defined by the three independent parameters R_p , a , and b (or R_p , a , and ε). An optimized (maximized effective collecting area) Wolter Type I telescope can be obtained if we require the grazing angles of reflection from the paraboloid and the hyperboloid to be equal near their point of intersection. This constraint reduces the number of independent parameters defining the optical prescription to two.

For our purposes it is more convenient to choose the telescope radius at the intersection of the paraboloid and the hyperboloid, r_o , and the nominal focal length of the telescope f' as the parameters defining the optical prescription. The grazing angle at the joint is then given by

$$\alpha = \frac{1}{4} \arctan \left(\frac{r_o}{f'} \right) \quad (5)$$

The actual focal length, as measured from the system principal/nodal point, is slightly larger than the nominal focal length

$$f = f' + \frac{r_o^2}{2f'} \quad (6)$$

and the plate scale is the reciprocal of this focal length, expressed in arcsec per micrometers.

In addition to the telescope radius r_o and the nominal focal length f' , the remaining optical design parameters include the length of the paraboloid mirror L_p , the length of the hyperboloid mirror L_h , and the width of the gap between the two mirror elements. From these input parameters the actual dimensions of the mirror elements can be calculated as well as the obscuration ratio of the collecting aperture which determines both the geometrical collecting area and the diffraction-limited image characteristics.

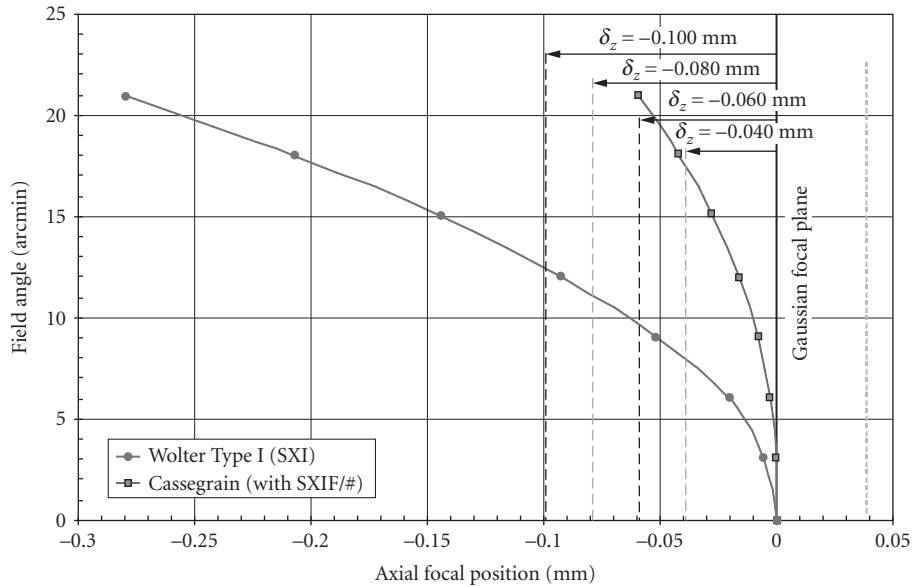


FIGURE 5 Field curvature exhibited by the SXI baseline design compared with that of a normal incidence Cassegrain telescope with the same focal ratio.

The classical Wolter Type I x-ray telescope design produces an ideal on-axis geometrical point image (zero spherical aberration); however, field curvature is a dominant limiting factor determining the off-axis performance of grazing incidence x-ray telescopes if a flat detector or focal plane is used. The severe field curvature of a classical Wolter Type I grazing incidence x-ray telescope is demonstrated in Fig. 5 by using the prescription of the solar x-ray imager (SXI) baseline design. SXI is a complementary, add-on instrument designed primarily for use on the GOES next generation geosynchronous weather satellites. However, its modular design is suitable for installation on many other spacecraft platforms. Its primary mission is to continuously observe the full solar disc, including coronal holes, active regions, flares, and coronal mass ejections.³⁵ The axial focal position was determined by minimizing the geometrical root-mean-square (rms) image size obtained from geometrical ray trace data.

If covering a significant field-of-view, the focal plane of such systems is frequently despaced to improve the off-axis performance; although, this results in a degraded (defocused) on-axis image. Geometrical optical performance from ray trace data is conveniently expressed in terms of rms image radius in arcsec. This quantity is calculated and plotted as a function of field angle for several different axial positions of the focal plane, see Fig. 6. Also shown for comparison is the performance curve that would be achieved with a curved detector conforming to the optimally curved focal surface.

Note that the curve for the best focal surface in Fig. 6 appears to have a linear and a quadratic component. This is consistent with findings of Van Speybroek and Chase.²³ For small field angles, the linear component dominates and will be associated with a comalike aberration.^{36,37} Similarly, the quadratic component of the curve will be associated with an astigmatism-like aberration.^{36,37} The curve corresponding to the Gaussian image plane also appears to consist of a linear and a quadratic component. The linear component (coma) is same as for the best focal surface as evidenced by the slope at small field angles. However, the quadratic component is significantly larger since it contains a contribution from both astigmatism and field curvature. Without attempting to develop any rigorous aberration theory, we will, throughout this chapter, refer to the image degradation (indicated by rms image size) that is linear with field angle as coma, and the degradation that is quadratic with field

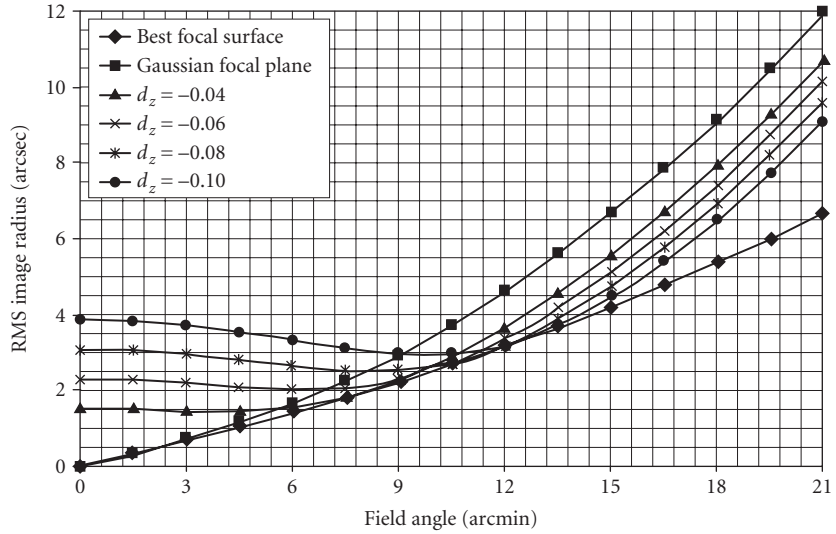


FIGURE 6 Geometrical performance of a classical Wolter Type I telescope for different axial positions of the focal plane.

angle as a combination of astigmatism and field curvature. Similarly, we will consider any image degradation on-axis to be caused by a combination of defocus and spherical aberration. Clearly, we are not trying to distinguish between various orders of aberrations. For example, the linear component of these curves includes all orders of linear coma. Likewise, the quadratic component includes third-order field curvature and astigmatism as well as all higher-order aberration terms that have a quadratic dependence on field angle; this includes the fifth-order aberration usually referred to as oblique spherical aberration.^{25,29–34} There are also, no doubt, cubic and higher-order contributions to the curves in Fig. 6; however, they do not appear to play a significant role for field angles less than 21 arcmin.

Despacing the focal plane of the classical Wolter Type I grazing incidence telescope clearly balances field curvature with defocus, thus improving the wide-field performance at the expense of the small-field performance. There are no additional design variables available to further correct or balance aberrations. This classical Wolter Type I design produces a stigmatic image on-axis in the Gaussian focal plane and has thus been used in virtually every x-ray stellar telescope built in the last 40 years, including the Einstein Observatory,³⁸ ROSAT,³⁹ and AXAF (the Chandra Observatory).⁴⁰

However, for wide-field x-ray imaging systems, stigmatic imaging on-axis is no longer an appropriate image quality requirement. Consider a solar physics application where the imaging system is a staring x-ray telescope pointed at the center of the sun. Sunspots or solar flares can appear anywhere on the solar disc; hence, the *field-weighted-average resolution element* as degraded by all error sources is an appropriate image quality criterion.⁴¹ This led to a departure from the conventional Wolter Type I grazing incidence x-ray telescope design which optimizes the on-axis resolution and is thus used for most stellar x-ray telescopes. Instead, a family of optimal hyperboloid-hyperboloid (HH) grazing incidence x-ray telescope designs was developed, where each member of the family is optimized for a different operational field-of-view (OFOV).⁴²

A schematic diagram of the resulting generalized Wolter Type I grazing incidence x-ray telescope is shown in Fig. 7. Note that the front focus of the primary mirror does *not* coincide with the rear focus of the secondary mirror as is the case with the classical Wolter Type I design. This *confocal delta* is indicated as the quantity Δ_{ps} in Fig. 7. Similarly, the system focal plane does *not* lie at the front focus of the secondary mirror. This displacement is indicated as Δf .

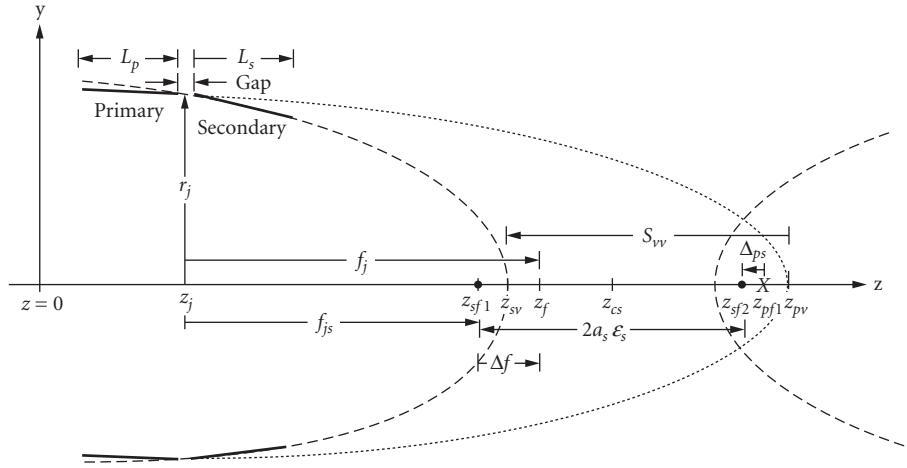


FIGURE 7 Hyperboloid-hyperboloid grazing incidence x-ray telescope design.

The above HH grazing incidence x-ray telescope thus has five design variables that can be varied in the design optimization process; the vertex radii of curvature of the two mirrors, their conic constants, and the vertex-to-vertex separation. For the SXI first-order design parameters a detailed comparison of the geometrical performance of the optimally despaced classical Wolter Type I design, the optimally despaced Wolter-Schwarzschild (WS) design, and the optimum HH design was performed. Figure 8 shows that the optimally despaced WS design and the optimum HH design always significantly

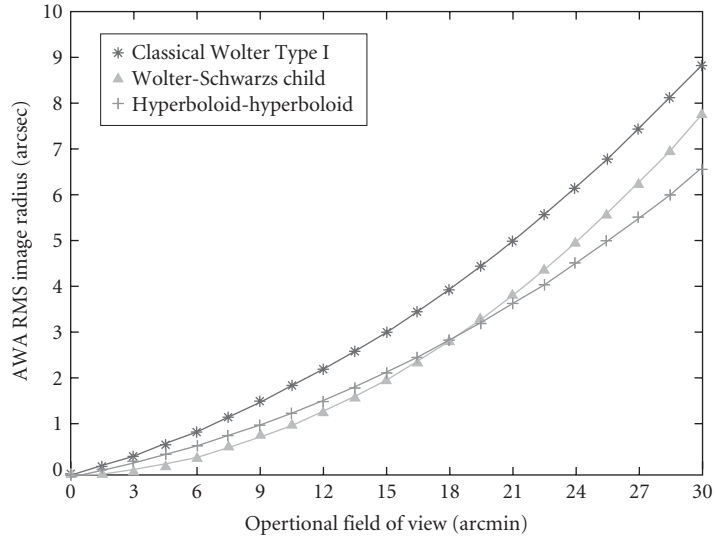


FIGURE 8 Comparison of the field weighted-average rms image radius versus OFOV for three different types of grazing incidence x-ray telescopes for wide-field imaging applications.

outperforms the optimally despaced classical Wolter Type I design. However, the optimally despaced WS design only outperforms the optimum HH design for OFOVs less than approximately 18 arcmin.⁴² Note that for OFOVs greater than approximately 18 arcmin, the *optimum HH design outperforms the optimally despaced WS design*.

44.4 IMAGE ANALYSIS FOR GRAZING INCIDENCE X-RAY OPTICS

A complete systems engineering analysis of x-ray telescope performance requires that we look at the effects of aperture diffraction, geometrical aberrations, surface scattering, and all other potential error sources such as assembly and alignment errors and metrology errors that appear in the mirror manufacturer's error budget tree as shown in Fig. 9. Detector effects must also be accurately modeled, particularly if it utilizes a staring mosaic detector array.

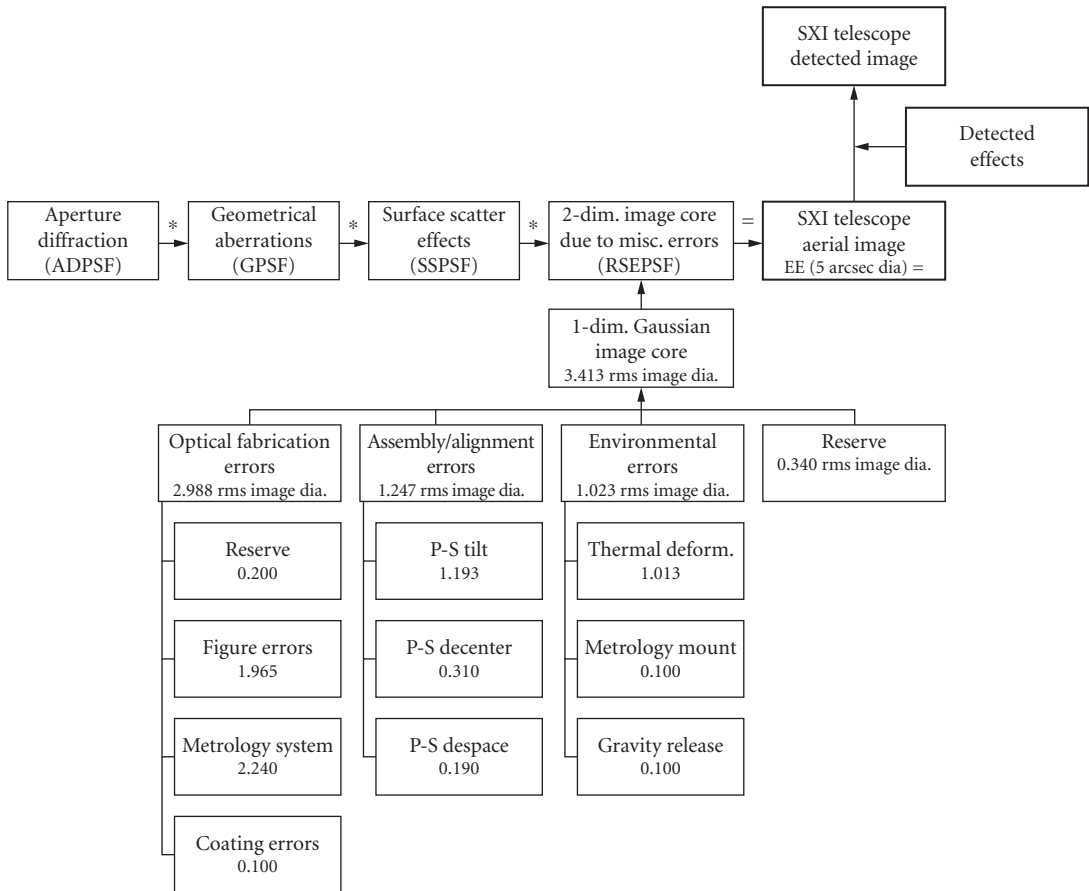


FIGURE 9 Error budget tree for manufacturing the SXI mirrors (* is a symbolic notation for the convolution operation).

Aperture diffraction effects are not necessarily negligible since the effects of the high obscuration ratios inherent to grazing incidence optics off-sets the effects of the very short x-ray wavelengths.^{43,44} Furthermore, most conventional, commercially available optical design and analysis codes do not adequately model the diffraction effects of highly obscured annular apertures. A semiempirical expression that does not require the high sampling density and associated computational problems when performing numerical convolutions can be used.⁴⁴

Likewise, most commercially available optical design and analysis codes do not provide aberration coefficients for highly obscured annular apertures. However, a high density spot diagram for each field angle of interest can be produced with a standard ray trace program (such as Code V or ZEMAX), the ray intercept data can be used to construct a ray *density* function in the focal plane which can be referred to as a geometrical point spread function (GPSF).

When light is reflected from an imperfect optical surface, the reflected radiation consists of a specularly reflected component and a diffusely reflected (or scattered) component. The scattered light behavior can be calculated from the surface characteristics (see Vol. I, Chap. 8, "Surface Scattering," by Eugene L. Church and Peter Z. Takacs). The surface autocovariance (ACV) function or the surface power spectral density (PSD) function completely determines the scattered light behavior. For the SXI program, we used NASA's EEGRAZE code to calculate the scatter profile for the SXI telescope. Either assumed or measured surface metrology data (PSD function) must be supplied as input for the EEGRAZE code. This scatter profile was then used to construct the surface scatter point spread function (SSPSF).

And finally, the effects of all other potential error sources appearing in the SXI mirror manufacturer's error budget tree were modeled as a single contribution to the final rms image core diameter. The aerial image produced by the x-ray telescope was then modeled by the *convolution* of the individual PSFs associated with the respective error sources as illustrated schematically in Fig. 10.⁴⁵

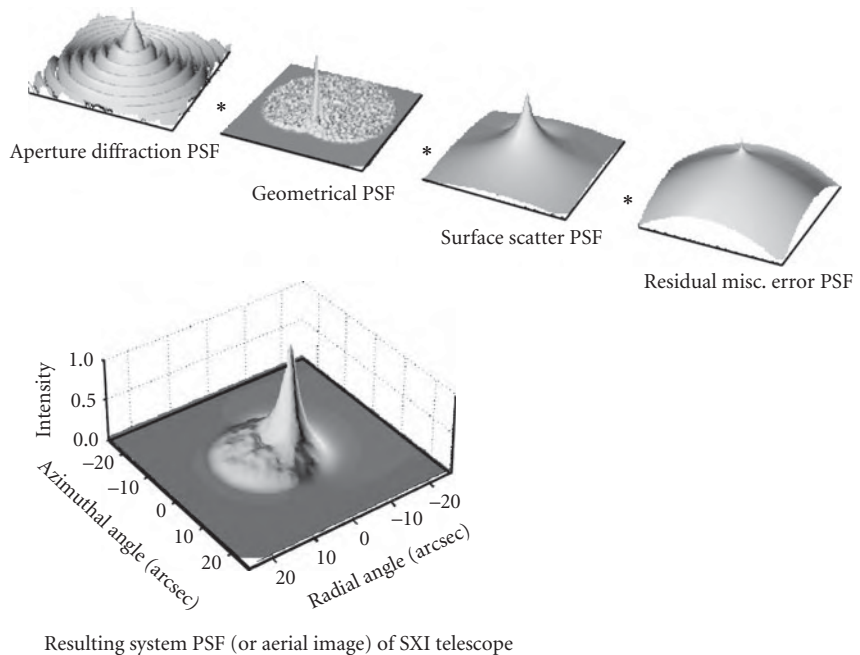


FIGURE 10 Illustration of the PSFs for the individual image degradation mechanisms and their resulting convolution (*aerial image*) of the SXI telescope for a field angle of 15 arcmin and a wavelength of 44.7 Å. (See also color insert.)

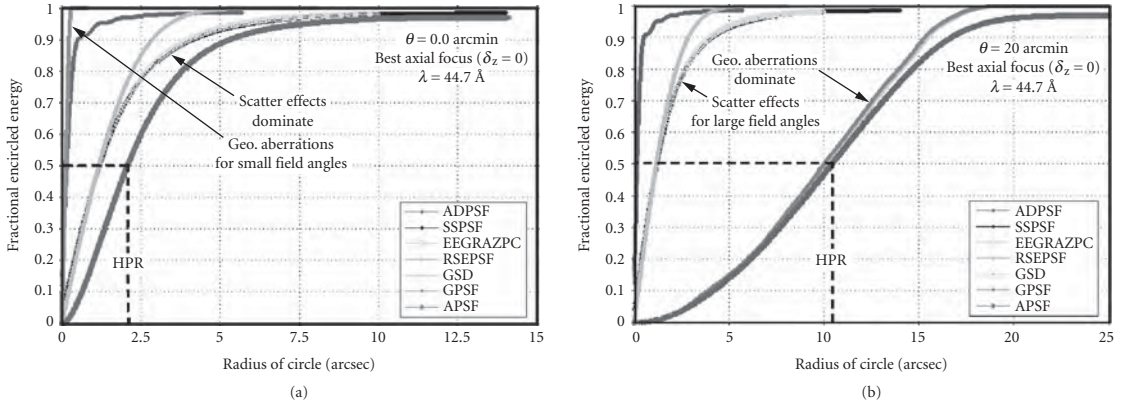


FIGURE 11 The fractional encircled energy of the aerial image and the four functions contributing to it provide insight into the image quality of a grazing incidence x-ray telescope. Note that scatter effects dominate geometrical aberrations for small field angles and geometrical aberrations dominate scatter effects for large field angles. (See also color insert.)

Fractional encircled energy plots of these four functions and the aerial image provide insight concerning the relative effects of the various image degradation mechanisms as shown in Fig. 11a and 11b. Note that for small field angles the scattering effects dominate geometrical aberrations, whereas for large field angles the geometrical aberrations dominate the scattering effects. Note also that the half power radius (HPR) of the on-axis aerial image is 2.0 arcsec and the HPR of the aerial image at a 20 arcmin field angle is 10.5 arcsec.

When a single detector is scanned over an aerial image, the detected image (in the scan direction) can be modeled by the convolution of the aerial image with the detector; or conversely, one can multiply the modulation transfer function (MTF) of the imaging system by the detector MTF. However, mosaic detector arrays employ a discrete sampling interval that causes these systems to exhibit a particular kind of *local shift variance* which causes the appearance of the reconstructed image to vary with the location of the aerial point spread function (PSF) relative to the sampling (i.e., pixel) grid.⁴⁶ The MTF approach to system performance analysis is thus not directly applicable to systems utilizing staring mosaic detector arrays.

Figure 12 illustrates the example of a Gaussian aerial PSF slightly larger than a detector pixel being sampled (averaging over each detector pixel) to produce a *detected* point spread function (DPSF).

An interpolation scheme is then used to reconstruct a smooth DPSF. This is done for three situations: (1) when the aerial PSF is precisely “registered” at the center of a detector pixel, (2) when the aerial PSF is positioned on the boundary between two detector pixels, and (3) when the aerial PSF is positioned at a point where four detector pixels meet. This registration error can result in as much as a 40 percent variation in the calculated HPR of the reconstructed DPSF.

For an application where the telescope is being operated as a staring telescope recording fine detail in an extended image (random location of aerial PSF on pixel), the “average unregistered” detected point spread function (AUDPSF) is given by the convolution of the registered detected point spread function (RDPSF) by the unit cell of the sampling grid.⁴⁷

Figure 13 illustrates the calculation of both the reconstructed registered DPSF and the reconstructed average unregistered DPSF. Since the aerial PSF is represented as a dense numerical array, the averaging over the individual pixels is referred to as a “binning” operation. Care is taken to precisely “register” the sampling detector grid by positioning it so as to maximize the signal produced by a given pixel. We then use a cubic interpolation technique to reconstruct the RDPSF. Finally, we convolve by the unit cell of the sampling grid to produce the AUDPSF.⁴⁷

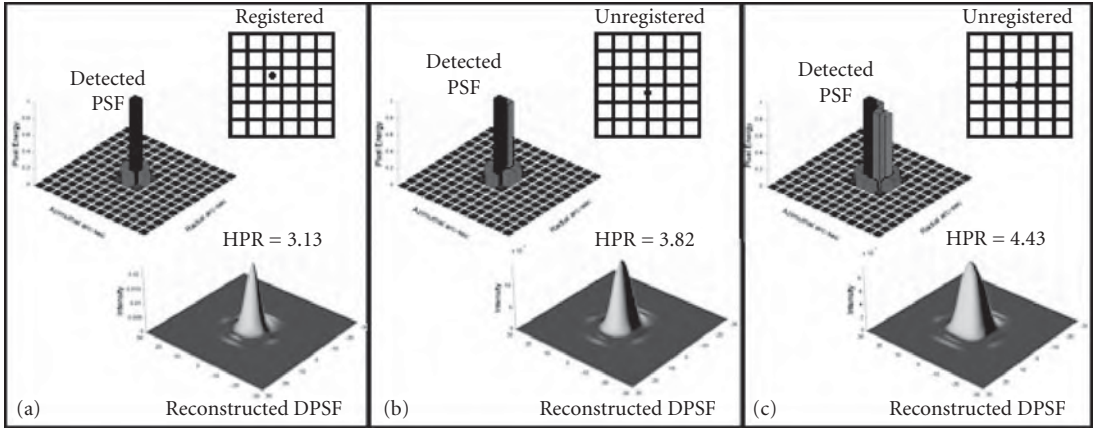


FIGURE 12 The DPSF and the reconstructed DPSF for: (a) precisely “registered” aerial PSF; (b) aerial PSF centered on boundary between two pixels; and (c) aerial PSF positioned where four pixels meet. (See also color insert.)

We can now calculate the field-weighted-average HPR of either the aerial image or the AUDPSF for any specific operational field-of-view (OFOV).

$$\text{HPR}_{fwa} = \frac{1}{A_T} \int_{\theta=0}^{\text{OFOV}} \text{HPR}(\theta) 2\pi\theta d\theta \quad (7)$$

where $A_T = \pi(\text{OFOV})^2$

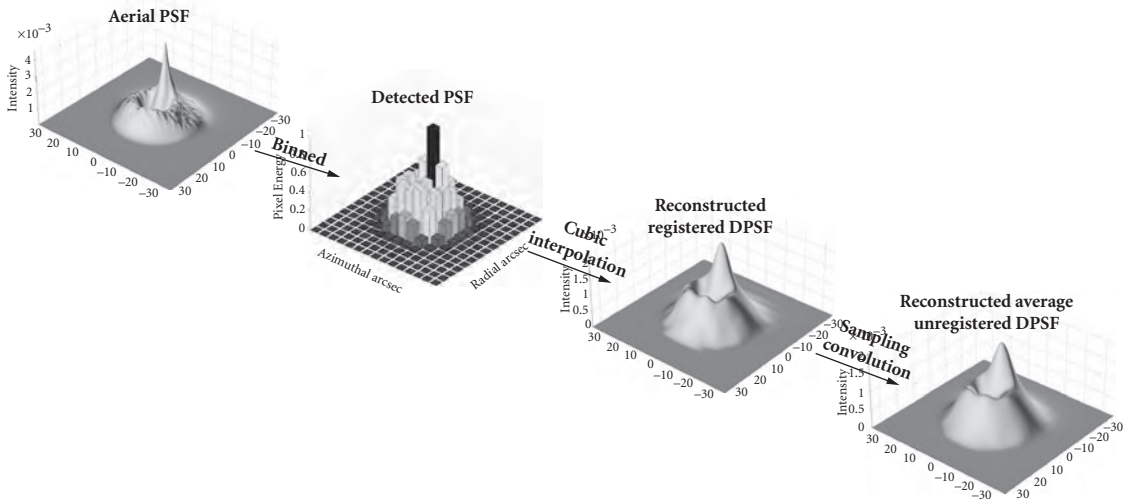


FIGURE 13 A graphical illustration of the numerical computation technique for modeling both the reconstructed “registered” DPSF and the reconstructed “average unregistered” DPSF is indicated. (See also color insert.)

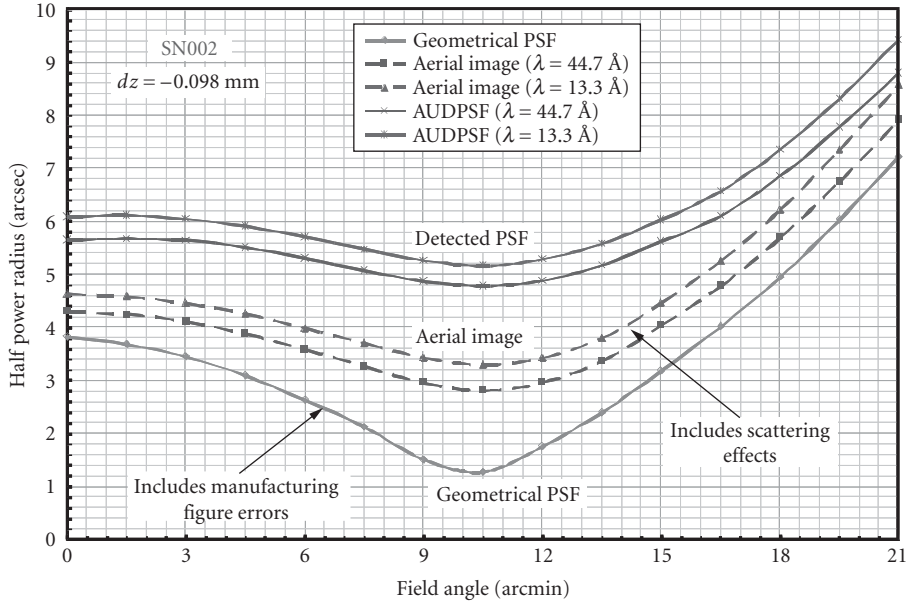


FIGURE 14 Comparison of the predicted HPR versus field angle of the geometrical PSF, the aerial image, and the average unregistered detected PSF for the SXI telescope at two different wavelengths (44.7 and 13.3 Å). (See also color insert.)

Defining a spatial (angular) resolution element as a circle of radius $\text{HPR}(\theta)$, we can also calculate the number of spatial resolution elements N in that OFOV

$$N = \text{number of res. ele.} = 2 \int_{\theta=0}^{\text{OFOV}} \frac{\theta}{\text{HPR}^2(\theta)} d\theta \quad (8)$$

Minimizing the HPR_{fwa} for a given OFOV maximizes the number of spatial resolution elements N , and thus also maximizes the amount of information contained in the image.

A comparison of the predicted HPR versus field angle of the geometrical PSF (geometrical aberrations only, as determined by ray trace analysis), the aerial image (including all system errors except detector effects), and the averaged unregistered detected PSF for the SN002 SXI telescope that was launched on GOES-13 is illustrated in Fig. 14.^{48,49} Predictions were performed for two different wavelengths, 44.7 and 13.3 Å. These curves provide insight concerning the relative effect of various error sources upon the image quality, and certainly demonstrate the inadequacy of merely performing a geometrical ray trace analysis of these grazing incidence x-ray telescopes. Similar graphs were obtained for each of the other “as-manufactured” SXI telescopes.⁵⁰

44.5 VALIDATION OF IMAGE ANALYSIS FOR GRAZING INCIDENCE X-RAY OPTICS

The SXI was launched on GOES-13 on May 24, 2006, and the “first light” image was recorded with on July 6, 2006. Figure 15 shows one of the first raw on-orbit images, which was described as “exquisite” by at least one solar physicist.

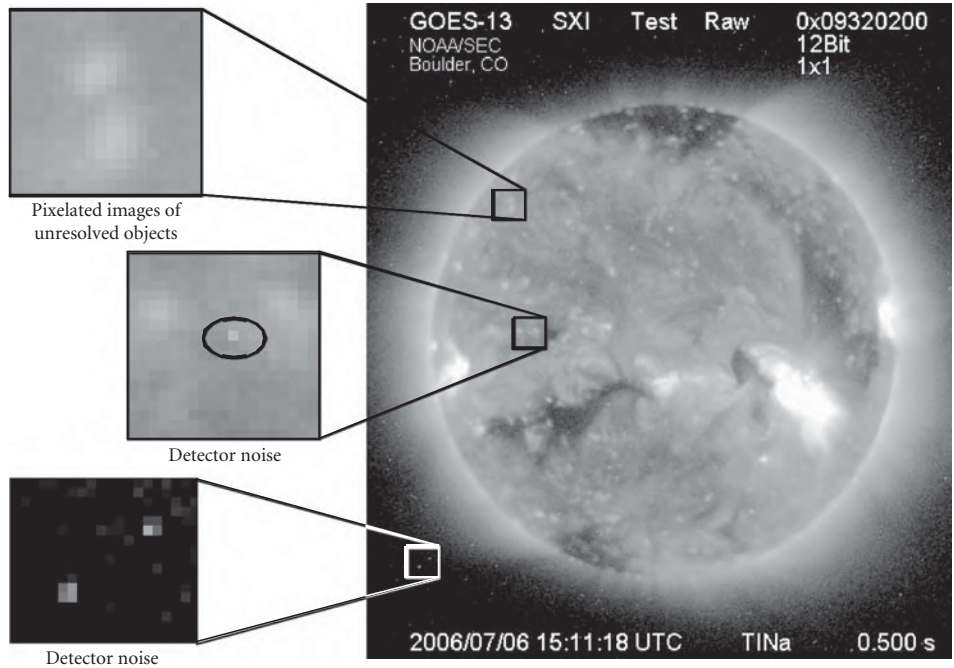


FIGURE 15 On-orbit solar image with three small regions highlighted and magnified for detailed inspection. This allows one to distinguish between images of unresolved bright features on the sun and merely detector noise. (See also color insert.)

Upon close inspection, it is reasonable to assume that the abundance of small features in Fig. 15 consist of the images of *unresolved bright spots* on the sun. This assumption becomes more obvious as we highlight and magnify small areas in the raw on-orbit image. As examples, we illustrate three separate regions of the solar image that we have highlighted and magnified. The first highlighted area shows two closely spaced pixelated images of unresolved features on the sun. Note that they are 6 to 8 pixels in diameter which is consistent with the numerical predictions of the detected point spread function in Fig. 13. The second highlighted area shows a very small bright spot which, when magnified, is obviously merely detector noise (a single detector pixel with unusually high responsivity). Similarly the third highlighted area depicts detector noise which should not be mistaken with actual bright spots in the solar corona as they would have to be at least as large as a point spread function.

In order to extract quantitative data from the raw on-orbit images to compare with our image quality predictions, the solar disc was divided into seven radial zones. The sizes of small prominent features in each radial zone were then measured (and normalized by the known size of the solar disc) and averaged. The apparent (measured) diameter of these features was assumed to be equal to twice their half power radius (HPR).⁵¹

These average-measured HPRs were then compared to the results of the detailed image quality predictions illustrated earlier by superposing these average values for each radial zone on the performance prediction curves presented in Fig. 14. This extremely simple and straightforward comparison of raw on-orbit experimental data with the detailed and exhaustive numerical predictions of our systems engineering analysis of image quality is displayed in Fig. 16. The excellent agreement of this quantitative data with image quality predictions provides an experimental validation of the detailed systems engineering analysis of image quality required for grazing incidence x-ray optics.

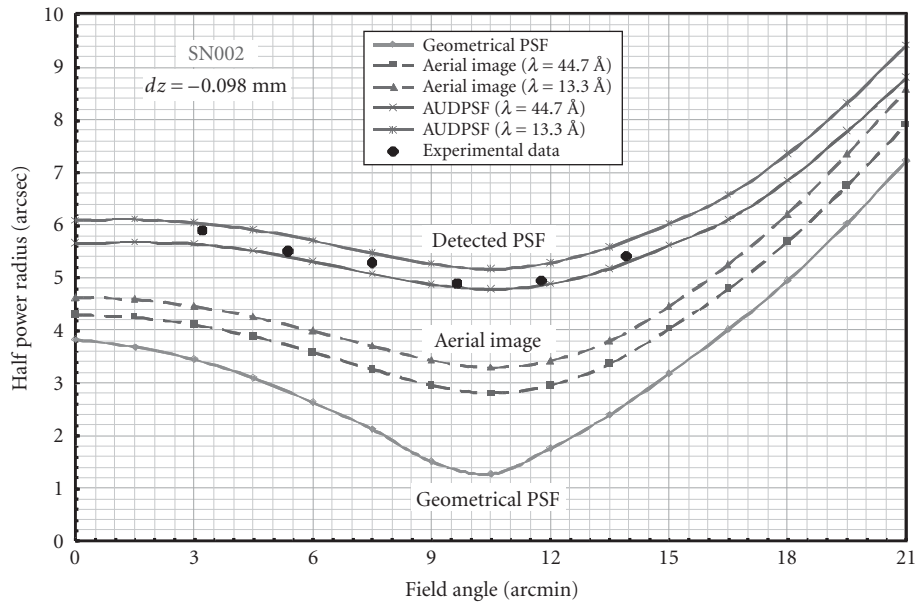


FIGURE 16 Experimental validation of an exhaustive systems engineering analysis of image quality for grazing incidence x-ray telescopes, including the modeling of surface scatter and detector effects. (See also color insert.)

44.6 REFERENCES

1. P. Kirkpatrick and A. V. Baez, "Formation of Optical Images by X-Rays," *J. Opt. Soc. Am.* **38**:776 (1948).
2. H. Wolter, "Mirror Systems with Glancing Incidence on Image Producing Optics for X-Rays," *Ann. Phys.* **10**:94 (1952).
3. J. E. Harvey, K. L. Lewotsky, and A. Kotha, "Performance Predictions of a Schwarzschild Imaging Microscope for Soft X-Ray Applications," *Opt. Eng.* **35**:2423–2436 (Aug. 1996).
4. B. R. Aschenbach, "X-Ray Telescopes," *Rep. Prog. Phys.* **48**:579–629 (1985), The Institute of Physics, Great Britain.
5. L. P. Van Speybroeck, "Grazing Incidence Optics for the U.S. High Resolution X-Ray Astronomy Program," *Opt. Eng.* **27**:1398–1403 (1988).
6. A. Slomba, R. Babish, and P. Glenn, "Mirror Surface Metrology and Polishing for AXAF/TMA," *Proc. SPIE* **597**:40 (1985).
7. B. R. Aschenbach, "Design, Construction, and Performance of the ROSAT High-Resolution Mirror Assembly," *Appl. Opt.* **27**:1404–1413 (1988).
8. R. Giacconi, et al., "The Einstein (HEAO 2) X-Ray Observatory," *Astrophys. J.* **230**:540 (1979).
9. L. P. Van Speybroeck, "Einstein Observatory (HEAO B) Mirror Design and Performance," *Proc. SPIE* **184**:2 (1979).
10. J. E. Harvey, "Recent Progress in X-Ray Imaging," presented at the *AIAA Space Programs and Technologies Conference*, Huntsville, AL, Sept. 1990.
11. S. Bowyer and J. Green, "Fabrication, Evaluation, and Performance of Machined Metal Grazing Incidence Telescopes," *Appl. Opt.* **27**:1414–1422 (1988).
12. R. C. Catura, E. G. Joki, D. T. Roethig, and W. J. Brookover, "Lacquer Coated X-Ray Optics," *Proc. SPIE* **640**:140–144 (1986).

13. J. A. Nousek, et al., "Diamond-Turned Lacquer-Coated Soft X-Ray Telescope Mirrors," *Appl. Opt.* **27**:1430–1432 (1988).
14. R. Petre and P. J. Serlemitsos, "Conical Imaging Mirrors for High-Speed X-Ray Telescopes," *Appl. Opt.* **24**:1833 (1985).
15. R. Petre, P. J. Serlemitsos, F. E. Marshall, K. Jahoda, and H. Kunieda, "In Flight Performance of the Broad-Band X-Ray Telescope," *Proc. SPIE* **1546**:72–81 (1991).
16. Y. Matsui, M. P. Ulmer, and P. Z. Takacs, "X-Ray and Optical Profiler Analysis of Electroformed X-Ray Optics," *Appl. Opt.* **27**:1558–1563 (1988).
17. O. Citterio, et al., "Optics for the X-Ray Imaging Concentrators Aboard the X-Ray Astronomy Satellite SAX," *Proc. SPIE* **830**:139 (1987).
18. W. Egle, H. Bulla, P. Kaufmann, B. Aschenbach, and H. Brauningner, "Production of the First Mirror Shell for ESA's XMM Telescope by Application of a Dedicated Large Area Replication Technique," *Opt. Eng.* **29**:1267 (1990).
19. E. Spiller, "Reflective Multilayer Coatings in the Far UV Region," *Appl. Opt.* **15**:2333 (1976).
20. A. B. C. Walker, Jr., T. W. Barbee, Jr., R. B. Hoover, and J. F. Lindblom, "Soft X-Ray Images of the Solar Corona with a Normal-Incidence Cassegrain Multilayer Telescope," *Science* **241**:1781 (1988).
21. N. M. Ceglio (ed.), Thirty-one papers presented at the 1991 SPIE conference on "Multilayer Optics for Advanced X-Ray Applications," *Proc. SPIE* **1547** (1991).
22. H. Wolter, "Generalized Schwarzschild Mirror Systems with Glancing Incidence as Optics for X-Rays," *Ann. Phys.* **10**:286 (1952).
23. L. P. Van Speybroeck and R. C. Chase, "Design Parameters of Paraboloid-Hyperboloid Telescopes for X-Ray Astronomy," *Appl. Opt.* **11**:440–445 (1972).
24. R. C. Chase and L. P. Van Speybroeck, "Wolter-Schwarzschild Telescopes for X-Ray Astronomy," *Appl. Opt.* **12**:1042–1044 (1973).
25. W. Werner, "Imaging Properties of Wolter I Type X-Ray Telescopes," *Appl. Opt.* **16**:764–773 (1977).
26. C. E. Winkler and D. Korsch, "Primary Aberrations for Grazing Incidence," *Appl. Opt.* **16**:2464–2469 (1977).
27. W. Cash, D. L. Sheeley, and J. H. Underwood, "Space Optics Imaging X-Ray Optics Workshop," *Proc. SPIE* **184**:228 (1979).
28. D. Korsch, *Reflective Optics*, Chapter 11, Academic Press Inc., Boston, MA, 1991, pp. 282–284.
29. K. Nariai, "Geometrical Aberrations of a Generalized Wolter Type I Telescope," *Appl. Opt.* **26**:4428–4432 (1987).
30. K. Nariai, "Geometrical Aberrations of a Generalized Wolter Type I Telescope 2: Analytical Study," *Appl. Opt.* **27**:345–350 (1988).
31. T. T. Saha, "Transverse Ray Aberrations for Paraboloid-Hyperboloid Telescopes," *Appl. Opt.* **24**:1856–1863 (1985).
32. T. T. Saha, "Transverse Ray Aberrations of Wolter Type I Telescopes," *Proc. SPIE* **640**:10–19 (1986).
33. T. T. Saha, "General Surface Equations for Glancing Incidence Telescopes," *Appl. Opt.* **26**:658–663 (1987).
34. T. T. Saha, "Aberrations for Grazing Incidence Telescopes," *Appl. Opt.* **27**:1492–1498 (1988).
35. P. L. Bornman, D. Speich, J. Hirman, V. Pizzo, R. Grubb, Balch, and G. Heckman, "GOES Solar X-Ray Imager: Overview and Operational Goals," in *GOES-8 and Beyond*, E. R. Washwell, ed., *Proc. SPIE* **2812**:309–319 (1996).
36. H. H. Hopkins, *Wave Theory of Aberrations*, Clarendon Press, Oxford, 1950.
37. V. N. Majahan, *Optical Imaging and Aberrations*, SPIE Optical Engineering Press, Bellingham, WA, 1998.
38. L. P. Van Speybroeck, "Einstein Observatory (HEAO B) Mirror Design and Performance," in *Space Optics-Imaging X-Ray Optics Workshop*, M. Weisskopf, ed., *Proc. SPIE* **184**:2–11 (1979).
39. B. Aschenbach, "Design, Construction, and Performance of the ROSAT High-Resolution Mirror Assembly," *Appl. Opt.* **27**:1404–1413 (1988).
40. L. P. Van Speybroeck, "Grazing Incidence Optics for the U.S. High-resolution X-Ray Astronomy Program," *Opt. Eng.* **27**:1398–1403 (1988).
41. P. L. Thompson and J. E. Harvey, "Development of an Image Quality Criterion for Wide-Field Applications of Grazing Incidence X-Ray Telescopes," *Proc. SPIE* **3766-13**:162–172 (1999).
42. J. E. Harvey, A. Krywonos, P. L. Thompson, and T. T. Saha, "Grazing Incidence Hyperboloid-Hyperboloid Designs for Wide-Field X-Ray Imaging Applications," *Appl. Opt.* **40**:136–144 (Jan. 2001).

43. J. E. Harvey, "Diffraction Effects in Grazing Incidence X-Ray Telescopes," *J. X-Ray Sci. Tech.* **3**:68–76 (1991).
44. P. L. Thompson and J. E. Harvey, "A Systems Engineering Analysis of Aplanatic Wolter Type I X-Ray Telescopes," *Opt. Eng.* **39**:1677–1691 (Jun. 2000).
45. J. E. Harvey and A. Krywonos, "A Systems Engineering Analysis of Image Quality," *Proc. SPIE* **4093B-50**:379–388 (2000).
46. S. K. Park, R. Schowengerdt, and M. Kaczynski, "Modulation-Transfer-Function Analysis for Sampled Image Systems," *Appl. Opt.* **2**:2572–2582 (Aug. 1, 1984).
47. G. D. Boreman, *Modulation Transfer Function in Optical and Electro-Optical Systems*, SPIE Press, *Tutorial Texts in Optical Engineering* **TT52**:41 (2001).
48. J. E. Harvey, M. Atanassova, and A. Krywonos, "Including Detector Effects in the Design of Grazing Incidence X-Ray Telescopes," presented at *SPIE's International Symposium on Astronomical Telescopes and Instrumentation*, Glasgow, Scotland, June 2004; published in *Proc. SPIE* **5497-76**:90–99 (Jun. 2004).
49. M. I. Atanassova, "Optimizing the Performance of As-Manufactured Grazing Incidence X-Ray Telescopes Using Mosaic Detector Arrays," Ph.D. Dissertation, University of Central Florida, Orlando, FL, 2005.
50. J. E. Harvey, M. Atanassova, and A. Krywonos, "Systems Engineering Analysis of Five 'As-Manufactured' SXI Telescopes," *Proc. SPIE* **5867-15**:114–124 (Aug. 2005).
51. J. E. Harvey, A. Krywonos, M. Atanassova, and P. L. Thompson, "The Solar X-Ray Imager (SXI) on GOES-13: Design, Analysis, and On-Orbit Performance," presented at *SPIE's International Symposium on Optics and Photonics*, San Diego, CA, August 2007; published in *Proc. SPIE* **6689-11**:6890I–66890I-9 (Aug. 2007).

ABERRATIONS FOR GRAZING INCIDENCE OPTICS

Timo T. Saha

*NASA/Goddard Space Flight Center
Greenbelt, Maryland*

45.1 GRAZING INCIDENCE TELESCOPES

Large number of grazing incidence telescope configurations have been designed and studied (see Chaps. 33 and 47 in this volume). Wolter¹ telescopes are commonly used in astronomical applications. Wolter telescopes consist of a paraboloidal primary mirror and a hyperboloidal or an ellipsoidal secondary mirror. There are eight possible combinations of Wolter telescopes.² Out of these possible designs only type 1 and type 2 telescopes are widely used. Type 1 telescope is typically used for x-ray applications and type 2 telescopes are used for EUV applications.

Wolter-Schwarzschild (WS) telescopes³ offer improved image quality over a small field of view. The WS designs are stigmatic and free of third-order coma and, therefore, the point spread function (PSF) is significantly better over a small field of view. Typically the image is more symmetric about its centroid. The eight designs of WS telescopes have not been widely used because the surface equations are complex parametric equations complicating the analysis, and typically the resolution requirements are too low to take full advantage of the WS designs.

There are several other design options. Most notable are wide-field x-ray telescope designs. Polynomial designs were originally suggested by Burrows⁴ and hyperboloid-hyperboloid designs for solar physics applications were designed by Harvey.⁵

No general aberration theory exists for grazing incidence telescopes that would cover all the design options. Several authors have studied the aberrations of grazing incidence telescopes.⁶⁻⁹ A comprehensive theory of Wolter type 1 and 2 telescopes has been developed.^{10,11} Later this theory was expanded to include all possible combinations of grazing incidence and also normal incidence paraboloid-hyperboloid and paraboloid-ellipsoid telescopes.¹² In this chapter the aberration theory of Wolter-type telescopes is briefly reviewed.

45.2 SURFACE EQUATIONS

The surface equations of the grazing incidence telescopes can be combined to a general form covering large number of design options. General surface equations have been developed for grazing incidence telescopes.² Unfortunately, these equations are too complex to be a basis for the aberration

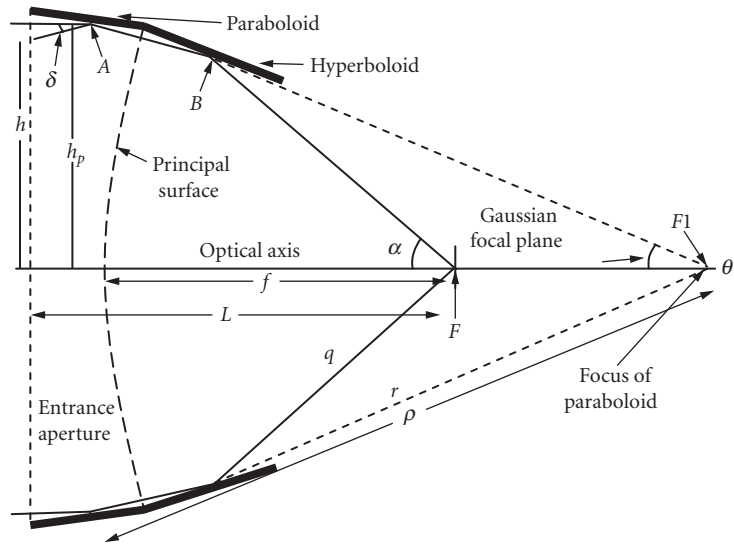


FIGURE 1 Cross-section of Wolter type 1 telescope showing the optical components, ray paths, and the parameters.

theory. The equations of the Wolter telescopes presented in the cylindrical coordinate system are relatively simple. Assuming an incoming ray hits the telescope entrance aperture at radial and azimuthal location (h, β) , it strikes the primary mirror at a location A , the secondary mirror at a location B , and the focal plane at a location F , as shown in Figs. 1 and 2. The extension of the ray which hits the secondary at the location B would intersect the optical axis at $F1$, at the common focus of the paraboloid and hyperboloid.

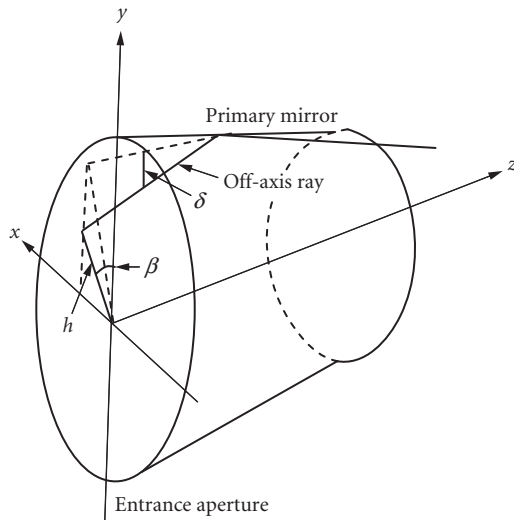


FIGURE 2 Primary mirror of the grazing incidence telescope showing entrance aperture and incoming off-axis ray.

Assuming the primary and the secondary mirror are surfaces of revolution and the primary mirror focus coincides with the secondary mirror focus, then the surface equation of the primary mirror can be written as

$$1/\rho = (\cos \theta - 1)/R_1 \quad (1)$$

where ρ is the distance from the paraboloid-hyperboloid focus F1 to a point on the primary mirror, θ is an angle this ray makes with the optical axis, and R_1 is the vertex radius of curvature of the primary mirror.

The secondary hyperboloid can be expressed either as a function of angle α or θ

$$1/q = (\varepsilon \cos \alpha - 1)/R_2 \quad (2)$$

$$1/r = (1 - \varepsilon \cos \theta)/R_2 \quad (3)$$

where q is the distance from the focus F to the point B on the secondary mirror, α is an angle q makes with the optical axis, R_2 is the vertex radius of curvature of the secondary mirror, ε is the eccentricity of the secondary, and r is the distance from the point B on the secondary to the focus F1 of the primary.

The principal surface^{13,14} of the Wolter telescopes is defined by the intersection points of the extensions of the incoming rays and the extensions of the rays reflected on the secondary, q . The principal surface goes through the intersection points of the primary mirror and secondary mirror. The principal surface of the Wolter telescopes is always a paraboloid.¹³ It is useful to define the focal length f of the telescope to be the distance from the vertex of the principal surface to the telescope focus F . Quite often the focal length is defined as the axial distance from telescope focus to the primary-secondary surface intersection plane.

A useful relation for the paraboloid-hyperboloid or paraboloid-ellipsoid telescopes is¹⁵

$$h_p = 2f \tan(\alpha/2) \quad (4)$$

This equation ties the telescope object side to the image side. Equation (4) shows that the Wolter telescopes do not satisfy the Abbe's sine condition

$$h_p = f \sin \alpha \quad (5)$$

If the angle α is small, as is the case in x-ray optical systems, then the trigonometric terms could be expanded in α . The difference in Eqs. (4) and (5) would be in the third-order term indicating that the third-order coma should be small in the Wolter telescopes and the telescopes nearly satisfy the Abbe's sine condition.

45.3 TRANSVERSE RAY ABERRATION EXPANSIONS

Transverse ray aberration (TRA) expansions are expressed as functions of entrance aperture coordinates (h, β) shown in Fig. 2. The expansions presented here are based on the format introduced by Cox.¹⁶ The derivation is simple but rather lengthy. In the derivation an off-axis ray making an angle δ is mathematically traced from the primary mirror to the secondary mirror and to the focal plane. The surface intersection points on the primary and on the secondary are solved with respect to an on-axis ray and expanded in δ and h . The image plane intersection point (H_x, H_y) of this ray are the final image coordinates. They are then expressed as functions

of entrance aperture coordinates (h, β) and image height $H_0 (=f \tan(\delta))$. The TRA expansions of Wolter telescopes are

$$H_x = H_0(A_2 h^2 \sin(2\beta) + B_2 h^4 \sin(2\beta)) + H_0^2 \sin\beta [A_3 h + B_3 h^3 + C_1 h^5 + \cos^2(\beta) B_4 h^3] + H_0^3 \sin(2\beta) (B_5 h^2 + C_2 h^4)/2 \quad (6)$$

$$H_y = H_0[1 + A_2 h^2 (2 + \cos(2\beta)) + B_2 h^4 (1.5 + \cos(2\beta) - \lambda_1 h^4)] + H_0^2 \cos\beta \{A_4 h + h^3 [B_3 + \lambda_2 + B_4 (1 + \cos^2(\beta))]\} + H_0^3 \cos^2(\beta) (B_5 h^2 + C_2 h^4) \quad (7)$$

In the derivation and resulting expansions only significant terms for grazing incidence telescopes are kept. The image height terms (H_0) higher than third order are dropped. The radial height, h , terms higher than fourth order are also dropped. The parameters of the TRA expansions are listed as follows in Eqs. (8) through (19) as functions of the telescope basic parameters f , R_1 , R_2 , ε , and L . The parameter L is the length of the telescope from the entrance aperture to the focal plane.

$$A_2 = 1/(4f^2) \quad (8)$$

$$A_3 = [K/(2f) - 1/R_1 - 1/R_2]/f \quad (9)$$

$$A_4 = [3K/(2f) - 1/R_1 - 1/R_2]/f \quad (10)$$

$$B_2 = 1/(8f^4) \quad (11)$$

$$B_3 = [K/f - 1/R_1 - 1/R_2 - 2f/(R_1 R_2)]/(4f^3) \quad (12)$$

$$B_4 = K/(2f^4) \quad (13)$$

$$B_5 = [-K + 1 + 2(R_1/R_2) + (R_1/(2f))]/(f^2 R_1 R_2) \quad (14)$$

$$\lambda_1 = 1/(16f^4) \quad (15)$$

$$\lambda_2 = [-K/(4f) + 1/R_1 + 1/R_2]/f^3 \quad (16)$$

$$C_1 = -1/(4f^4 R_1 R_2) \quad (17)$$

$$C_2 = (-1/R_1 + 1/R_2)/(2f^2 R_1 R_2) \quad (18)$$

$$K = [L - 2\varepsilon R_2/(\varepsilon^2 - 1)]/f + R_1/R_2 \quad (19)$$

The Wolter telescopes are stigmatic on-axis and, therefore, the designs are free of all orders of spherical aberration terms. The first term in the expansion is the third-order coma. The coma coefficient is proportional to the inverse square of the focal length and the coma term is inversely proportional to the square of the telescopes' f -number. It does not depend on the location of the entrance aperture or the other parameters.

The third-order aberration terms A_3 and A_4 are proportional to second order of the image height (H_0) and first-order aperture height h . Both terms depend on the location of the entrance aperture. Astigmatism and field curvature can be derived from these terms.¹⁰

All fifth-order terms are represented in the expansions relative to Cox's work.¹⁶ As the third-order spherical aberration, the fifth-order spherical aberration term is zero since the Wolter telescopes are stigmatic. Coefficients B_2 and λ_1 represent fifth-order circular coma. The term including the B_3 , B_4 , and λ_2 is the so-called *astralate aberration*.¹⁶ If coefficient $B_4 = 0$, the term is called *fifth-order oblique spherical aberration*. The terms represented by B_5 coefficient is the fifth-order elliptical coma aberration. Approximations suitable for grazing incidence telescopes were made in the derivation of B_5 coefficient.

Two seventh-order terms C_1 and C_2 proportional to $H_0^2 h^5$ and $H_0^3 h^4$ are approximations. Exact solutions are very complex formulas of the basic parameters.

Typically the seventh-order terms and the fifth-order terms become more important when the grazing angles of the surfaces decrease. Expanding the TRA equations as a function of radial height h is not the best choice in case of grazing incidence systems. For example, expanding the TRA equations as a function of Δh ($h - h_{\text{int}}$), where h_{int} is the radial height at the primary secondary intersection plane, could lead to fewer terms and aberration coefficients more meaningful for the grazing incidence telescopes.

The RMS image size can be represented as a function of the aberrations coefficients.¹⁷ The resulting equation is rather complex function of the aberration coefficients and the field angle.

45.4 CURVATURE OF THE BEST FOCAL SURFACE

All combinations of Wolter telescopes suffer from large curvature of the best focal surface. This limits the field of view of the telescopes. The shape of the best focal surface is parabolic. The radius of curvature R_d of this surface can be estimated from the TRA equation¹¹

$$1/R_d = f \left[A_3 + A_4 + (R_{\text{max}}^2 + R_{\text{min}}^2)(B_3 + B_4 + \lambda_2/2) + \frac{C_1}{3} \frac{R_{\text{max}}^6 - R_{\text{min}}^6}{R_{\text{max}}^2 - R_{\text{min}}^2} \right] \quad (20)$$

where R_{max} and R_{min} are the maximum and minimum radial heights of the entrance aperture, respectively. If only the third-order terms A_3 and A_4 are included, the equation represents third-order field curvature.¹⁰ In case of grazing incidence telescopes the fifth-order term ($B_3 + B_4 + \lambda_2/2$) is comparable to the third-order term. If the grazing angles are small, even the seventh-order term C_1 cannot be omitted.

Alternative equation for the shape of the best focal surface is given by Shealy.¹⁷ In Shealy's paper the RMS image radius is formally calculated from the TRA equations [Eqs. (6) and (7)].

45.5 ABERRATION BALANCING

In case of Wolter type 1 telescopes the aberration equations suggest that for the optimum design the primary and the secondary should be as close to each other as possible. Separating the primary and the secondary, increases the radial heights on the primary, and therefore the image size.

The best focal surface of the Wolter type 1 telescopes always curves toward the telescope. The largest term in the aberration coefficients is the sum of the inverse of the radii of curvatures ($1/R_1 + 1/R_2$). For Wolter type 1 telescopes, the radii of curvatures are both negative and these quantities in the terms A_3 , A_4 , and B_3 cannot cancel each other.

In case of Wolter type 2 telescopes, R_1 is negative and R_2 is positive. The radius of curvature can be optimized. It turns out that for all the practical designs the R_2 is always smaller than R_1 and the radius of curvature of the best focal surface is negative and curving toward the telescope. The aberrations are optimized when the primary and secondary are as close to the primary-secondary surface intersection point as possible.

The field of curvature could be improved by moving the entrance aperture away from the telescope. The K parameter would get bigger since the length L of the telescope would increase. Having the entrance aperture far in front of the telescope may not be practical design. For example, the vignetting would increase rapidly as a function of the off-axis angle.

The aberration equations [Eqs. (6) and (7)] presented in this paper are derived in terms of conventional parameters. The equations are shown as functions of the entrance aperture coordinates using the formulation introduced by Cox.¹⁶ The TRA polynomials and the OPD-polynomial have also been derived as functions of the exit pupil coordinates¹² using the traditional formulation shown, for example, in *Handbook of Optics*.¹⁸ The derivation includes all the terms shown in this paper (as a function of entrance aperture coordinates). The expansions are valid for all the combinations of Wolter telescopes and also for all the combinations of normal incidence paraboloid-hyperboloid or paraboloid-ellipsoid telescopes.

45.6 ON-AXIS ABERRATIONS

Rigid-Body Motions

Rigid-body motions and low spatial frequency errors of the primary and secondary are the most important on-axis image aberrations. These errors typically degrade the on-axis resolution of the grazing incidence telescopes and limit the encircled energy performance of the telescopes. Glenn¹⁹ introduced an orthonormal set of Legendre-Fourier (L-F) polynomials for cylindrical mirrors which are used to describe the low-order errors of the primary and secondary mirrors. The L-F polynomials have been implemented in the optical surface analysis code (OSAC) ray trace code.²⁰

The TRA aberration expansions have been derived for the rigid body motions and low order L-F polynomials.²¹ The rigid-body motions of the primary and secondary mirrors are despace, decenter, tilt, and defocus errors. The TRA equations can be derived following the similar principle used in the derivation of TRA expansions of off-axis aberrations shown in Sec. 45.5.

The defocus term (Δz) can be expressed as^{11,21}

$$H_x = -\frac{\Delta z}{f} h \sin(\beta) \quad (21)$$

$$H_y = -\frac{\Delta z}{f} h \cos(\beta) \quad (22)$$

In the equations only the first-order term in radial height h is kept and the higher-order terms are omitted. The defocus terms are proportional to the radial height h . Therefore, this term can be used to optimize the off-axis aberration terms that are also proportional to $h \sin(\beta)$ - $h \cos(\beta)$ pair. This principle was used to find the best focal surface and the radius of curvature of the best focal surface.

Despace surface errors of the primary ($n = 1$) and secondary ($n = 2$) can be expressed as surface radial height errors Δh_n and surface axial errors Δz_n as

$$\Delta h_n = 0 \quad (23)$$

$$\Delta z_n = \text{constant} \quad (24)$$

The despace image terms of the primary mirror can be approximated by

$$H_x = -\Delta z_1 \frac{4fR_1^2}{h^3} \sin(\beta) \quad (25)$$

$$H_y = -\Delta z_1 \frac{4fR_1^2}{h^3} \cos(\beta) \quad (26)$$

The despace image terms of the secondary mirror are

$$H_x = \Delta z_2 \left(\frac{4fR_2^2}{h^3} + \frac{h}{f} \right) \sin(\beta) \quad (27)$$

$$H_y = \Delta z_2 \left(\frac{4fR_2^2}{h^3} + \frac{h}{f} \right) \cos(\beta) \quad (28)$$

The leading term in the expansions is now proportional to inverse of h^3 . The off-axis aberrations shown above do not have terms proportional to inverse of h . This is because the primary mirror and secondary mirror are stigmatic and confocal.

Decentering the mirror leads to radial height error of the primary ($n = 1$) and secondary ($n = 2$) that can be written in terms of radial error as

$$\Delta h_n = e_{01n} \cos(\beta) + f_{01n} \sin(\beta) \quad (29)$$

where e_{01n} and f_{01n} are the amount of decenter error. The approximate TRA aberrations of the decentered primary mirror for the cosine component are

$$H_x = e_{011} \frac{2fR_1}{h^2} \sin(2\beta) \quad (30)$$

$$H_y = e_{011} \frac{2fR_1}{h^2} \cos(2\beta) \quad (31)$$

The equations of decentered secondary are similar

$$H_x = -e_{012} \frac{2fR_1}{h^2} \sin(2\beta) \quad (32)$$

$$H_y = e_{012} \left[1 - \frac{2fR_1}{h^2} \cos(2\beta) \right] \quad (33)$$

The tilt error of the mirrors can also be expressed in terms of radial height error and axial translation of the primary ($n = 1$) and secondary ($n = 2$). The radial and axial errors are

$$\Delta h_n = z_n (E_{11n} \cos(\beta) + F_{11n} \sin(\beta)) \quad (34)$$

$$\Delta z_n = -h_n (E_{11n} \cos(\beta) + F_{11n} \sin(\beta)) \quad (35)$$

where E_{11n} and F_{11n} are the amount of tilt error in radians when the mirror is rotated about x axis and y axis, respectively. The approximate expansions of the primary mirror are

$$H_x = -E_{111} f \left(\frac{h_{10}}{h} \right)^2 \sin(2\beta) \quad (36)$$

$$H_y = -E_{111} f \left[1 + \left(\frac{h_{10}}{h} \right)^2 \cos(2\beta) \right] \quad (37)$$

where h_{10} is the radial height of the primary mirror at the center of the mirror. The equations of the secondary mirror are

$$H_x = E_{112} \frac{fh_{10}h_{20}}{h^2} \sin(2\beta) \quad (38)$$

$$H_y = E_{112} f \frac{h_{10}}{h_{20}} \left[1 + \left(\frac{h_{10}}{h} \right)^2 \cos(2\beta) \right] \quad (39)$$

where h_{20} is the radial height of the secondary mirror at the center of the mirror.

The $\sin(2\beta)$ - $\cos(2\beta)$ relationship of the primary and secondary components shown in the equations for the decenter [Eqs.(30) to (33)] and tilt [Eqs. (36) to (39)] errors is typical of coma. Note that the components are now inversely proportional to h^2 . The off-axis coma term is directly proportional to h^2 .

Axial and Circumferential Slope Errors and TRA Equations

Assuming the primary mirror figure error Δh_1 and the secondary mirror figure error Δh_2 are known as functions of the surface axial coordinate and circumferential coordinate. Then, if the grazing angles of the mirrors are small, the on-axis aberrations can be evaluated easily from the surface slope errors. The TRA equations of the primary mirror of the axial surface errors are

$$H_x = -2f \frac{\partial \Delta h_1}{\partial z_1} \sin(\beta) \quad (40)$$

$$H_y = -2f \frac{\partial \Delta h_1}{\partial z_1} \cos(\beta) \quad (41)$$

where $\partial\Delta h_1/\partial z_1$ is its axial slope error. The secondary mirror TRA equations for the axial slope errors are

$$H_x = 2q \frac{\partial\Delta h_2}{\partial z_2} \sin(\beta) \quad (42)$$

$$H_y = 2q \frac{\partial\Delta h_2}{\partial z_2} \cos(\beta) \quad (43)$$

where q is the distance from the secondary to the telescope focus and $\partial\Delta h_2/\partial z_2$ is the axial slope error of the secondary. The q variable can be approximated by q_0 that is the distance from the center point of the surface to the telescope focus.

The approximated TRA equations for the primary mirror circumferential slope errors are

$$H_x = f \sin(2I_1) \frac{1}{h} \frac{\partial\Delta h_1}{\partial\beta} \cos(\beta) \quad (44)$$

$$H_y = -f \sin(2I_1) \frac{1}{h} \frac{\partial\Delta h_1}{\partial\beta} \sin(\beta) \quad (45)$$

where I_1 is the grazing angle on the primary and $\partial\Delta h_1/(h\partial\beta)$ is the circumferential slope error of the primary mirror. The approximate TRA equations of the secondary for the circumferential slope error are

$$H_x = -2f \sin(I_2) \frac{h_2}{h} \frac{\partial\Delta h_2}{\partial\beta} \cos(\beta) \quad (46)$$

$$H_y = 2f \sin(I_2) \frac{h_2}{h} \frac{\partial\Delta h_2}{\partial\beta} \sin(\beta) \quad (47)$$

where I_2 is the grazing angle of the surface, h_2 is the radial height of the surface, and $\partial\Delta h_2/h_2\partial\beta$ is the circumferential slope of the surface. The variables I_2 and h_2 can be approximated by the grazing angle and radial height at the midpoint on the surface.

The TRA equations of the circumferential errors include the grazing angle I_1 or I_2 . Typically, the grazing angles are small (0.5 to few degrees). Therefore, the circumferential errors have miniscule effect on the image at the focal plane compared to the axial slope errors and, in many cases, can be ignored.

45.7 REFERENCES

1. H. Wolter, "Mirror Systems with Glancing Incidence as Image-Producing Optics for X-Rays," *Ann. Phys.* **10**:94 (1952).
2. T. T. Saha, "General Surface Equations for Glancing Incidence Telescopes," *Appl. Opt.* **26**:658–663 (1987).
3. H. Wolter, "Generalized Schwarzschild Systems of Mirrors with Glancing Reflection as Optical System for X-Rays," *Ann. Phys.* **10**:286 (1952).
4. C. Burrows, R. Burg, and R. Giacconi, "Optimal Grazing Incidence Optics and Its Applications to Wide-Field X-Ray Imaging," *Astrophys. J.* **392**:760–765 (1992).
5. J. E. Harvey, A. Krywonos, P. L. Thompson, and T. T. Saha, "Grazing Incidence Hyperboloid-Hyperboloid Designs for Wide-Field X-Ray Imaging Applications," *Appl. Opt.* **40**:136–144 (2001).
6. L. VanSpeybroeck and R. Chase, "Design Parameters of Paraboloid-Hyperboloid Telescopes for X-Ray Astronomy," *Appl. Opt.* **11**:440 (1972).
7. H. Wolter, "Estimation of Image Aberrations for X-Ray Telescopes," *Opt. Acta* **18**:425 (1971).
8. W. Werner, "Imaging Properties of Wolter Type 1 X-Ray Telescopes," *Appl. Opt.* **16**:764 (1977).
9. C. Winkler and D. Korsch, "Primary Aberrations for Grazing Incidence," *Appl. Opt.* **16**:2464 (1977).
10. T. T. Saha, "Transverse Ray Aberrations of Paraboloid-Hyperboloid Telescopes," *Appl. Opt.* **24**:1856–1863 (1985).

11. T. T. Saha, "Transverse Ray Aberrations of Wolter Type 1 Telescopes," *Proc. SPIE* **640**:10–19 (1986).
12. T. T. Saha, "Aberrations for Grazing Incidence Telescopes," *Appl. Opt.* **27**:1492–1498 (1988).
13. H. P. Brueggemann, *Conic Mirrors*, Focal Press, London, 1968.
14. J. D. Mangus, "Optical Design of Glancing Incidence XUV Telescopes," *Appl. Opt.* **9**:1019 (1970).
15. T. T. Saha, "A Generalized Sine Condition and Performance Comparison of Wolter Type 2 and Wolter-Schwarzschild Telescopes," *Proc. SPIE* **444**:112–117 (1984).
16. A. Cox, *A System of Optical Design*, Focal Press, London, 1964.
17. D. L. Shealy and T. T. Saha, "Formula for the RMS Blur Circle Radius of Wolter Telescopes Based on Aberration Theory," *Appl. Opt.* **29**:2433–2439 (1990).
18. W. G. Driscoll, ed., *Handbook of Optics*, McGraw-Hill, New York, 1978.
19. P. Glenn, "Set of Orthonormal Surface Error Descriptors for Near-Cylindrical Optics," *Opt. Eng.* **23**:384–390 (1984).
20. R. J. Noll, P. Glenn, and J. F. Osantowski, "Optical Surface Analysis Code (OSAC)," *Proc. SPIE* **362**:78–85 (1983).
21. T. T. Saha, "Image Defects from Surface and Alignment Errors in Grazing Incidence Telescopes," *Opt. Eng.* **29**:1296–1305 (1990).

This page intentionally left blank

X-RAY MIRROR METROLOGY

Peter Z. Takacs

*Brookhaven National Laboratory
Upton, New York*

46.1 GLOSSARY

d	sampling distance on surface
DFT	two-sided discrete Fourier transform
f	spatial frequency
$F[*]$	Fourier transform operator
H	transfer function in frequency domain
$K(m)$	bookkeeping factor
M	surface slope function
PSD	power spectral density function
Rq, Mq	root-mean-square (RMS) roughness, and slope error
$S_1(f)$	one-sided height profile power spectral density function
$S_1'(f)$	one-sided slope profile power spectral density function
$W(n)$	window function in spatial domain
w	sensor pixel width projected onto surface
$Z(x)$	surface height profile
$z(x)$	residual surface height profile after detrending

46.2 INTRODUCTION

The earliest use of mirrors for producing images with x rays dates from the work of Kirkpatrick and Baez in 1948 who used an orthogonal pair of shallow cylindrical mirrors to produce a focal spot.¹ Interest in x-ray mirror fabrication technology grew rapidly in the 1970s and 1980s as a result of several factors: the initiation of a number of space-based x-ray telescope projects, the development of diamond machining of precision optical components, the development of bright laboratory x-ray sources for x-ray microscopy, and the development of dedicated synchrotron radiation source

facilities with many available beam lines. Owing to the nature of grazing incidence x-ray optics, conventional interferometric techniques for surface figure measurement are not easily employed in the testing of these aspheric surfaces. Surface profilometry, either contact or noncontact, has been the method of choice for measuring the figure of Wolter telescopes, x-ray microprobe optics, and synchrotron radiation mirrors. This has necessitated the development of specialized instruments with measurement capabilities far beyond those of the typical coordinate-measuring machine (CMM) found in precision machining shops. Grazing incidence x-ray mirror metrology requires measurement precision and accuracy in the nanometer range, while typical CMM machines are only able to achieve measurements at the micron level.

46.3 SURFACE FINISH METROLOGY

Surface roughness reduction has been a major consideration in the use of grazing incidence optics in x-ray telescope applications^{2,3} and in synchrotron beam line instrumentation.⁴ Optical systems used to reflect x rays at extreme grazing incidence angles are often made from far off-axis segments of cylinders, paraboloids, and toroids. These aspheric surfaces usually have a long tangential radius and a short sagittal radius, which necessitates fabrication processes that depart from conventional optical polishing techniques. Conventional full-contact pitch lap polishing works well for flat and spherical surfaces, but is difficult to achieve for aspheric surfaces. Novel figuring techniques, such as single-point diamond turning and ductile grinding, leave large microroughness levels that need to be reduced by polishing.

Measuring the roughness of aspheric optics at the sub-nanometer level was difficult before the advent of noncontact optical profilers. The collection of papers by Bennett⁵ chronicles the development of various surface roughness-measurement instruments and techniques over the past several decades. Of particular relevance to x-ray optics are the noncontact optical-measurement methods. Early optical profilers were based around interference microscopes with images recorded on film or by eye. Roughness was estimated by observing irregularities in the shape of multiple-beam interference fringes, as in the fringes of equal chromatic order (FECO) technique⁶ or by analyzing differential interference contrast (Nomarski) microscope images.⁷ Contact methods were more quantitative. Stylus profilers, such as the Talystep and Dektak, used a small radius diamond stylus that was dragged across the surface to record the surface topography. Analog data was recorded on a chart recorder. The stylus profilers were difficult to use for x-ray optics because they could only handle small samples and were not suited for measuring the surface of large optics. There was always the risk of damaging the surface of an optic coated with a soft metal-reflecting layer with the diamond stylus. Methods were developed to make replicas of surface roughness with collodion films that could then be analyzed in electron microscopes or by stylus instruments. This hybrid measurement technique was not very repeatable and suffered from accuracy problems. Sommargren developed a noncontact optical scanning profiler with Angstrom-level accuracy that utilized a precision rotary stage to scan a probe beam around a central stationary reference beam.⁸ The first noncontact optical profiling instrument based upon a computer-controlled phase-measuring interferometer (PMI) was developed by Koliopoulos^{9,10} and became available as a commercial instrument in the early 1980s. Similar instruments are now available from several sources.

Digital optical profilers revolutionized surface roughness-measurement technology for x-ray optics. Since they were true noncontact measurements, it was possible to configure the microscope-based PMI instruments to measure the entire surface of full-size mirrors. Manufacturers were able to get rapid feedback about their polishing processes and were able to make significant improvements in producing low-scatter superpolished surfaces for x-ray optics. In parallel with these instrument developments, advances in x-ray scattering theory were made to connect the performance of grazing incidence optics at x-ray wavelengths with roughness measurements made by the various surface-profiling techniques at visible wavelengths. A critical part of connecting the surface-finish measurements to the functional performance at x-ray wavelengths was the understanding of the measurement bandwidth of each profiling technique (frequency footprint) and how it is related to scattering at grazing incidence through the power spectral density function.^{11–17} (See Chap. 44 in this volume.)

46.4 SURFACE FIGURE METROLOGY

Full-Aperture Techniques

Conventional optical interferometry of grazing incidence aspheric optics is difficult to implement. Speer used a Linnik interferometer arrangement to visualize the figure error on toroidal surfaces at EUV wavelengths,¹⁸ but most interferometry techniques at grazing incidence on cylindrical aspheres result in poor spatial sampling of errors along the surface owing to the extreme foreshortening of the surface in the aperture. Grazing angle of incidence testing of long flat and large radius spherical mirrors can be done by using double-pass Fizeau techniques.¹⁹ This geometry allows one to foreshorten the long axis of a mirror to fit within the interferometer aperture that is usually 4 or 6 inches in diameter. Careful system calibration needs to be done with this test geometry to ensure that subtle systematic errors are corrected in order to achieve nanometer accuracy. Full-aperture measurements of x-ray telescope optics at x-ray wavelengths have been made at specialized test facilities, such as the X-Ray Calibration Facility at Marshall Space Flight Center that has a 518-m long source distance, and the PANTER facility at the Max-Planck-Institute near Munich. These are not interferometric tests, rather they measure directly the quality of the image of an x-ray point source.

Height Profilometry

Practical figure metrology methods for aspheric optics are generally based on some type of profilometry. Surface-profiling instruments for grazing incidence metrology can be classified into two broad categories: height-measuring instruments and slope-measuring instruments, based on the fundamental measurement provided by the machine. Height-measuring instruments measure the distance between the test surface and a reference surface; slope-measuring instruments measure either the differential phase shift between two closely spaced probe beams or the angle of a reflected laser beam directly, and are usually based upon some variant of the optical lever principle.^{20–22} A number of surface profilers with nanometer precision were originally developed for microroughness investigations. Their scan ranges were limited to a few millimeters. Bennett et al.²³ describe a number of commercially available mechanical stylus-based profiling instruments that have been modified to extend the measurement length to the 100- to 200-mm range. Carl Zeiss has developed a 3D coordinate-measuring machine, the M400, which utilizes a laser-based distance interferometer on all three axes with 10-nm positioning resolution.²⁴ This machine is capable of measuring slope errors with a repeatability of 0.1 arc sec (0.5 μ rad) over the $400 \times 600 \times 200$ mm measurement volume. Other scanning stylus profilers were developed at the National Physical Laboratory²⁵ and at Cranfield Institute.²⁶ A noncontact fringe-scanning technique was developed at Perkin-Elmer to measure the distance between an aspheric test surface and a toroidal test plate. The fringe scanner measured the position of Fizeau fringes formed in the air gap between the two surfaces and converted the air-gap profile into a surface height profile.^{27,28} This instrument was used to measure the axial profile of the HEAO-2 and AXAF Wolter telescope mirrors.

Slope Profilometry

Slope-profiling instruments generally operate on the principle of the optical lever, where an angular deviation of a probe beam is magnified and recorded with a high-gain amplifier. Jones chronicles the development of sensitive optical lever instruments capable of measuring nanoradian angular deflections.²² An optical lever device was developed by George Random to measure slope errors on x-ray optics several centimeters in length.^{29,30} The Random Devices Slope Scanner used a clever air-bearing slide mechanism to move the probe beam along a circular arc that matched the curvature of the part under test.³¹ This motion always kept the surface near the focus of the laser beam spot. The deflection of the reflected beam was measured with a resolution of a few tenths of an arc second by the differential signal from a pair of photodiodes. As DeCew notes, the limitation in the measurement accuracy was not in the optical system, but was in the tolerances of the mechanical translation system.

Makosh and coworkers developed a number of surface-profiling instruments at IBM in Germany based on the principle of differential heterodyne interferometry.^{32–34} Two probe beams generated by a Wollaston prism were focussed onto a test surface. The spot separation could be varied from 2 μm to 1.5 μm by changing the Wollaston prism or the microscope objective. The phase shift between the two spots was a measure of the local surface slope, and the slope was integrated to provide surface height roughness information. They point out that the differential height measurement is insensitive to mechanical vibrations and air turbulence and is suitable for long-distance measurements.

A slope-measuring system was developed by Ennos and Virdee at NPL that was based on the principle of laser autocollimation.^{35,36} An unfocused laser beam was reflected off a cylindrical asphere and an astigmatic line image was focused onto a position-sensitive quad cell detector so that the detector sensed the slope errors in the axial direction of the scan. An innovation in the Ennos and Virdee system was the incorporation of an optical reference beam to monitor changes in the pointing direction of the laser beam during the scan. The measurement precision was on the order of 0.5 μrad , corresponding to a displacement of the line focus by 0.1 μm on the detector, but beam pointing drifts of 1.5 μrad over the 10-min scan period needed to be corrected from the measured profiles. The instrument could reproducibly measure height profiles to within ± 5 nm over 16-mm travel lengths.

Bristow and Arackellian reported on the development of a slope-measuring profiler based upon a modification of a surface microroughness-profiling instrument.³⁷ The original was based on the principle of Nomarski differential interference contrast but was modified to act as a differential polarization interferometer.³⁸ The MP2000 had an intrinsic lateral resolution on the order of 1 μm , owing to the focal properties of the microscope objective used to focus the beam onto the surface. An astigmatic autofocus control system was added to the instrument to extend the depth of focus and allow the profiler to scan over longer distances on curved parts.³⁹ The basic instrument had a maximum scan range of 100 mm.

Glenn described a novel surface-profiling instrument that, unlike the height- and slope-measuring instruments, measured surface curvature directly on aspheric parts.^{40,41} The Bauer Model 100 used a laser pencil beam that was passed through a calcite prism to generate two spatially separated and orthogonally polarized beams. Both beams were directed onto the test surface by a steerable mirror so that normal incidence could be maintained over surface slope changes of up to $\pm 30^\circ$. The reflected beams were directed to separate quad-cell detectors, one of which could be translated to accommodate the average surface curvature. The differential measurement between each separate beam position was a direct measurement of the variation of the surface curvature away from the average. Because the instrument measures the intrinsic curvature of the surface, it is immune to vibration or tilt errors between the optical head and the part during the scan period. This instrument is particularly well-suited for azimuthal circularity measurements on x-ray telescope mirrors and mandrels.

Weingärtner, Schulz, Estler, and coworkers at the PTB in Braunschweig, Germany, have developed a number of precision figure-measuring techniques based on multiple-beam slope profilometry with large lateral shear distances.^{42–47} Their measurement techniques and analysis algorithms correct for errors in the translation stages and rigid-body motions of the test piece during the measurement. Schulz has extended this work to combine a scanning micointerferometer that measures distance at 16 discrete positions with an autocollimator to measure pitch error during the scan.^{48,49} A sketch of an autocollimator is shown in Fig. 1. The interferometer is scanned over the surface with a 0.2-mm step size, determined by the interferometer sensor resolution. The resultant height measurements are stitched together to produce a high-resolution surface profile with sub-nanometer uncertainty.

The long trace profiler (LTP) is optimized for measuring the surface figure of synchrotron beam line optics. Its design is based upon the principle of the pencil beam interferometer, originally developed by von Bieren in 1982.^{50,51} Qian designed an improved version of the beam-splitting system for the original LTP I at BNL that produces zero-optical path difference probe beams with a spacing that could be adjusted from complete overlap to several millimeters apart.^{52–54} The optical head design is quite simple, with no internal moving parts during the measurement to ensure long-term stability and maintain high accuracy. An advantage of the pencil beam interferometer over microscope-based systems is the effectively infinite depth of field of the probe beam, which relaxes the tolerances on test piece alignment and allows for rapid setup and testing. An LTP slope profile of a single-crystal silicon cylinder mirror is shown in Fig. 2. This mirror exhibits several significant defects and was returned to the manufacturer for rework.

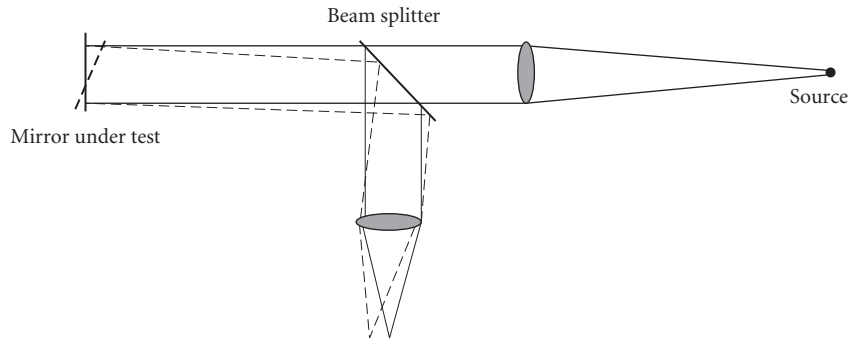


FIGURE 1 Sketch of a visible light autocollimator.

The LTP can be operated in two basic configurations. The standard configuration places the optical head on a linear translation stage and moves the entire optical head over the surface that is being measured. The alternative configuration places the stationary optical head off to the side with the test beam directed horizontally to a penta prism that sends the probe beam down to the test surface. The penta prism is scanned over the surface; the optical head remains fixed. This is the original measurement method proposed by von Bieren.⁵⁰ The nano-optic-measuring (NOM) machine at BESSY II also uses a scanning penta prism with a fixed autocollimator providing the main probe beam.⁵⁵ Each configuration has advantages and disadvantages. The moving optical head requires a high-quality translation stage to minimize pitch angle errors during the scan. A reference beam must be used to correct for residual pitch errors during the scan caused by the sag of the translation stage and irregularities in the motion. The moving penta prism, however, always redirects the incident beam by exactly 90° , independent of

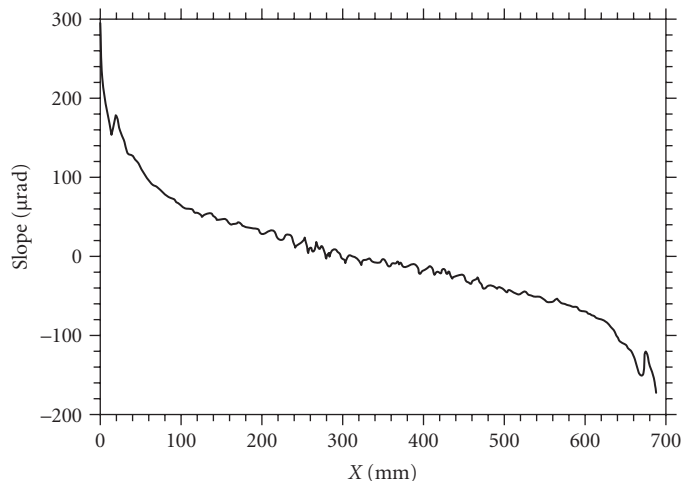


FIGURE 2 Slope profile measurement on a 700-mm long silicon cylinder mirror made with an LTP. Sampling step size is 2 mm. Mean has been subtracted from the data (detrend 0). Profile shows that the surface has an overall convex curvature (tilted profile down to the right) with significant edge roll-off (change in slope at each end). Also, a polishing defect with a 20-mm period is evident in the center of the surface. The slope profile emphasizes high-frequency surface defects. (See also color insert.)

the errors in the translation stage. Use of a reference beam is optional in this case.^{56–62} Use of the penta prism also greatly relaxes the tolerances on the translation mechanism. The penta prism LTP can be configured to provide measurements in difficult environments, such as inside complete x-ray telescope cylinders^{60,63–66} and in situ at SR beam lines during actual operations.^{61,67}

46.5 PRACTICAL PROFILE ANALYSIS CONSIDERATIONS

Nomenclature

Most profiling instruments today produce a map of surface height or slope topography over a 2D area. Early profiling instruments, optical and contact, were restricted to single 1D line profiles. Some authors refer to surface profilometry over a 2D area as *3D profilometry* and over a line as *2D profilometry*, which introduces some ambiguity into the nomenclature. For purposes of clarity, 1D will refer to a linear scan over one line and 2D will refer to a scan over an area.

Detrending

Most x-ray mirrors are made today with extremely smooth surfaces by processes that produce isotropic roughness. In this case it is sufficient to consider surface statistics in one dimension to be descriptive of the entire surface. Most x-ray mirrors have overall figures that are either plano, spherical, or cylindrical, or are a far off-axis segment of a conic section. The current trend is to produce aspheric surfaces by bending substrates that are prefigured with a long-radius sphere or cylinder. In order to assess the underlying residual roughness, it is usually necessary to subtract the best-fit conic section from the measured data before computing statistical properties. The magnitude of the overall figure is usually many orders of magnitude greater than the residual surface roughness and needs to be removed by mathematical manipulation of the measured profile. The residual profile z can be viewed as the difference between the measured profile, Z_{raw} , and the deterministic figure:

$$z(x_n) = Z_{\text{raw}}(x_n) - [A + Bx_n + Cx_n^2 + Dx_n^3 + \dots] \quad (1)$$

where the profile is sampled at discrete points, x_n , separated by equal intervals, d , and the figure is described by some polynomial function of position. The process of removing the deterministic part of the measured profile is called *detrending* and we denote removing the mean as a “detrend0” or “D0” process, a first order detrend as a “detrend1” or “D1” process, and so forth. Most often, the detrending function is determined by a least-squares fit to the measured data. Any convenient polynomial function can be used to describe this function. In certain cases it may be more useful to do a straight-line detrend between the endpoints of the profile. The simple polynomial shown in Eq. (1) has terms that can be identified with various rigid body alignments: A = piston (DC) offset, B = tilt. The second order curvature term C is an intrinsic surface property, independent of body orientation. The third-order term D is related to the ellipticity of the surface, but its value depends on which part of the ellipse is being viewed. For instruments that measure slope M , directly, the slope space profiles can be detrended in a manner analogous to Eq. (1), but the identification of the polynomial coefficients with rigid-body parameters differs from the distance space identifications.

Simple polynomials are not orthogonal over the measured region, so the values of lower-order terms change when higher-order terms are added to the detrending. Other orthogonal polynomial functions, such as sinusoids, Chebyshev, and Legendre may be better suited for linear traces, while Zernike polynomials may be better suited for 2D area detrending, depending upon the application. The simple polynomial, however, has a simple connection to the radius of curvature of the surface when $R \gg L$, (L = the trace length): $R = 1/2C$. For the other polynomials, some combination of two or more coefficients is required to estimate the radius. Note that this discussion has involved fitting a 1D

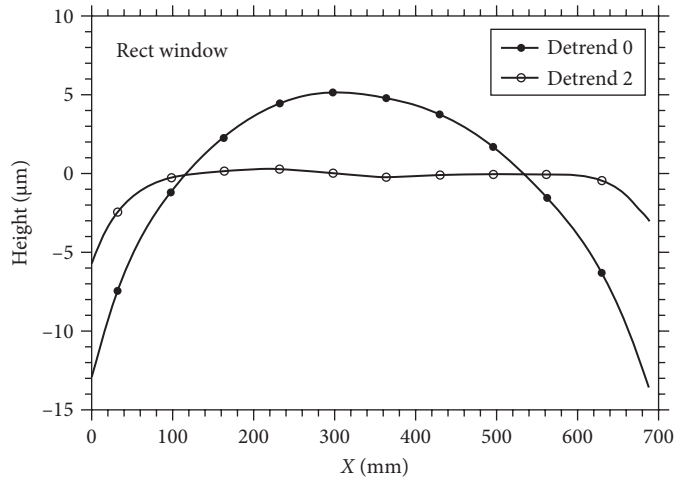


FIGURE 3 Height profile calculated by integrating the slope profile of Fig. 1. Solid circles: mean height subtracted (detrend 0); open circles: second order polynomial subtracted (detrend 2). The radius of curvature extracted from the second order term coefficient is 3.572 km. The residual profile shows that the surface has a “kink” in the center that separates it into two distinct segments with slightly different slopes. This low frequency defect is not evident in the slope profile of Fig. 1. (See also color insert.)

polynomial to a linear profile that is only a function of one coordinate. Equation (1) can be generalized to two variables when fitting a 2D polynomial to the 2D surface. However, when a 2D polynomial is fit to an area, the individual row or column profiles will generally not be fully 1D detrended. In computing 1D statistical quantities from 2D data, one must be aware that the extracted profiles may need further detrending.

A typical detrending example is shown in Fig. 3. The detrend0 height profile has only had the piston (DC) term removed. The radius of curvature can be extracted by further detrending with a second order polynomial function ($D2$). The radius of curvature of the central region of the surface between 100 and 600 mm is derived from the coefficient of the x^2 term, giving $R = 3.572$ km as the best-fit radius. When this fit is subtracted from the $D0$ curve, the edge roll-off at each end of the mirror becomes visible. The region over which the detrending polynomial is applied, and over which the statistical roughness and slope-error parameters are computed, depends upon how the mirror will be used in a synchrotron beamline and on the clear aperture of the illuminated region. Surface errors outside the clear aperture can generally be ignored and should be excluded from the statistics.

Power Spectral Density Function

Various statistical quantities can be computed from profiler data. Standards exist that define the various surface texture parameters that can be derived from surface profile measurement.^{68,69} For high-performance x-ray optics, the most useful descriptor of surface roughness is the power spectral density function (PSD). Church and collaborators have shown that the PSD computed from normal incidence visible light profilometry measurements can be used to predict the performance of grazing incidence optics at x-ray wavelengths in SR beam lines. (See Chap. 8 in Vol. I and Chap. 44 of this volume for a discussion on the connection between PSD and optical scatter.) A detailed description of the definition of the PSD function and related issues involving calculations from sampled data can be found in SEMI standard MF1811-0704⁷⁰ and in the volume on scattered light by Stover.⁷¹

The following paragraphs highlight parameters derived from profile measurements important for x-ray optic characterization.

The “periodogram estimator” for a 1D profile $Z(n)$ is used to define the 1D PSD function^{70–72}

$$S_1(m) = \frac{2d}{N} |\text{DFT}(m)|^2 \cdot K(m) \quad (2)$$

where DFT is the *two-sided discrete Fourier transform* of the N real data points:

$$\text{DFT}(m) = \sum_{n=1}^N e^{i2\pi \frac{(n-1)(m-1)}{N}} \cdot Z(n) \cdot W(n) \quad (3)$$

d = sampling period in one direction, N is the number of sampled points, m is the spatial frequency index, where $f_m = (m-1)/Nd$ is the value of the spatial frequency, $W(n)$ is a window function (see following section), and $K(m)$ is a bookkeeping factor to ensure that Parseval’s theorem is satisfied, that is, that the variance of the distance space profile is equal to the variance of the frequency space spectrum (see the following section).

A number of useful bandwidth-limited statistical parameters and functions can be derived from the PSD function in frequency space. The RMS roughness, Rq , over a given spatial frequency bandwidth is given by

$$Rq^2(\text{low}, \text{hi}) = \frac{1}{Nd} \sum_{m=\text{low}}^{\text{hi}} S_1(m) \quad (4)$$

where low and hi are the indices corresponding to the desired spatial frequency range. When the frequency indices correspond to the full bandwidth, this number is identically equal to the RMS roughness computed in distance space (Parseval’s theorem). Measurements from instruments that have different bandwidths can be compared by restricting the calculation of RMS roughness to the bandwidth that is common to both instruments.

The “Bookkeeping Factor” for the PSD

Most discrete Fourier transform algorithms used in computing libraries today can efficiently calculate the transform for an arbitrary number of real or complex data points. See the section in *Numerical Recipes*⁷³ for a discussion of practical considerations in Fourier transform calculation. The form of the DFT shown in Eq. (3) is known as the “two-sided” DFT, since the m -index runs over all N frequency terms. But since we are starting with N real numbers, the $|\text{DFT}|^2$ will be symmetric about the Nyquist frequency, defined to be $f_{\text{Ny}} = 1/2d$ and will consist of only $N/2$ independent numbers (assuming N is even—see the following discussion). The numbers beyond the Nyquist frequency correspond to the negative frequencies in the input signal and have the same amplitude as the positive frequencies when the input data are real numbers. We can then effectively ignore the numbers beyond the Nyquist frequency and only need consider terms over one half of the N points. When we do this, we restrict m to range over approximately $N/2$ points. However, the missing power in the negative frequencies needs to be added to the positive frequency terms, hence the factor of 2 in the numerator of Eq. (2).

An additional bookkeeping consideration depends on whether N is an odd or even number. In order to ensure that the total power, according to Parseval’s theorem, is satisfied in both distance and frequency space, careful consideration must be given to the terms at the extremes of the frequency interval. In our realization of the Fourier transform, the DC term occurs at frequency index $m = 1$. The fundamental frequency is at $m = 2$ with a value of $1/Nd = 1/L$, where L is the total trace length. The difficulty arises in specifying the index of the Nyquist frequency for a set of N discrete sampled points that can be even or odd. When N is even, the Nyquist frequency term occurs at a single index position, $m = (N/2) + 1$. When N is odd, the Nyquist frequency power is split between adjacent points

$m = 1 + (N - 1)/2$ and $m = 2 + (N - 1)/2$. When the terms in the two-sided spectrum above the Nyquist frequency are discarded and the remaining terms are doubled to generate the one-sided spectrum, the resultant DC term for both N even and odd must be reduced by $1/2$, and the Nyquist term for the $N =$ even case must also be reduced by $1/2$. Hence the bookkeeping factor, $K(m)$:

$$\begin{aligned} \text{for } N \text{ even} \quad K(m) &= \begin{cases} 1/2 & \text{for } m = 1 \text{ or } m = (N/2) + 1 \\ 1 & \text{for all other } m \end{cases} & \text{where } m = 1, 2, \dots, (N/2) + 1 \\ \\ \text{for } N \text{ odd} \quad K(m) &= \begin{cases} 1/2 & \text{for } m = 1 \\ 1 & \text{for all other } m \end{cases} & \text{where } m = 1, 2, \dots, \left(\frac{N-1}{2}\right) + 1 \end{aligned}$$

Windowing

Proper preparation of raw input profile data is necessary to obtain meaningful results from statistical calculations. Detrending is usually required to remove the gross figure terms from the measured data so that the underlying surface roughness can be seen. Even after the gross figure has been removed by detrending, edge discontinuities in the residual profile can introduce spurious power into the spectrum and hide the true nature of the underlying roughness spectrum. Methods have been developed in the signal-processing literature to deal with this “leakage” problem.^{73–75} These methods are collectively known as “prewhitening” techniques. A simple method to preprocess the data is known as “windowing.” A window function, $W(n)$, in distance space is applied to the residual profile before the PSD is computed to smooth the spectrum somewhat and to minimize the spectral leakage from edge discontinuities. The window function used should be normalized to unity so as not to introduce any additional scale factors that would distort the magnitude of the spectrum. Most common window functions tend to enhance the lowest two or three spatial frequencies which are generally related to the deterministic figure components, but the power in these frequencies should already have been minimized by the detrending process. We prefer to use the Blackman window for processing smooth-surface residual profile data. The Blackman window, normalized to unit area, is defined as

$$W(n) = \sqrt{\frac{2}{1523}} [21 - 25\cos(2\pi(n-1)/N) + 4\cos(4\pi(n-1)/N)] \quad (6)$$

The Blackman window applied to the detrended profiles of Fig. 3 are shown in Fig. 4. One can see that this window function forces the edge discontinuities to go to zero. This has the beneficial effect of reducing the discontinuity between the derivatives of the profile at each end point. Discontinuities at the end points, or large spurious spikes in the data, produce large high-frequency ringing effects in the DFT coefficients that distort the underlying surface spectrum. Figure 5 shows the results of computing the PSD for the windowed and unwindowed profiles of Figs. 3 and 4. Application of the window function effectively eliminates the contamination of the high frequency content in each profile, even for significant edge discontinuities.

Instrument Transfer Function Effects

The ideal surface-profiling instrument has a unity transfer function response over an infinite spatial period bandwidth. In other words, the measurement does not distort the intrinsic surface properties. In the real world, however, all measuring instruments have a response function that varies over

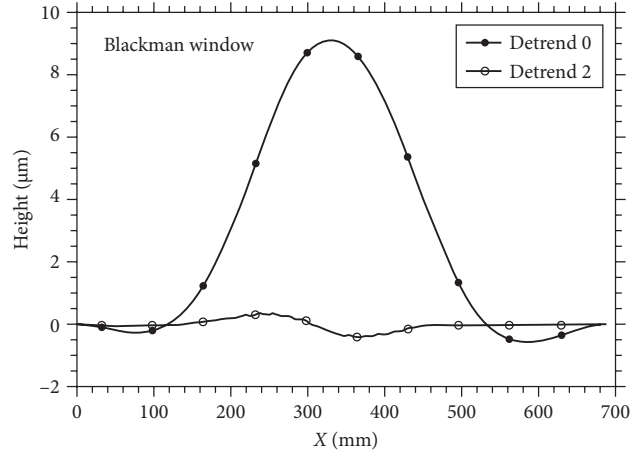


FIGURE 4 The height profiles of Fig. 3 with a Blackman window applied. The edge discontinuities are minimized by this function. Although the shape of the profile is distorted, the average statistical properties of the underlying function are not changed. (See also color insert.)

a limited bandwidth. For optical profiling instruments, the transfer function is limited mainly by the numerical aperture of the objective and the sampling properties of the detector. The transfer function of an unobscured objective for incoherent illumination is given by⁷⁶

$$H_{\text{obj}} = \frac{\pi}{2} (\cos^{-1} \Omega - \Omega \sqrt{1 - \Omega^2}) \quad (7)$$

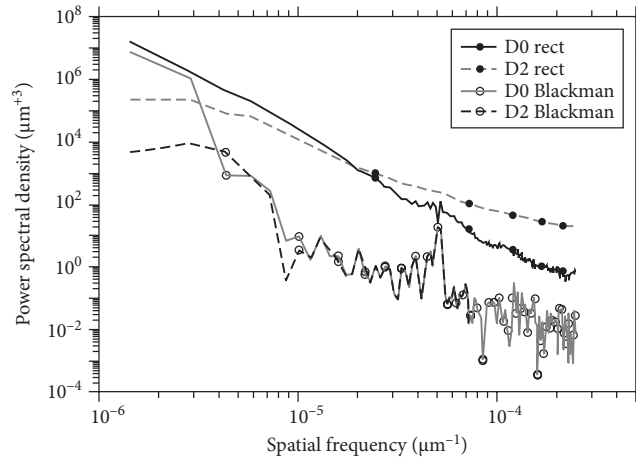


FIGURE 5 PSD curves computed from the four profiles in Figs. 3 and 4. The two upper curves from the unwindowed data show severe contamination effects due to the strong edge discontinuities that introduce spurious power into all frequencies. The lower curves show how the Blackman filter eliminates the discontinuity contamination, allowing the underlying surface spectral characteristics to become visible. (See also color insert.)

where $\Omega = \lambda f / 2NA$ and NA is the numerical aperture of the objective lens. The cutoff frequency where H_{obj} goes to zero is at $f_{\text{cutoff}} = 2NA/\lambda$; the transfer function is zero for all higher frequencies. This function may need to be modified if the objective lens in the profiling instrument contains a central obscuration, as in the case of a Mirau objective. The transfer function of an ideal 1D linear sensor array is given by

$$H_{\text{arr}} = \frac{\sin(\pi w f)}{\pi w f} \quad (8)$$

where w is the pixel width in one dimension. In most cases the pixel width is equal to the sampling distance, $w = d$, and the attenuation at the Nyquist frequency, $f = 1/2d$, is $\sin(\pi/2)/(\pi/2) = 0.63$. Since the transfer function of the array is still above zero beyond the Nyquist frequency, and the optical cutoff frequency is also generally beyond the Nyquist frequency, the measured spectral density will usually contain aliasing from frequencies beyond the Nyquist. This usually results in a flattening of the measured spectrum at the highest spatial frequencies. This effect is obvious when measurements are made with more than one magnification objective with an optical profiler. The combined optical and pixel sampling transfer functions can be used as an inverse filter to restore the high-frequency content of the measured spectrum and give a better estimate of the intrinsic surface power spectral density:

$$\bar{S} = S_{\text{meas}} \cdot H_{\text{obj}}^{-1} H_{\text{arr}}^{-1} \quad (9)$$

There are practical limitations of this technique, such as the need to avoid singularities in the inverse filter that cause the resultant PSD estimate to blow up. In practice, one must impose a practical cutoff frequency before the Nyquist frequency or before the zero in the sampling function to avoid significant distortion of the spectrum by noise and aliasing.⁷⁷ Other complications to this simple restoration filter approach occur when, unknown to the user, the signals from the sensor are preprocessed inside the measuring instrument, such as when adjacent rows of pixels are averaged to smooth out amplifier gain differences in 2D array sensors. In this case, the H_{arr} filter needs to be modified with correction factors.^{78,79}

Slope Measurement Analysis

Surface slope profiles can be derived from surface height measurements by finite difference calculation:

$$M(x_n) = \frac{1}{d} [Z(x_{n+1}) - Z(x_n)] \quad n = 1, 2, \dots, (N-1) \quad (10)$$

Note that there is always one less slope point than there are height points in this calculation. Conversely, height profiles can be generated from slope profiles by numerical integration:

$$Z(x_n) = Z_0 + d \sum_{i=1}^n M(x_i) \quad n = 1, 2, \dots, N \quad (11)$$

Note that this latter calculation involves an arbitrary constant, Z_0 , which corresponds to a rigid body piston orientation of the part. Surface slope and height profiles can also be computed by Fourier differentiation and integration. The formal relationships between the slope and height transforms are

$$\begin{aligned} F[M(x)]_m &= i2\pi f_m F[Z(x)]_m \\ F[Z(x)]_m &= \frac{-i}{2\pi f_m} F[M(x)]_m \end{aligned} \quad (12)$$

where $F[*]$ is the Fourier transform operator. Care must be exercised in implementing these expressions with a DFT to ensure that the frequency terms f_m encompass both the negative and positive frequencies around zero and multiply the corresponding transformed height and slope numbers. Also, the DC term at $f = 0$ must be excluded from the denominator in the slope-to-height transform.

Of particular interest to users of x-ray optics is the slope PSD function, S' , which is related to the height PSD by

$$S'_1(m) = (2\pi f_m)^2 S_1(m) \quad (13)$$

The slope S' spectrum can also be calculated directly from data that is generated by profilers that measure slope by substituting the slope data, $M(n)$, for the $Z(n)$ data in Eq. (3). Bandwidth-limited RMS slope numbers, M_q , can be calculated from the slope S' function in a manner analogous to Eq. (4). Conversely, a measured slope profile PSD curve can be used to predict the height spectrum by the inverse of the above expression:

$$S_1(m) = \frac{1}{(2\pi f_m)^2} S'_1(m) \quad (14)$$

Care must be exercised in this case to exclude the DC term ($f_m = 0$) from the calculation, as it will cause the result to be nonsense.

46.6 REFERENCES

1. P. Kirkpatrick and A. V. Baez, "Formation of Optical Images by X-Rays," *J. Opt. Soc. Am.* **38**:766–774 (1948).
2. B. Aschenbach, "Design, Construction, and Performance of the Rosat High-Resolution X-Ray Mirror Assembly," *Appl. Opt.* **27**(8):1404–1413 (1988).
3. L. Van Speybroeck, *AXAF Mirror Fabrication—Element Fabrication*, http://hea-www.harvard.edu/asc/news_02/subsection3_4_3.html. (1994).
4. M. Howells and P. Z. Takacs, "Use of Diamond Turned Mirrors for Synchrotron Radiation," *Nucl. Instrum. Methods* **195**:251–257 (1982).
5. J. M. Bennett, *Surface Finish and Its Measurement*, Washington, D.C.: Optical Society of America, p. 918, 1992.
6. S. Tolansky, *Multiple-Beam Interferometry of Surfaces and Films*, Oxford: Clarendon Press, p. 187, 1948.
7. J. S. Hartman, R. L. Gordon, and D. L. Lessor, "Quantitative Surface Topography Determination by Nomarski Reflection Microscopy. 2: Microscope Modification, Calibration, and Planar Sample Experiments," *Appl. Opt.* **19**(17):2998–3009 (1980).
8. G. E. Sommargren, "Optical Heterodyne Profilometry," *Appl. Opt.* **20**:610 (1981).
9. C. L. Koliopoulos and J. C. Wyant, "Profilometer for Diamond-Turned Optics Using a Phase-Shifting Interferometer," *J. Opt. Soc. Am.* **70**(12):1591–1591 (1980).
10. B. Bhushan, J. C. Wyant, and C. L. Koliopoulos, "Measurement of Surface Topography of Magnetic Tapes by Mirau Interferometry," *Appl. Opt.* **24**:1489–1497 (1985).
11. E. L. Church, "The Precision Measurement and Characterization of Surface Finish," *Precision Surface Metrology, Proc. SPIE* **429**, J. C. Wyant, ed., pp. 86–95 (1983).
12. E. L. Church and P. Z. Takacs, "Use of an Optical-Profilometer Instrument for the Measurement of the Figure and Finish of Optical-Quality Surfaces," *WEAR* **109**:241–257 (1986).
13. E. L. Church and P. Z. Takacs, "Statistical and Signal Processing Concepts in Surface Metrology," *Optical Manufacturing, Testing, and Aspheric Optics, Proc. SPIE* **645**:107–115 (1986).
14. E. L. Church and P. Z. Takacs, "The Interpretation of Glancing-Incidence Scattering Measurements," *Grazing Incidence Optics, Proc. SPIE* **640**:126–133 (1986).
15. E. L. Church and P. Z. Takacs, "Spectral and Parameter Estimation Arising in the Metrology of High-Performance Mirror Surfaces," *Proc. of ICASSP '86*, S. Saito, ed., pp. 185–188 (1986).

16. E. L. Church and P. Z. Takacs, "Effects of the Optical Transfer Function in Surface Profile Measurements, *Surface Characterization and Testing II*, J. E. Grievenkamp and M. Young, eds., *Proc. SPIE* **1164**:46–59 (1989).
17. E. L. Church and P. Z. Takacs, "Prediction of Mirror Performance from Laboratory Measurements," in *X-ray/EUV Optics for Astronomy and Microscopy*, R. Hoover, ed., *Proc. SPIE* **1160**:323–336 (1989).
18. R. J. Speer, M. Chrisp, D. Turner, S. Mrowka, and K. Tregidgo, "Grazing Incidence Interferometry: The Use of the Linnik Interferometer for Testing Image-Forming Reflection Systems," *Appl. Opt.* **18**(12):2003–2012 (1979).
19. P. H. Langenbeck, "New Developments in Interferometry—VI. Multipass Interferometry," *Appl. Opt.* **8**(3):543–552 (1969).
20. W. H. Wilson and T. D. Eps, *Proc. Phys. Soc.* **32**:326 (1920).
21. R. V. Jones, "Some Points in the Design of Optical Levers," *Proc. Phys. Soc. B* **64**:469–482 (1951).
22. R. V. Jones, *Instruments and Experiences: Papers on Measurement and Instrument Design*, Wiley Series in Measurement Science and Technology, P. H. Sydenham, ed., John Wiley & Sons, Ltd., p. 485, 1988.
23. J. M. Bennett, V. Elings, and K. Kjoller, "Recent Developments in Profiling Optical Surfaces," *Appl. Opt.* **32**(19):3442–3447 (1993).
24. K. Becker and E. Heynacher, "M400—A Coordinate Measuring Machine with 10 nm Resolution," *In-Process Optical Metrology for Precision Machining*, *Proc. SPIE* **802**:209–216 (1987).
25. M. Stedman and V. W. Stanley, "Machine for the Rapid and Accurate Measurement of Profile," *Proc. SPIE* **163**:99–102 (1979).
26. W. J. Wills-Moreen and P. B. Leadbeater, "Stylus Profilometry of Large Optics," presented at San Diego, Calif., *Advanced Optical Manufacturing and Testing*, *Proc. SPIE* **1333**:183–194 (1990).
27. P. S. Young, "Fabrication of the High-Resolution Mirror Assembly for the HEAO-2 X-Ray Telescope," *Proc. SPIE* **184**:131–138 (1979).
28. A. Sarnik and P. Glenn, "Mirror Figure Characterization and Analysis for the Advanced X-Ray Astrophysics Facility/Technology Mirror Assembly (AXAF/TMA) X-Ray Telescope," *Grazing Incidence Optics for Astronomical and Laboratory Applications*, *Proc. SPIE* **830**, S. Bowyer and J. C. Green, eds., pp. 29–36 (1987).
29. J. K. Silk, *A Grazing Incidence Microscope for X-Ray Imaging Applications*, *Annals of the New York Academy of Sciences* **342**:116–129 (1980).
30. R. H. Price, "X-Ray Microscopy Using Grazing Incidence Reflection Optics," *Low Energy X-Ray Diagnostics*, AIP Conf. Proc. **75**, D. T. Attwood and B. L. Henke, eds., pp. 189–199 (1981).
31. A. E. DeCew, Jr. and R. W. Wagner, "An Optical Lever for the Metrology of Grazing Incidence Optics," *Optical Manufacturing, Testing, and Aspheric Optics*, *Proc. SPIE* **645**:127–132 (1986).
32. G. Makosch and B. Solf, "Surface Profiling by Electro-Optical Phase Measurements," *High Resolution Soft X-Ray Optics*, *Proc. SPIE* **316**, E. Spiller, ed., pp. 42–53 (1981).
33. G. Makosch and B. Drollinger, "Surface Profile Measurement with a Scanning Differential AC Interferometer," *Appl. Opt.* **23**(24):4544–4553 (1984).
34. G. Makosch, "LASSI—a Scanning Differential AC Interferometer for Surface Profile and Roughness Measurement," *Surface Measurement and Characterization*, *Proc. SPIE* **1009**, J. M. Bennett, ed., pp. 244–253 (1988).
35. A. E. Ennos and M. S. Virdee, "High Accuracy Profile Measurement of Quasi-Conical Mirror Surfaces by Laser Autocollimation," *Proc. Eng.* **4**:5–9 (1982).
36. A. E. Ennos and M. V. Virdee, "Precision Measurement of Surface Form by Laser Autocollimation," *Industrial Applications of Laser Technology*, *Proc. SPIE* **398**:252–257 (1983).
37. T. C. Bristow and K. Arackellian, "Surface Roughness Measurements using a Nomarski Type Scanning Instrument," *Metrology: Figure and Finish*, *Proc. SPIE* **749**:114–118 (1987).
38. J. M. Eastman and J. M. Zavislan, "A New Optical Surface Microprofiling Instrument," *Precision Surface Metrology*, *Proc. SPIE* **429**:56–64 (1983).
39. T. C. Bristow, G. Wagner, J. R. Bietry, and R. A. Auriemma, "Surface Profile Measurements on Curved Parts," *Surface Characterization and Testing II*, J. Grievenkamp and M. Young, eds., *Proc. SPIE* **1164**:134–141 (1989).
40. P. Glenn, "Angstrom Level Profilometry for Sub-Millimeter to Meter Scale Surface Errors," *Advanced Optical Manufacturing and Testing*, *Proc. SPIE* **1333**, G.M. Sanger, P. B. Ried, and L. R. Baker, eds., pp. 326–336 (1990).

41. P. Glenn, "Robust, Sub-Angstrom Level Mid-Spatial Frequency Profilometry," *Advanced Optical Manufacturing and Testing, Proc. SPIE 1333*, G. M. Sanger, P. B. Ried, and L. R. Baker, eds., pp. 175–181 (1990).
42. I. Weingartner, M. Schulz, and C. Elster, "Novel Scanning Technique for Ultra-Precise Measurement of Topography," *Proc. SPIE 3782*:306–317 (1999).
43. M. Schulz, P. Thomsen-Schmidt, and I. Weingaertner, "Reliable Curvature Sensor for Measuring the Topography of Complex Surfaces," *Optical Devices and Diagnostics in Materials Science, Proc. SPIE 4098*:84–93 (2000).
44. P. Thomsen-Schmidt, M. Schulz, and I. Weingaertner, "Facility for the Curvature-Based Measurement of the Nanotopography of Complex Surfaces," *Proc. SPIE 4098*:94–101 (2000).
45. I. Weingartner and C. Elster, "System of Four Distance Sensors for High-Accuracy Measurement of Topography," *Prec. Eng.* **28**(2):164–170 (2004).
46. I. Weingaertner, M. Wurm, R. Geckeler, et al., "Novel Scheme for the Ultra-Precise and Fast Measurement of the Nanotopography of Large Wafers," *Proc. SPIE 4779*:13–22 (2002).
47. J. Illelmann, R. Geckeler, I. Weingaertner, et al., "Topography Measurement of Nanometer Synchrotron-Optics," *Proc. SPIE 4782*:29–37 (2002).
48. M. Schulz and C. Elster, "Traceable Multiple Sensor System for Measuring Curved Surface Profiles with High Accuracy and High Lateral Resolution," *Opt. Eng.* **45**(6) (2006).
49. M. Schulz, C. Elster, and I. Weingartner, "Coupled Distance Sensor Systems for High-Accuracy Topography Measurement: Accounting for Scanning Stage and Systematic Sensor Errors," *Prec. Eng.* **30**(1):32–38 (2006).
50. K. von Bieren, "Pencil Beam Interferometer for Aspherical Optical Surfaces," *Laser Diagnostics*, S. Holly, ed., *Proc. SPIE 343*:101–108 (1982).
51. K. von Bieren, "Interferometry of Wavefronts Reflected Off Conical Surfaces," *Appl. Opt.* **22**:2109–2114 (1983).
52. P. Z. Takacs, S. Qian, and J. Colbert, "Design of a Long-Trace Surface Profiler," *Metrology—Figure and Finish, Proc. SPIE 749*, B. Truax, ed., pp. 59–64 (1987).
53. P. Z. Takacs, S. K. Feng, E. L. Church, S. Qian, and W. Liu, "Long Trace Profile Measurements on Cylindrical Aspheres," *Advances in Fabrication and Metrology for Optics and Large Optics, Proc. SPIE 966*, J. B. Arnold and R. A. Parks, eds., pp. 354–364 (1988).
54. P. Z. Takacs, K. Furenlid, R. DeBiaise, and E. L. Church, "Surface Topography Measurements over the 1 meter to 10 micrometer Spatial Period Bandwidth," *Surface Characterization and Testing II, Proc. SPIE 1164*, J. E. Grievenkamp and M. Young, eds., pp. 203–211 (1989).
55. H. Lammert, F. Siewert, and T. Zeschke, *The Nano-Optic-Measuring Machine NOM at BESSY—Further Improvement of the Measuring Accuracy*, in *BESSY Annual Report 2005*. BESSY GmbH: Berlin, Germany. pp. 481–486 (2006).
56. S. Qian, W. Jark, and P. Z. Takacs, "The Penta-Prism LTP: A Long-Trace-Profiler with Stationary Optical Head and Moving Penta-Prism," *Rev. Sci. Instrum.* **66**:2187 (1995).
57. S. Qian, W. Jark, P. Z. Takacs, K. J. Randall, and W. Yun, "In-Situ Surface Profiler for High Heat Load Mirror Measurement," *Opt. Eng.* **34**(2):396–402 (1995).
58. S. Qian, W. Jark, and P. Z. Takacs, "The Penta-Prism LTP: A Long-Trace-Profiler with Stationary Optical Head and Moving Penta-Prism," *Rev. Sci. Instrum.* **66**:2562–2569 (1995).
59. S. Qian, W. Jark, P. Z. Takacs, K. J. Randall, Z. Xu, and W. Yun, "In-Situ Long Trace Profiler for Measurement of Mirror Profiles at Third Generation Synchrotron Facilities," *Rev. Sci. Instrum.* **67**:3369 (1996).
60. S. Qian, H. Li, and P. Z. Takacs, "Penta-Prism Long Trace Profiler (PPLTP) for Measurement of Grazing Incidence Space Optics," *Multilayer and Grazing Incidence X-Ray/EUV Optics III, Proc. SPIE, 2805*, R. Hoover and A. B. C. Walker, Jr., eds., pp. 108–114 (1996).
61. S. Qian, W. Jark, G. Sostero, A. Gambitta, et al., *Penta-prism LTP Detects First In-Situ Distortion Profile*, *Synchrotron Radiation News* **9**(3), pp. 42–44 (1996).
62. S. Qian, W. Jark, G. Sostero, A. Gambitta, et al., "Advantages of the In-Situ LTP Distortion Profile Test on High-Heat-Load Mirrors and Applications," *Proc. SPIE 2856*, L. E. Berman and J. Arthur, eds., pp. 172–182 (1996).
63. H. Li, X. Li, M. W. Grindel, and P. Z. Takacs, "Measurement of X-Ray Telescope Mirrors Using A Vertical Scanning Long Trace Profiler," *Opt. Eng.* **35**(2):330–338 (1996).

64. H. Li, P. Z. Takacs, and T. Oversluizen, "Vertical Scanning Long Trace Profiler: a Tool for Metrology of X-Ray Mirrors," *Materials, Manufacturing, and Measurement for Synchrotron Radiation Mirrors, Proc. SPIE* **3152**:180–187 (1997).
65. M. Gubarev, T. Kester, and P. Z. Takacs, "Calibration of a Vertical-Scan Long Trace Profiler at MSFC," *Optical Manufacturing and Testing IV, Proc. SPIE* **4451**, H. P. Stahl, ed., pp. 333–339 (2001).
66. P. Z. Takacs, H. Li, X. Li, and M. W. Grindel, "3-D X-Ray Mirror Metrology with a Vertical Scanning Long Trace Profiler," *Rev. Sci. Instrum.* **67**, G. K. Shenoy and J. L. Dehmer, eds., pp. 3368–3369 (1996).
67. P. Z. Takacs, S. Qian, K. J. Randall, W. Yun, and H. Li, "Mirror Distortion Measurements with an In-Situ LTP," *Advances in Mirror Technology for Synchrotron X-Ray and Laser Applications, Proc. SPIE* **3447**:117–124 (1998).
68. ASME, *ASME B46.1-2002—Surface Texture (Surface Roughness, Waviness, and Lay)*, The American Society of Mechanical Engineers: New York (2002).
69. ISO, *ISO 4287:1997 Geometrical Product Specifications (GPS)—Surface Texture: Profile Method—Terms, Definitions and Surface Texture Parameters*, International Organization for Standardization, 1997.
70. SEMI, *SEMI MF1811-0704 Guide for Estimating the Power Spectral Density Function and Related Finish Parameters from Surface Profile Data*, Semiconductor Equipment and Materials International: San Jose, Calif., 2004.
71. J. C. Stover, *Optical Scattering: Measurement and Analysis*, 2d ed., Bellingham, WA: SPIE Press, p. 340, 1995.
72. E. L. Church and H. C. Berry, "Spectral Analysis of the Finish of Polished Optical Surfaces," *WEAR* **83**:189–201 (1982).
73. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes—the Art of Scientific Computing*, Cambridge: Cambridge University Press, p. 818, 1986.
74. D. E. Newland, *An Introduction to Random Vibrations, Spectral and Wavelet Analysis*, 3d ed., New York: Wiley, p. 477, 1993.
75. C. K. Yuen and D. Fraser, *Digital Spectral Analysis*, London: Pitman Publishing Ltd., p. 156, 1979,
76. M. Born and E. Wolf, *Principles of Optics*, 6th ed., New York: Pergamon Press, p. 808, 1980.
77. E. L. Church, T. V. Vorburger, and J. C. Wyant, "Direct Comparison of Mechanical and Optical Measurements of the Finish of Precision Machined and Optical Surfaces," *Opt. Eng.* **24**(3):388–395 (1985).
78. V. V. Yashchuk, et al., "Cross-Check of Different Techniques for Two-Dimensional Power Spectral Density Measurements of X-Ray Optics," *Advances in Metrology for X-Ray and EUV Optics, Proc. SPIE* **5921**, L. Assoufid, P. Z. Takacs, and J. S. Taylor, eds., p. 59210G (2005).
79. V. V. Yashchuk, A. D. Franck, S. C. Irick, M. R. Howells, A. A. MacDowell, and W. R. McKinney, "Two-Dimensional Power Spectral Density Measurements of X-Ray Optics with the Micromap Interferometric Microscope," *Nano- and Micro-Metrology, Proc. SPIE* **5858**, H. Ottevaere, P. DeWolf, and D. S. P. O. Wiersma, eds., p. 58580A (2005).

This page intentionally left blank

ASTRONOMICAL X-RAY OPTICS

Marshall K. Joy and Brian D. Ramsey

*National Aeronautics and Space Administration
Marshall Space Flight Center
Huntsville, Alabama*

Over the past three decades, grazing incidence optics has transformed observational x-ray astronomy into a major scientific discipline at the cutting edge of research in astrophysics and cosmology. This chapter summarizes the design principles of grazing incidence optics for astronomical applications, describes the capabilities of the current generation of x-ray telescopes and the techniques used in their fabrication, and explores avenues of future development.

47.1 INTRODUCTION

The first detection of a cosmic x-ray source outside of our solar system was made during a brief rocket flight less than 50 years ago.¹ This flight, which discovered a bright source of x rays in the Scorpius Constellation, was quickly followed by other suborbital experiments and in 1970, the first dedicated x-ray astronomy satellite, UHURU, was launched into an equatorial orbit from Kenya.² UHURU, which used mechanically collimated gas-filled detectors, operated for just over 2 years and produced a catalog of 339 cosmic x-ray sources.

While UHURU significantly advanced the discipline, the real revolution in x-ray astronomy came about with the introduction of grazing-incidence optics aboard the Einstein observatory in 1978.³ Focusing optics provide an enormous increase in signal to noise ratio by concentrating source photons into a tiny region of the detector, thereby reducing the detector-area-dependent background to a very small value. Despite a modest collecting area, less than that of the UHURU, the Einstein observatory had two-to-three orders of magnitude more sensitivity, enabling emission from a wide range of sources to be detected and changing our view of the x-ray sky.

Since that time, payloads have increased in capability and sophistication. The current “flagship” x-ray astronomy missions are the U.S.-led Chandra observatory and the European-led XMM-Newton observatory. Chandra represents the state of the art in astronomical x-ray optics with sub-arcsecond on-axis angular resolution and about 0.1 m² of effective collecting area.⁴ Its sensitivity is over five orders of magnitude greater than that of UHURU, despite having only slightly greater collecting area; in deep fields Chandra resolves more than 1000 sources per square degree.

The XMM-Newton observatory, designed for high throughput, has nearly 0.5 m^2 of effective collecting area, and 15-arcsecond-level angular resolution.⁵

In addition to providing a considerable increase in sensitivity and enabling fine imaging, x-ray optics also permit the use of small-format, high-performance focal-plane detectors. Both Chandra and XMM feature fine-pixel silicon imagers with energy resolutions an order of magnitude greater than the earlier gas-filled detectors. Currently planned missions will utilize imaging x-ray calorimeters that will offer one to two orders of magnitude further spectroscopic improvement.

This chapter describes the optics used in, or with potential for use in, x-ray astronomy. The missions described above use mirror geometries based on a design first articulated by Wolter⁶ and these, and their fabrication techniques, are described in Sec. 47.2. An alternate mirror configuration, termed Kirkpatrick-Baez, which has not seen use yet in astronomy but offers potential future benefit, is described in Sec. 47.3. Payloads designed to extend the range of x-ray focusing optics into the hard-x-ray region are detailed in Sec. 47.4. Finally, new developments offering the promise of ultra-high angular resolution for future missions are discussed in Sec. 47.5.

47.2 WOLTER X-RAY OPTICS

Optical Design and Angular Resolution

Wolter optics are formed by grazing-incidence reflections off two concentric conic sections (a paraboloid and hyperboloid, or a paraboloid and ellipsoid—see Chap. 44). The most common case (Wolter type I) is conceptually similar to the familiar Cassegrain optical telescope: the incoming parallel beam of x rays first reflects from the parabolic section and then from the hyperbolic section, forming an image at the focus (Fig. 1). To increase the collecting area, reflecting shells of different diameters are nested, with a common focal plane.

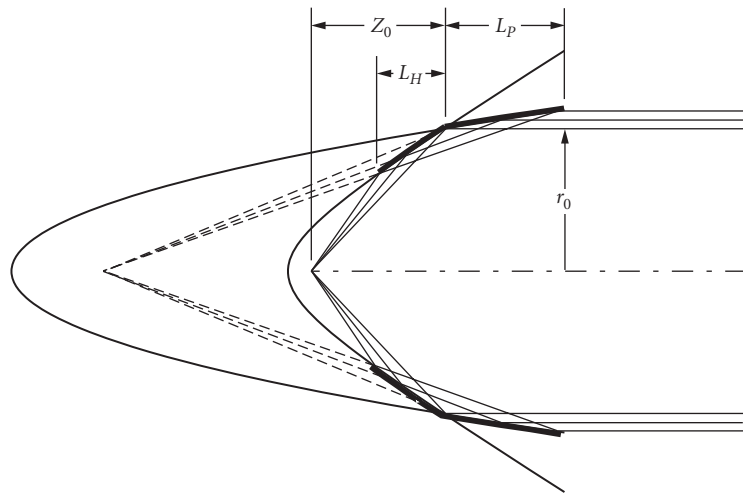


FIGURE 1 Geometry of a Wolter type I x-ray optic. Parallel light incident from the right is reflected at grazing incidence on the interior surfaces of the parabolic and hyperbolic sections; the image plane is at the focus of the hyperboloid.

A Wolter I optic can be described by four quantities:

1. The focal length, Z_0 , which is defined as the distance from the intersection of the paraboloid and hyperboloid to the focal point
2. The mean grazing angle α , which is defined in terms of the radius of the optic at the intersection plane r_0 :

$$\alpha \equiv \frac{1}{4} \arctan \frac{r_0}{Z_0} \quad (1)$$

3. The ratio of the grazing angles of the paraboloid and the hyperboloid ξ , measured at the intersection point
4. The length of the paraboloid L_p

Wolter optics produce a curved image plane, and have aberrations which can cause the angular resolution of the optic to be significantly worsened for x-ray sources that are displaced from the optical axis of the telescope (see Chap. 45); for these reasons, designs are usually optimized using detailed ray-trace simulations (see Chap. 35). However, a good approximation to the optimum design can be readily obtained using the results of Van Speybroeck and Chase.⁷ The highest x-ray energy that the optic must transmit largely determines the mean grazing angle (see “X-Ray Reflectivity” section), which in turn constrains the focal ratio of the optic [Eq. (1)]. The grazing angles on the parabolic and hyperbolic sections are usually comparable, so $\xi \approx 1$. With the diameter and length of the optic as free parameters, the curves in Van Speybroeck and Chase can be used to estimate the angular resolution and the collecting area for different designs. Very high resolution and good off-axis performance is possible with Wolter I optics. Figure 2 presents sub-arcsecond angular resolution images from the Chandra X-Ray Observatory and wide field images from the XMM-Newton Observatory.

Mirror Figure and Surface Roughness

Irregularities in the mirror surface will cause light to be scattered out of the core of the x-ray image, degrading the angular resolution of the telescope. If an incoming x ray strikes an area of the mirror surface that is displaced from the ideal height by an amount σ , the resulting optical path difference is given by

$$\text{OPD} = 2\sigma \sin \alpha \quad (2)$$

and the corresponding phase difference is

$$\Delta = \frac{4\pi\sigma \sin \alpha}{\lambda} \quad (3)$$

where α is the grazing angle and λ is the x-ray wavelength. For a uniformly rough mirror surface with a gaussian height distribution, RMS values of σ and Δ can be used to calculate the scattered intensity relative to the total intensity:^{8,9}

$$\frac{I_s}{I_0} = 1 - e^{-\Delta_{\text{RMS}}^2} \quad (4)$$

This result implies that high-quality x-ray reflectors must have exceptionally smooth surfaces: in order for the scattered intensity I_s to be small, Δ_{RMS} must be $\ll 1$. For a grazing angle of 0.5° and a wavelength of $1 \text{ \AA}$, a surface with an RMS microroughness of $4 \text{ \AA}$ will scatter approximately 20 percent of the incident intensity. A surface roughness of $8 \text{ \AA}$ will scatter more than 50 percent of the incident intensity at this x-ray wavelength and grazing angle.

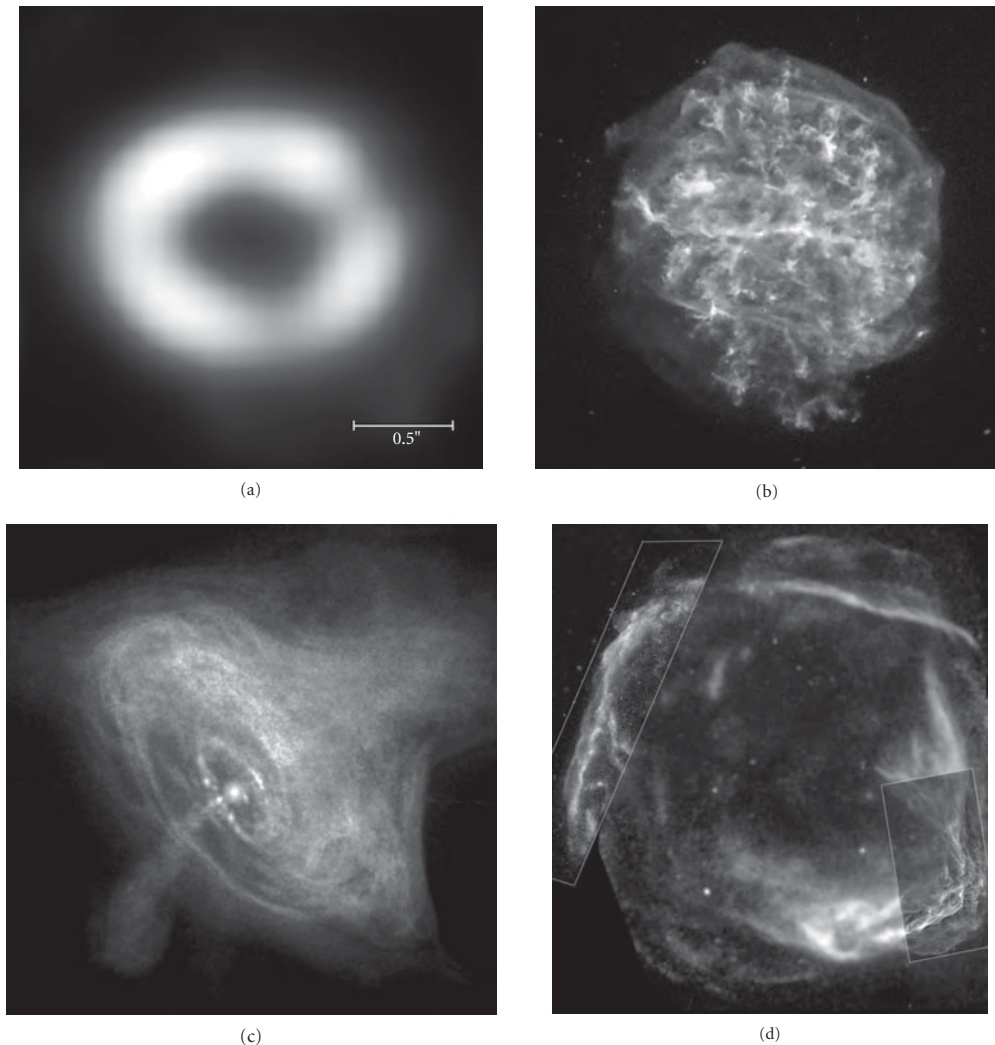


FIGURE 2 (a) A Chandra x-ray image of the fiery ring surrounding the 1987 supernova explosion in the Dorado constellation. Subarcsecond angular resolution is required to resolve the structure surrounding the supernova remnant (<http://chandra.harvard.edu/photo/2005/sn87a/>). (NASA/CXC/PSU/S.Park & D. Burrows.) (b) A Chandra x-ray image of the supernova remnant G292.0+1.8. The colors in the image encode the x-ray energies emitted by the supernova remnant; the center of G292.0+1.8 contains a region of high energy x-ray emission from the magnetized bubble of high-energy particles that surround the pulsar, a rapidly rotating neutron star that remained behind after the original, massive star exploded (<http://chandra.harvard.edu/photo/2007/g292/>). (NASA/CXC/Penn State/S.Park *et al.*) (c) Chandra x-ray image of the Crab Nebula—the remains of a nearby supernova explosion which was seen on Earth in 1054 AD. At the center of the bright nebula is a rapidly spinning neutron star, or pulsar, that emits pulses of radiation 30 times a second (<http://chandra.harvard.edu/photo/2002/0052/>). (NASA/CXC/ASU/J. Hester *et al.*) (d) XMM-Newton and Chandra x-ray images RCW 86, an expanding ring of debris that was created after a massive star in the Milky Way collapsed onto itself and exploded. Both the XMM-Newton and Chandra images show low-energy x-rays in red, medium energies in green and high energies in blue. The Chandra observations focused on the northeast (left-hand) and southwest (lower right) side of RCW 86, and show that x-ray radiation is produced both by high-energy electrons accelerated in a magnetic field (blue) as well as heat from the blast itself (red). These images demonstrate the large field of view and moderate angular resolution of XMM-Newton, compared to the smaller field of view and high angular resolution provided by Chandra (<http://chandra.harvard.edu/photo/2007/2snr/>). (NASA/CXC/ESA/Univ. of Utrecht/J. Vink *et al.*) (See also color insert.)

To calculate the effect of surface imperfections on the x-ray point spread function in detail, it is necessary to measure the distribution of the deviations on all spatial scales of the optic, from the microroughness on submicron scales to the overall slope error of the full mirror. The surface deviations are characterized by the power spectral density, and the resulting x-ray scattering can be calculated using the methods described by Church,¹⁰ Aschenbach,⁸ O'Dell et al.,¹¹ and Hughes et al.¹²

X-Ray Reflectivity

For most astronomical applications, x-ray telescopes are required to have high reflection efficiency across the operational energy band, and also to function at the highest possible energy for a given focal length and diameter. The overall reflection efficiency and the high energy response can both be increased by applying a high density coating to the polished reflecting surfaces of the optics, such as nickel, gold, platinum, or iridium (the latter element having the highest density and the best reflectivity).¹³ As discussed earlier, the uniformity and smoothness of the reflective coating is critical to the imaging performance of the optic.

The reflection efficiency of an x-ray optic is strongly dependent on energy, due to the presence of atomic absorption edges and the rapid decrease in efficiency at high energies. The reflectivity can be readily calculated using the Fresnel reflection formulae expressed in terms of the complex dielectric constant of the reflecting surface, κ .¹³ Dielectric constants can be easily derived from experimentally determined atomic scattering factors, f , using the relation

$$\kappa = 1 - \frac{r_e \rho N_A}{\pi A_w} \lambda^2 f \quad (5)$$

where r_e is the classical electron radius, N_A is Avogadro's number, and ρ and A_w are the density and molecular weight of the reflecting material. An extensive compilation of atomic scattering factors for elements 1 to 92 over the energy range 50 eV to 30 keV is given by Henke, Gullikson, and Davis¹⁴ (also see Chap. 36).

Effective Collecting Area

The useful collecting area of a grazing incidence optic depends on several factors: the size and number of mirror shells, the projection of the grazing incidence mirrors onto the sky, blockage of the aperture by support structures, x-ray reflection efficiency, and vignetting of off-axis sources. The effective collecting area A_e is the convolution of all of these terms; it is highly energy dependent due to the mirror reflectivity, and falls to zero at the high energy cutoff of the optic.

Technologies for Fabricating Wolter-Type X-Ray Optics

There are currently three primary methods for fabricating nested Wolter x-ray optics for astronomical applications. The approach that has produced the highest angular resolution involves the fabrication of Zerodur glass shells, a few centimeters thick, which are ground, figured, highly polished, and coated on the interior reflecting surface. Optics of this type have been built and flown on the Einstein observatory,³ ROSAT,¹⁵ and the Chandra X-Ray Observatory.¹⁶ Einstein and ROSAT each had approximately 5 arcsec angular resolution, while Chandra has sub-arcsecond on-axis resolution. These optics represent what might be called the traditional approach to mirror fabrication which involves meticulous figuring and polishing of relatively heavy substrates. The resulting mirrors deliver superb performance, but in an effort to avoid the high cost necessarily incurred by this labor-intensive work and the significant weight of the mirrors and their support structures, other fabrication techniques have been developed that trade angular resolution for light weight, higher throughput, and lower cost.

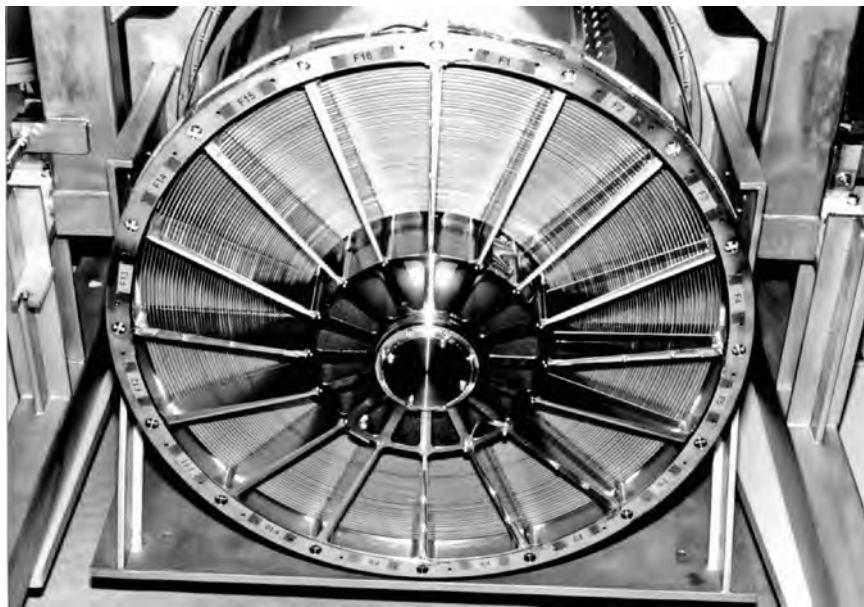


FIGURE 3 An x-ray optics module for the XMM observatory. Fifty-eight electroformed nickel Wolter I optics are nested to increase the effective x-ray collecting area. (See also color insert.) (Photo courtesy of ESA.)

A mirror-fabrication process used widely in x-ray astronomy is electroformed nickel replication (ENR). This technique makes use of a master or “mandrel” which is used to replicate identical mirror shells. The required parabolic and hyperbolic surfaces are machined into the outer surface of the mandrel, which is then highly polished and coated or chemically treated. A thin nickel shell is electroplated onto the mandrel; the shell and mandrel (which have different coefficients of thermal expansion) are then separated by cooling. The attraction of the process is that the resulting shells are full cylinders that contain both parabolic and hyperbolic segments of the Wolter geometry. This makes them inherently stable and results in reasonable angular resolution despite their thin (1 mm or less) walls. A good example of the use of the ENR process is in the mirrors for the XMM-Newton Observatory which has three optics modules, each containing 58 replicated Wolter-I shells (Fig. 3), which provide high effective collecting area and 15-arcsecond angular resolution.¹⁷

A third approach to mirror fabrication is to utilize reflector segments mounted in a housing to give the desired mirror geometry. This technique opens the possibility of compact, very-light-weight, densely-packed optics modules. To date, many segmented optics have flown on satellite missions and all have utilized thin aluminum reflectors coated (typically via an epoxy replication off a smooth mandrel) to improve their surface roughness. The current SUZAKU¹⁸ x-ray astronomy mission has several foil segment mirror modules (Fig. 4). Typically these have around 175 nested shells, each consisting of $8 \times 150 \mu\text{m}$ thick reflectors. The on-orbit measured angular resolution is around 2 arcminutes.

Recently there has been a significant optics development effort utilizing slumped glass segmented reflectors rather than aluminum. The attraction of the glass is that it is available with very smooth surfaces due to the fabrication process used. Certain glass types, notably float borosilicate glass, is obtainable with sub-5-Å smoothness making it an ideal x-ray reflector even for demanding multilayer coatings. The NuSTAR mission,¹⁹ recently selected as a NASA small explorer, utilizes float glass reflectors “slumped” by heating the glass to approximately 600°C on a suitably figured pyrex mandrel. The reflectors are coated with multilayers for hard-x-ray responsivity (see Sec. 47.4), and

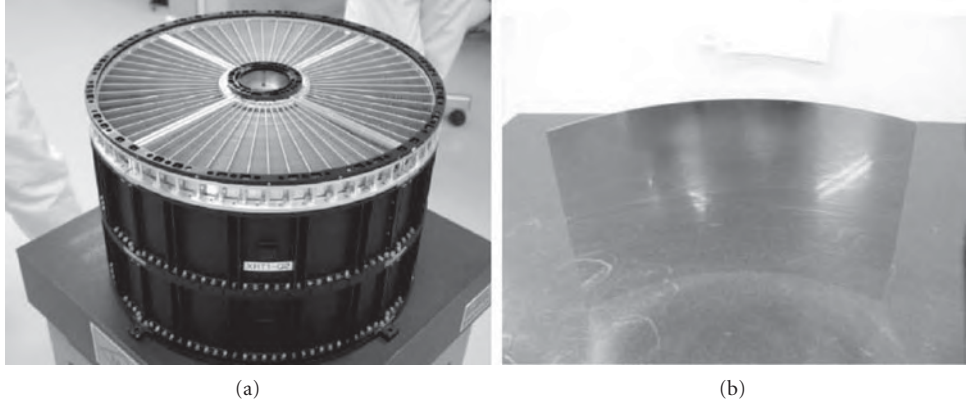


FIGURE 4 The segmented foil mirrors aboard the SUZAKU spacecraft. (a) A complete mirror module and (b) a single aluminum foil reflector coated with gold. (See also color insert.)

are assembled into a module through the use of machined precision graphite spacers inserted and glued between successive shells. The spacers are machined in situ, before each successive shell is glued, to prevent error buildup. The expected angular resolution for the NuSTAR optics is approximately 50 arcseconds.

47.3 KIRKPATRICK-BAEZ OPTICS

The first demonstration of grazing incidence x-ray imaging was performed by Kirkpatrick and Baez²⁰ using mirrors with one-dimensional parabolic curvature (Fig. 5a). An optic of this type focuses in one plane only, so two mirrors are mounted orthogonal to each other to achieve 2D imaging (Fig. 5b); they can also be stacked to increase the effective collecting area (Fig. 5c). Design studies of multielement Kirkpatrick-Baez systems have been carried out by Van Speybroeck et al.,²¹ Weisskopf,²² and Kast.²³

Kirkpatrick-Baez optics has not been widely used in astronomical applications because the required grazing angles are a factor of two larger than those of a Wolter design, for the same optic diameter and focal length. (This can be seen from Fig. 7: An x-ray reflecting in the horizontal plane receives the full angular displacement in a single reflection, while a Wolter optic achieves the same angular displacement via two half-angle reflections.) The larger grazing angles in the Kirkpatrick-Baez design significantly reduce the high energy efficiency of the optic. However, optimal packing schemes and ease of manufacture may eventually permit Kirkpatrick-Baez designs to overcome this disadvantage; for example, mass-produced flat plates can be substituted for the curved optics in cases where the curvature of the mirrors is small (Fig. 6). In the small angle approximation the projected width of the flat plate is

$$\delta = \frac{Lr_0}{2F} \quad (6)$$

where L is the length of the plate, r_0 is the distance of the plate from the optical axis, and F is the focal length. The limiting angular resolution due to the flat plate geometry, $\Delta\theta$, is

$$\Delta\theta(\text{arcsec}) \sim 10^5 \frac{Lr_0}{F^2} \quad (7)$$

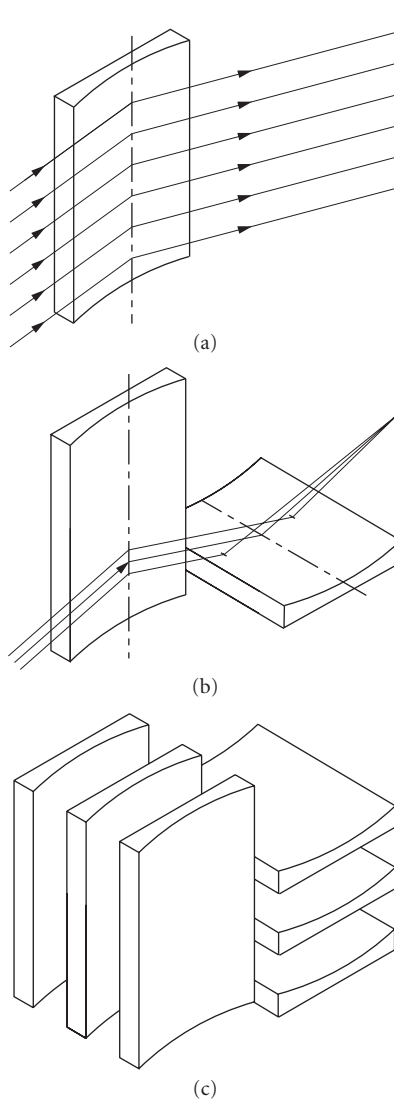


FIGURE 5 Kirkpatrick-Baez optics. (a) Mirrors with one-dimensional parabolic curvature focus in one plane only, producing a line image; (b) two mirrors mounted orthogonal to each other can produce 2D imaging; and (c) parabolic Kirkpatrick-Baez optics can be nested to increase the effective collecting area.

Equation (7) indicates that the error introduced by the flat plate approximation can be made reasonably small: $\Delta\theta = 10$ arcsec for a mirror 0.1 m in length and $r_0 = 0.1$ m away from the axis of a 10-m focal length optic. Analytic expressions for the number of flat plates, their spacing, and the resulting packing fraction are given by Weisskopf.²²

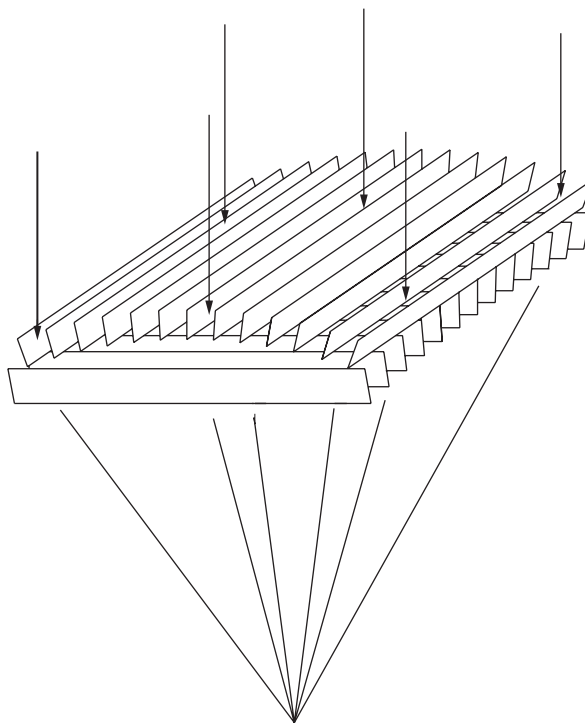


FIGURE 6 Flat plates can be substituted for the parabolic Kirkpatrick-Baez optics in cases where the curvature of the mirrors is small, and 2D imaging can be achieved by stacking two orthogonal sets of flat reflectors.

47.4 HARD X-RAY OPTICS

As noted in Sec. 47.1, x-ray optics has revolutionized the field of x-ray astronomy, with current state-of-the-art telescopes having more than five orders of magnitude greater sensitivity than the first satellite-borne instruments. This statement is only true, however, below approximately 10 keV where x-ray optics have been used extensively. The hard-x-ray region, above 10 keV, is very important for study as in this energy regime sources transit from thermal to nonthermal emission mechanisms. In addition, nuclear lines appear and obscured objects become visible. Despite this significance, this energy region remains relatively unexplored at high sensitivities and fine angular scales.

The difficulty inherent in obtaining high-energy response is that the critical angle, below which x rays can be reflected, is given by

$$\phi_c \sim 0.93 \cdot \lambda \cdot \sqrt{\rho}$$

where λ is the wavelength of the incident x ray (nanometers) and ρ is the density of the reflecting medium (grams per cubic centimeter). Thus at 6 keV, for example, a nickel surface will reflect x rays up to an angle of approximately 1/2 degree, but at 60 keV, this will be only 3 arcminutes and thus the projected collecting area of hard-x-ray mirror shells becomes very small.

There have recently been several balloon payloads developed with hard-x-ray telescope systems using slightly different approaches to obtaining high-energy response. The HERO balloon payload^{24,25} uses an array of shallow-graze-angle, iridium-coated optics fabricated using the electroformed nickel replication process, while the HEFT²⁶ and the InFOCUS/SUMIT²⁷ payloads use multilayers (see Chap. 41) to extend the graze angles over which useful reflectivity can be obtained. Both InFOCUS/SUMIT and HEFT utilize a segmented optics approach, the former with aluminum reflectors and the latter with slumped glass segments as described in “Technologies for Fabricating Wolter-Type X-Ray Optics” section.

There are also various satellite missions with hard-x-ray focusing optics that have been approved and are in the early stages of development. Among these are the US NuSTAR small explorer mission,^{19,28} the Japanese NeXT (ASTRO-h) mission.^{27,29}

NuSTAR is a small explorer mission that has recently been approved for flight with a planned launch in 2011. It is an outgrowth of the HEFT balloon program, described previously, but with an increased focal length and glass-mirror-module size to give an effective area around 200 cm² at 50 keV and useful response to above 70 keV. This, coupled with the long integration times available on orbit, results in a two-order-of magnitude increase in sensitivity over currently orbiting (non-focusing) hard-x-ray instruments. Launched in an equatorial orbit for low background, NuSTAR will have a 2-year mission lifetime.

The Japanese NeXT mission builds on the InFOCUS/SUMIT experience, and will utilize multilayer-coated aluminum foil optics for extended response up to approximately 60 keV. It is projected to fly in the 2013/2014 timeframe.

47.5 TOWARD HIGHER ANGULAR RESOLUTION

The next generation of x-ray missions requires ever-higher angular resolutions and collecting areas and it is difficult to construct grazing incidence x-ray optics that combine both of these. The Chandra X-ray Observatory currently delivers sub-arcsecond angular resolution, a superlative achievement for an x-ray telescope, but still an order of magnitude below the resolution achieved by optical observatories, and more than three orders of magnitude below the diffraction limit. Improving conventional Wolter x-ray optics to produce milli-arcsecond (or better) angular resolution is a formidable technical challenge. One alternative is to exploit an approach that is widely used in the radio and optical spectral regions: interferometry.

Several fundamental questions about the feasibility of x-ray interferometry are immediately apparent. First, interferometry requires that the phase errors be controlled to a fraction of a wavelength; is it possible to figure and control optics for x-ray wavelengths on the scale of Angstroms? Second, cosmic x-ray sources are notoriously faint (especially the cosmologically interesting sources at high redshift), and the effective collecting areas of x-ray telescopes are low; is it possible to build an x-ray interferometer that would collect a sufficient number of photons to construct a useful image in a finite amount of time?

The scientific case for very-high-angular resolution x-ray imaging proves to be strong. Many of the most interesting objects in the universe (including active galactic nuclei, accretion disks, and the black holes that power them) are very small (< 1 milli-arcsec), very hot ($T > 10^8$ K), and radiate a large amount of thermal energy ($B_\nu \propto T^4$) primarily in the x-ray spectral region. Obtaining the highest scientific returns, however, will require microarcsecond x-ray angular resolution.³⁰

The answers to some of the technical feasibility questions are interesting and rather unexpected. Figure 7 shows a schematic drawing of an x-ray interferometer, in which the beams from two x-ray telescopes are mixed to form fringes. Grazing incidence mirrors are used for both the telescope optics and the beam-combining optics.

How accurately must the grazing incidence optics be figured and controlled in order to achieve an acceptable phase error at x-ray wavelengths? For a $1/2$ degree grazing incidence optic, $\lambda = 2$ Å, and a $\lambda/10$ path difference, Eq. (2) indicates that the tolerance on surface deviations is large

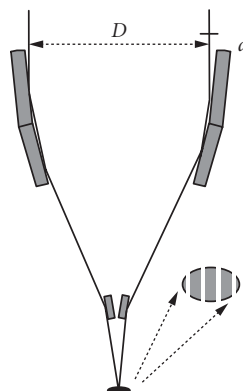


FIGURE 7 Schematic drawing of a grazing-incidence x-ray interferometer.

compared to the wavelength: $\sigma \approx 10 \text{ \AA}$. Thus, for typical grazing angles, the factor of $\sin\alpha$ in Eq. (2) relaxes the tolerances on the mirror surface by approximately two orders of magnitude. Significantly, this brings the required x-ray optics and metrology solidly into a regime that has been demonstrated in the laboratory, suggesting the future possibility of an ultra-high-angular resolution x-ray astronomy mission.

47.6 REFERENCES

1. R. Giacconi, H. Gursky, F. Paolini, and B. Rossi, "Evidence for X-Rays from Sources Outside the Solar System," *Phys. Rev. Lett.* **9**:435 (1962).
2. R. Giacconi, E. Kellogg, P. Gorenstein, H. Gursky, and H. Tananbaum, "An X-Ray Scan of the Galactic Plane from UHURU," *Ap. J.* **165**: L27 (1971).
3. R. Giacconi, G. Branduardi, U. Briel, A. Epstein, D. Fabricant, E. Feigelson, W. Forman, et al., "The Einstein/HEAO 2/X-Ray Observatory," *Ap. J.* **230**:540 (1979).
4. M. C. Weisskopf and S. L. O'Dell, "Calibration of the AXAF Observatory: Overview," *Proc. SPIE* **3113**:2 (1997).
5. P. Gondoin, B. R. Aschenbach, M. W. Beijersbergen, R. Egger, F. A. Jansen, Y. Stockman, and Tock, J. -P., "Calibration of the First XMM Flight Mirror Module: I. Image Quality," *Proc. SPIE* **3444**:278 (1998).
6. H. Wolter, "Grazing Incidence Mirror Systems as Imaging Optics for X-Rays," *Ann. Phys.* **10** (Ser. 6):94 (1952).
7. L. P. Van Speybroeck, and R. C. Chase, "Design Parameters of Paraboloid-Hyperboloid Telescopes for X-Ray Astronomy," *Appl. Opt.* **11**:440 (1972).
8. B. Aschenbach, "X-Ray Telescopes," *Rep. Prog. Phys.* **48**:579–629 (1985).
9. P. Beckmann, and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, Oxford: Pergamon (1963).
10. E. L. Church, "Role of Surface Topography in X-Ray Scattering," *Proc. SPIE* **184**: 196 (1979).
11. S. L. O'Dell, R. F. Elsner, J. J. Kolodziejczak, M. Weisskopf, J. P. Hughes, and L. P. van Speybroeck, "X-Ray Evidence for Particulate Contamination on the AXAF VETA-I Mirrors," *Proc. SPIE* **1742**:171 (1993).
12. J. P. Hughes, D. Schwartz, A. Szentgyorgyi, L. Van Speybroeck, and P. Zhao, "Surface Finish Quality of the Outer AXAF Mirror Pair Based on X-Ray Measurements of the VETA-I," *Proc. SPIE* **1742**:152 (1993).
13. R. F. Elsner, S. L. O'Dell, and M. C. Weisskopf, "Effective Area of the AXAF X-Ray Telescope: Dependence upon Dielectric Constants of Coating Materials," *J. X-Ray Sci. Technol.* **3**, 35–44 (1991).

14. B. L. Henke, E. M. Gullikson, and 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* **54**(2): 181–342 (1993). This data, along with useful computational tools, is also available on the internet at http://www-cxro.lbl.gov/optical_constants.
15. B. Aschenbach, "First Results from the X-Ray Astronomy Mission ROSAT," *Rev. Mod. Astron.* **4**:173 (1991).
16. M.C. Weisskopf, "Five Years of Operation of the Chandra X-Ray Observatory," *Proc. SPIE* **5488**:25 (2004).
17. Y. Stockman, P. Barzin, H. Hansen, E. Mazy, J. P. Tock, D. de Chambure, R. Laine, D. Kampf, et al., "XMM Flight Model Mirror Module Testing," *Proc. SPIE* **3766**:51 (1999).
18. K. Mitsuda, M. Bautz, H. Inoue, R. L. Kelley, K. Koyama, H. Kunieda, K. Makishima, et al., "The X-Ray Observatory Suzaku," *Publications of the ASJ* **59**:1–7 (2007).
19. F. A. Harrison, F. E. Christensen, W. Craig, C. Hailey, W. Baumgartner, C. M. H. Chen, J. Chonko, et al., "Development of the HEFT and NuSTAR Focusing Telescopes," *Exp. Astr.* **20**:131 (2005).
20. P. Kirkpatrick, and A. V. Baez, "Formation of Optical Images by X-Rays," *J. Opt. Soc. Am.* **38**:776 (1948).
21. L. P. Van Speybroeck, R. C. Chase, and T. F. Zehnpfennig, "Orthogonal Mirror Telescopes for X-Ray Astronomy," *Appl. Opt.* **10**:945 (1971).
22. M. C. Weisskopf, "Design of Grazing-Incidence X-Ray Telescopes," *Appl. Opt.* **12**:1436 (1973).
23. J. W. Kast, "Scanning Kirkpatrick-Baez X-Ray Telescope to Maximize Effective Area and Eliminate Spurious Images," *Appl. Opt.* **14**:537 (1975).
24. B. D. Ramsey, C. D. Alexander, J. A. Apple, C. M. Benson, K. L. Dietz, R. F. Elsner, D. Engelhaupt, et al., "First Images from HERO: A Hard-X-Ray Focusing Telescope," *J. Astrophys.* **568**:432–435 (2002).
25. B. D. Ramsey, R. Elsner, D. Engelhaupt, M. Gubarev, J. J. Dell Kolodziejczak, C. O. S. L. O Speegle, and M. C. Weisskopf, "The Development of Hard-X-Ray Optics at NASA/MSFC," *Proc. SPIE* **5168**:129–135 (2004).
26. J. E. Koglin, F. E. Christensen, J. Chonko, W. W. Craig, T. R. Decker, M. A. Jimenez-Garate, K. S. Gunderson, et al., "Development and Production of Hard X-Ray Multilayer Optics for HEFT," *Proc. SPIE* **4851**: 607–618 (2003).
27. Y. Ogasaka, K. Tamura, T. Miyazawa, Y. Fukaya, T. Iwahara, N. Sasaki, A. Furuzawa, et al., "Thin-Foil Multilayer-Supermirror Hard X-Ray Telescope for InFOC μ S/SUMIT Balloon Experiments and NeXT Satellite Program," *Proc. SPIE* **6688**:03 (2007).
28. J. E. Koglin, F. E. Christensen, W. Craig, T. R. Decker, C. J. Hailey, F. A. Harrison, C. Hawthorn, et al., "NuSTAR Hard X-Ray Optics," *Proc. SPIE* **5900**(31) (2005).
29. Y. Ogasaka, K. Tamura, R. Shibata, A. Furuzawa, T. Okajima, K. Yamashita, Y. Tawara, H. Kuneida, "NeXT Hard X-Ray Telescope," *Proc. SPIE* **5488**:148–155 (2004).
30. K. Gendreau, N. White, S. Owens, W. Cash, A. Shipley, and M. Joy, "The MAXIM X-Ray Interferometry Mission Concept Study," *Liege International Astrophysical Colloquia* **36**:11 (2001).

MULTIFOIL X-RAY OPTICS

Ladislav Pina

*Faculty of Nuclear Sciences and Physical Engineering
Czech Technical University
Prague, V Holesovickach 2*

48.1 INTRODUCTION

Improved source development and the growing number of types of x-ray sources have generated an increased interest in detection, spectroscopy, and x-ray imaging. Synchrotron, free electron laser (FEL), laser plasma, tokamak plasma, capillary discharge, and Z-pinch x-ray sources need various diagnostics and offer many applications in science and technology. X-ray imaging and x-ray optical systems play a major role. The study of refraction, reflection, and diffraction in the x-ray region have led to the development of new optical elements and systems. They allow extended use of x rays in microscopy, computed tomography, nondestructive testing (NDT), and material science, as well as in the semiconductor industry, arts, security, and astronomy. Demands for extremely wide field of view (FOV) and cost-effective optical systems have led to the concept of multifoil optics (MFO). The basic principles and designs of grazing incidence multifoil optics are reviewed in this chapter. Recent developments and typical applications for multifoil optics in astronomy, extreme ultraviolet (EUV) lithography, and microscopy are described. The requirements for these applications are discussed.

48.2 GRAZING INCIDENCE OPTICS

Atomic scattering functions for x rays had been studied by Henke¹ and are tabulated in Chap. 36. The real part of complex index of refraction in the x-ray region approaches unity so the focusing effect is small in comparison with classic visible radiation optics. The imaginary part of the index causes absorption that is not negligible for most materials. Therefore, refraction optics had played no significant role until recently (see Chap. 37). The theory of reflection of x rays from a mirror surface shows the possibilities and limitations for optics in the x-ray spectral region (see Chaps. 26 and 44). Reflection optics have been more suitable for most applications. However, x rays can reflect only at grazing incidence angles and, thus, only a relatively small solid angle of collection can be used. Different solutions can be applied to enlarge the collection angle such as nested Wolter-type optics described in Chaps. 44 to 46, or multifoil optics (MFO). The microroughness of the reflecting surface plays a critical role in the x-ray spectral region. Microroughness on no larger than an atomic scale is necessary for acceptable x-ray reflection² and optic figure errors no larger than micro-

nanometer scales are necessary for high-resolution imaging. Grazing incidence x-ray optics can be used to collect the radiation in a wide wavelength range. While grazing incidence optics are commonly used in space x-ray telescopes, they can be also successfully used for laboratory imaging, as well as for collecting x rays from the laboratory sources. The critical angle is relatively large in case of the EUV radiation. It can be up to 15° for gold-coated mirrors with surface microroughness below 1 nm and radiation wavelength around 10 nm, or better for some other materials. Beside flat mirrors, the most commonly used optics are toroidal mirrors, ellipsoidal mirrors, or parabolic mirrors. Kirkpatrick-Baez,³ Wolter,⁴ Schwarzschild, and polycapillary systems are also commonly used and based on reflection. Multilayer structures can be deposited on the surfaces of mirrors to achieve the desired spectral properties or higher solid angles of collection.

48.3 MULTIFOIL LOBSTER-EYE OPTICS

The first representative of the MFO is lobster-eye optics (LE). The name of these optics arises because lobsters have eyes which operate in a similar manner.⁵ See also Chap. 49.

Lobster-Eye Geometry

X-ray optics offer an excellent opportunity to achieve very wide fields of view. A one-dimensional lobster-eye geometry was originally suggested by Schmidt,⁶ based upon flat reflectors. The device consists of a set of flat reflecting surfaces. The plane reflectors are arranged in a uniform radial pattern around the perimeter of a cylinder of radius R (Fig. 1a). X rays from a given direction are focused to a line on the surface of a cylinder of radius $R/2$. The azimuthal angle is determined directly from the centroid of the focused image. Used at a glancing angle for x rays of wavelength 1 nm and longer, this device can be used to focus a sizable portion of an intercepted beam of parallel x rays. Focusing is not perfect and the image size is finite. On the other hand, this type of focusing device

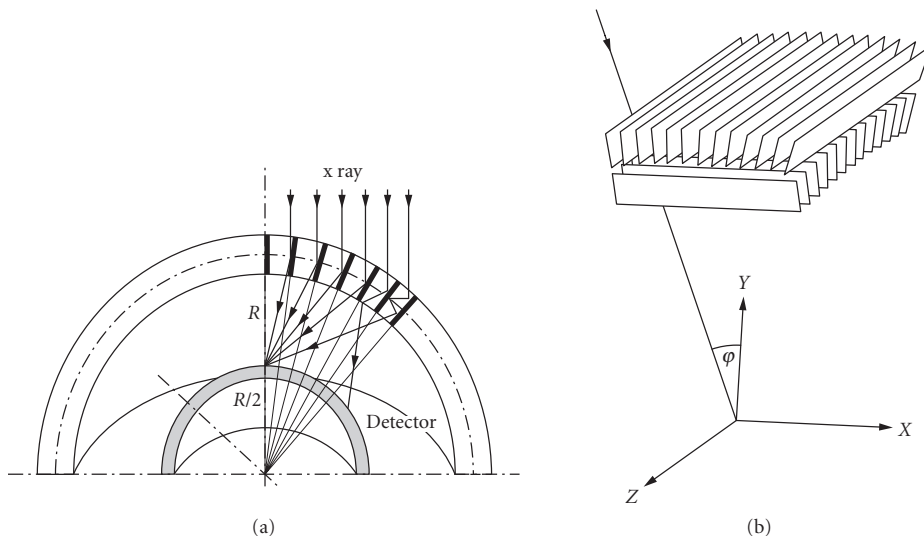


FIGURE 1 (a) Geometry of a lobster-eye x-ray optic. Parallel light incident from the top is reflected at grazing incidence on the set of flat mirrors. The image is formed on a sphere with $R/2$ radius. (b) The Schmidt lobster-eye objective in the double-focusing arrangement. (Courtesy of A. Inneman.)

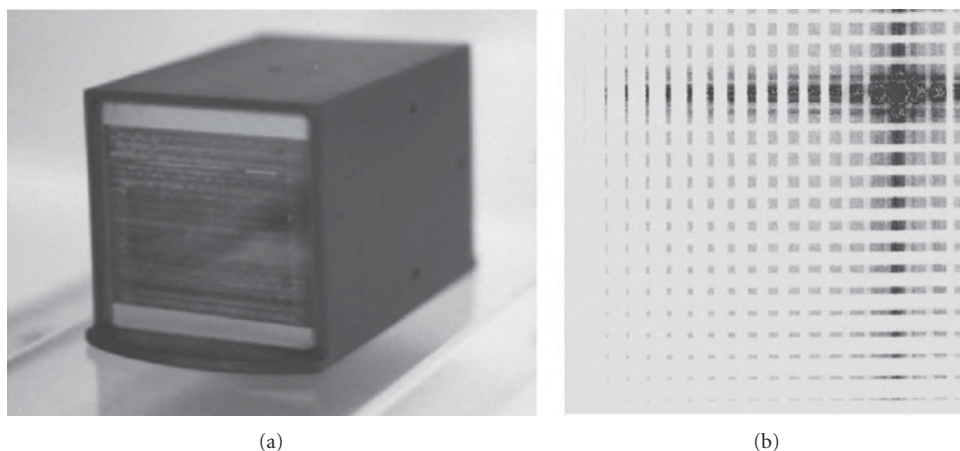


FIGURE 2 (a) The mini (24×24 mm, 0.1-mm-thick foils spaced by 0.3 mm) Schmidt LE module. (b) The mini Schmidt LE module focal spot image (8-keV x rays from microfocus x-ray tube, image area 12.3×12.3 mm). Maximum beam intensity in the focal spot is $680\times$ higher than the intensity of unfocused beam. (Courtesy of Reflex.)

offers a wide field of view, up to a maximum of 2π with coded apertures. It appears practically possible to achieve an angular resolution of the order of one-tenth of a degree or better. Two such systems in sequence, with orthogonal stacks of reflectors, form a double-focusing device (Figs. 1b and 2). Such device offers a field of view of up to 1000 square degrees at a moderate angular resolution.

48.4 MULTIFOIL KIRKPATRICK-BAEZ OPTICS

Schmidt LE systems with flat mirrors offer unsurpassed FOV, but have serious imaging and focusing limitations. Ideally, a good LE system should have mirrors as short as possible, very thin and densely packed. In laboratory applications, where the FOV is often not of the key importance, better focusing systems can be used. Kirkpatrick and Baez³ proposed combination of two mirrors in orthogonal



FIGURE 3 Kirkpatrick-Baez test Au-coated glass foils system (a) and VIS focal image (see also color insert) (b) studied for XEUS project. (Courtesy of Reflex.)

configuration (KB system) in order to achieve two-dimensional imaging. With this mirror combination they were the first to demonstrate experimentally grazing incidence x-ray imaging. Multifoil versions of KB systems (Fig. 3—example of MFO KB test module) have been studied for astronomical purposes by Van Speybroeck et al.,⁷ Weisskopf,⁸ Kast,⁹ and Gorenstein.¹⁰ A laboratory MFO system with elliptical mirrors was recently designed and tested in EUV lithography.

48.5 SUMMARY

Wide-field lobster-eye x-ray telescopes are expected to play an important role in future x-ray astrophysics missions and projects. These devices allow the study of novel science, including such important fields such as gamma ray bursts (GRB). Use of wide field x-ray optics will allow the signal-to-noise ratio to be increased compared to measurements with nonfocusing devices. The expected limiting sensitivity of LE telescopes is roughly 10^{-12} ergcm⁻²s⁻¹ for daily observation in the soft x-ray range. The scientific applications are expected to be very broad, covering numerous types and categories of variable and transient x-ray sources including x-ray binaries, AGN, blazars, Supernovae, x-ray counterparts of gamma ray bursts (including orphan afterglows), x-ray flashes, cataclysmic variables, and the like. Laboratory multifoil optical condenser allows extended use of x rays in microscopy and in EUV lithography.

48.6 REFERENCES

1. B. L. Henke, E. M. Gullikson, and 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, **54**(2):181–342 (1993). This data, along with useful computational tools, is also available on the internet at http://www-cxro.lbl.gov/optical_constants.
2. P. Beckmann and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces*, Oxford: Pergamon (1963).
3. P. Kirkpatrick and A. V. Baez, *J. Opt. Soc. Am.* **38**:776 (1948).
4. H. Wolter, *Ann. Phys.* **10**:94 (1952).
5. J. R. P. Angel, *Astrophys. J.* **233**:364–373 (1979).
6. W. H. K. Schmidt, *Nucl. Instr. Meth.* **127**:285 (1975).
7. L. P. Van Speybroeck, R. C. Chase, and T. F. Zehnpfennig, *Appl. Opt.* **10**:945 (1971).
8. M. C. Weisskopf, *Appl. Opt.* **12**:1436 (1973).
9. J. W. Kast, "Scanning Kirkpatrick-Baez X-Ray Telescope to Maximize Effective Area and Eliminate Spurious Images," *Appl. Opt.* **14**:537 (1975).
10. P. Gorenstein, *SPIE* **3444**:382 (1998).

PORE OPTICS

Marco W. Beijersbergen

*Cosine Research B.V./Cosine Science & Computing B.V.
Leiden University
Leiden, Netherlands*

49.1 INTRODUCTION

For x rays, the ratio of effective collecting area to the total reflecting surface area is small. This quantity, defined as the surface utility, for a Wolter-I optic, equals

$$s = R^2(\alpha, E) \sin(\alpha)/2 \quad (1)$$

where α is the grazing incidence angle on both mirrors in the Wolter-I optic, and $R(\alpha, E)$ the reflectivity at that grazing angle and the energy of the incident radiation. This quantity is only determined by the reflective properties of the surface, and is generally small. A typical value is 0.0015 for Au at 1 keV, which means that 1 cm² mirror only results in 0.0015 cm² effective collecting area. The challenge is therefore to create large surface area with high accuracy. For a given surface area the mass is determined by the density and thickness of the mirrors. However, thinner mirrors have lower stiffness and it is therefore more difficult to achieve a good figure and therefore angular resolution of the optic. An x-ray optic is therefore a trade-off between mass and resolution. The resolution can be improved by increasing the stiffness for a given mirror thickness. One possible solution is the use of a spacer that supports the mirrors. A more drastic implementation are pore optics, where the spacing of the spacers is of the same order as the spacing between the mirrors. There are currently three technologies that employ pore optics to implement x-ray pore optics: microchannel plate, silicon stacks, and micromachined silicon.^{1–15}

49.2 GLASS MICROPORE OPTICS

Introduction

Glass micropore optics are made from square glass fibers that are arranged in an appropriate configuration, after which the cores of the fibers are etched away such that thin mirrors remain. The resulting mirror thickness can be as low as 1 μm . With a glass density of about 2.3, this results in several orders of magnitude reduction in mass compared to 1-mm nickel shells or 0.1-mm-thick

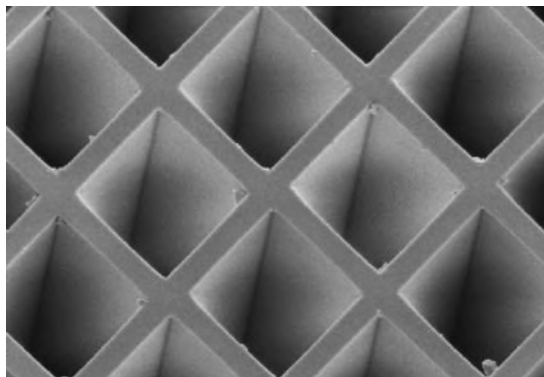


FIGURE 1 SEM image of a square-pore square-pack microchannel plate that acts as micropore optics. The pores are $20 \times 20 \mu\text{m}$, the wall thickness is $2 \mu\text{m}$.

aluminium foils. These optics can be produced with the same technology used for the manufacturing of microchannel plates, but using square instead of round glass blocks from which the fibers are drawn, as shown in Fig. 1. The process results in multifibers, as shown in Fig. 2, a square array of typically 50×50 fibers. These can be arranged in the desired geometry in a block. Plates are cut from the block and the core of the fibers are etched away, leaving square holes in a regular grid, as shown in Fig. 3.¹⁶

Glass micropore optics can be manufactured from different types of glass. In the etching process heavier elements are etched away preferentially and the resulting surface is mostly composed of SiO_2 . It has not yet been possible to coat the inside of the pores uniformly with a material with a higher atomic number.

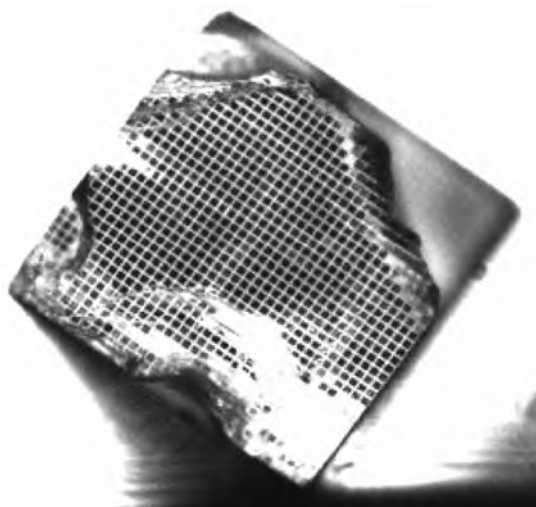


FIGURE 2 The front side of a multifiber with a size of $0.7 \times 0.7 \text{ mm}^2$. The front side of the unprocessed fiber is irregular and therefore not the entire area is properly focussed.

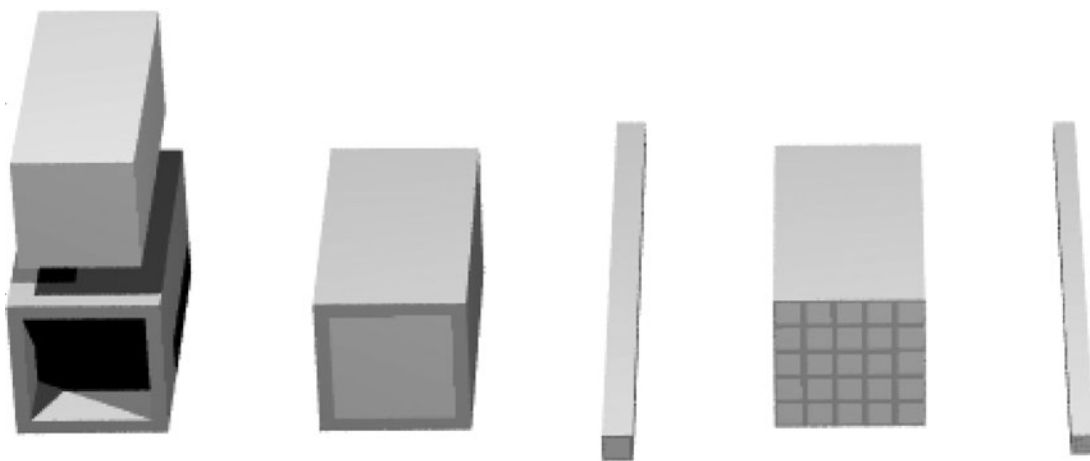


FIGURE 3 Manufacturing of a microchannel plate optics. A square core is inserted into a square cladding, and fibers are drawn. The fibers are stacked in a block, and multifibers are drawn.

In the drawing process the surface of the core glass is elongated by several orders of magnitude, thereby reducing the surface roughness significantly to about 1 to 2 nm. This makes them suitable for x-ray reflection. The typical pore size varies between 10 and 100 μm , and the wall thickness between 2 and 20 μm . This results in an open area ratio of typically 65 percent.

Lobster-Eye Optics

Multifibers can be arranged in a regular rectangular arrays, resulting in the so-called square-pore square-pack plates. If these are slumped over a spherical surface, a lobster-eye optic results,¹⁷ as shown in Fig. 4. Radiation that reflects twice inside the pore from orthogonal walls will be focused into a point. However, a large fraction of the incoming radiation will be reflected from one wall only

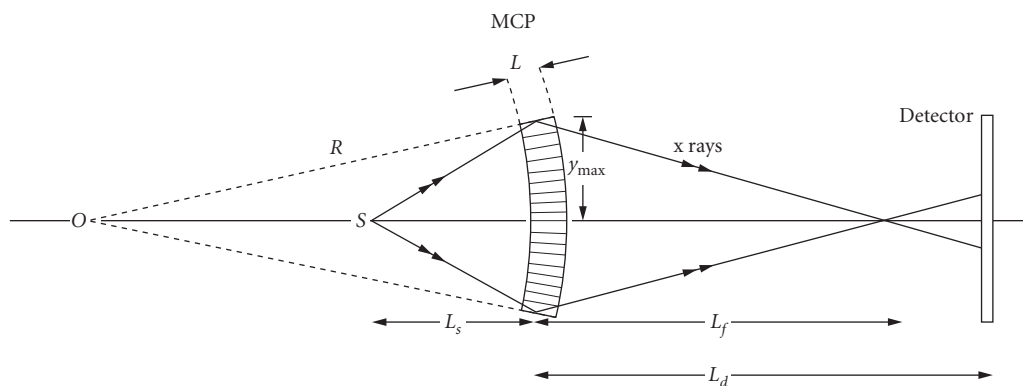


FIGURE 4 Creating a collimated beam with a slumped multichannel plate. (From Ref. 16.)

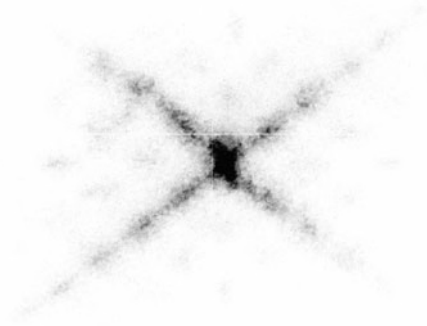


FIGURE 5 The crucifix image produced by a microchannel plate optic with square packing in a confocal arrangement. The half-energy width of the spot is about 6 arcmin. Data taken at the beam line of the Space Research Centre of Leicester University at 1 keV. (From Ref. 18.)

or go straight through the pores. The resulting image from a point source is called a crucifix image, as shown in Fig. 5.¹⁸ The crucifix point spread function (PSF) of the lobster optic results in a typical image resolution of a few arcminutes half-energy width. With a sufficiently large signal-to-noise ratio it is possible to deconvolve the image using the known PSF.

The lobster-eye optic only focuses light from an annulus on the optic. The collecting area will therefore be smaller than the total open fraction of the optic.

The field of view of a large curved lobster-eye optic is large. This optic is therefore well suited for sky survey instruments. A large-FOV instrument based on a lobster-eye optic is being developed for the International Space Station.¹⁹

The lobster-eye optic can be used for sources at finite distance, in which case a thin lens formula applies, as shown in Fig. 4,

$$\frac{1}{L_s} - \frac{1}{L_f} = \frac{2}{R_{\text{slump}}} \quad (2)$$

Glass micropore optics have been used to image x-ray fluorescence of biological and geological samples.^{20,21}

Wolter Optics

If the square fibers are stacked in concentric rings or sectors, as shown in Fig. 6, the radiation that is reflected from the azimuthal walls will result in a focus, as shown in Fig. 7. A second plate can be used to eliminate the extreme coma from a single plate, as shown in Fig. 8, resulting in proper imaging.^{22,23} In the case of a source at infinity the required slump radii are

$$f = \frac{1}{4} R_1 = \frac{3}{4} R_2 \quad (3)$$

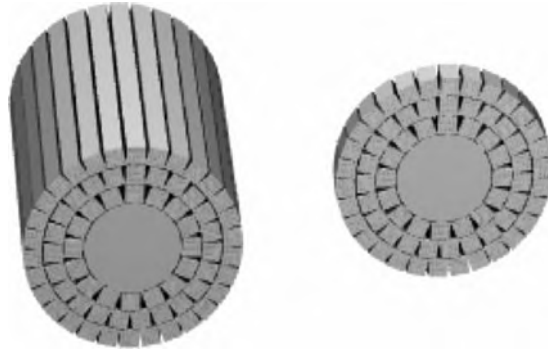


FIGURE 6 Radial microchannel plate optic are produced by stacking multifibers radially, cutting plates and etching the cores.

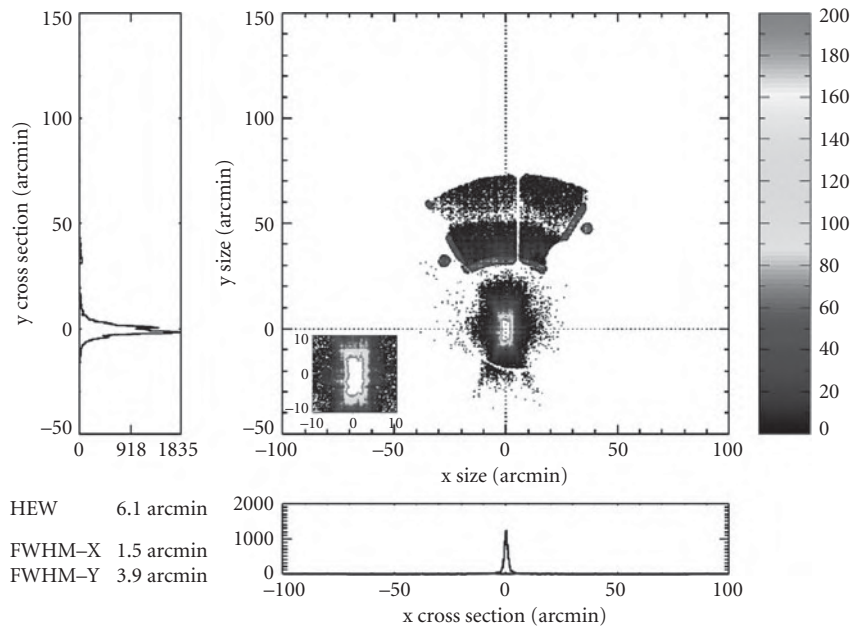


FIGURE 7 The focus of a single-plate radially packed microchannel plate optic. (See also color insert.)

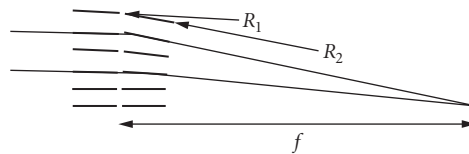


FIGURE 8 Two radially stacked plates that are slumped to different radii mimic a Wolter-I optical geometry and provide true focusing.

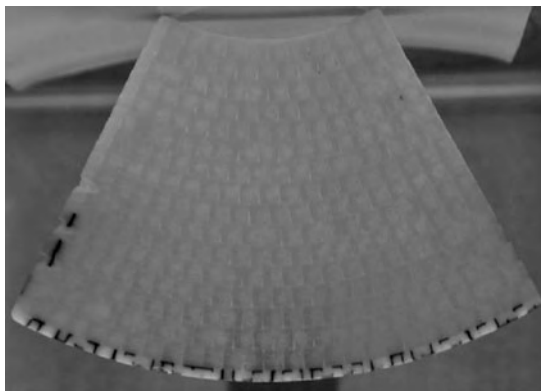


FIGURE 9 A segment of a radially stacked micropore optic. Two such plates behind each other act as a conical approximation to a Wolter-I optic. The width of the segment is 35 mm. (See also color insert.)

When the ratio of the pore size and thickness of the plate is properly chosen, the entire optic contributes to the focus of an on-axis source. The geometrical collecting area is therefore the open area ratio times the aperture.

The imaging performance of a glass micropore optic will be determined by the optical geometry, manufacturing errors that are imminent in the production of glass surfaces, as well as diffraction due to the pore structure. The focal plane is in the near field of the diffraction of the pores at typical x-ray wavelengths and Fresnel diffraction treatment is required to calculate the diffraction pattern. The actual geometry is basically a conical approximation to a Wolter-I. This will result in an image that has the size of a pore. However, in the case of concentric rings of multifibers, the multifibers will not be appropriately slumped in the azimuthal direction, and the focal spot will have the size of a multifiber. For a 1-mm multifiber size and a focal length of 1 m this corresponds to 3 arcmin. Diffraction at x-ray energies can be comparable depending on energy and pore size.²⁴

The optic can be created in off-axis segments, so that a larger collecting area in a single focus can be obtained, as shown in Fig. 9. Such an optic is being developed for an x-ray imaging spectrometer for BepiColombo, a European mission to Mercury^{25,26} and could be used for medium-resolution timing experiments.²⁷

49.3 SILICON PORE OPTICS

Silicon pore optics are produced by stacking silicon plates with ribs on the backside onto a conical mandrel, as shown in Fig. 10. Silicon wafers produced by the semiconductor industry are used. The chemomechanical polishing process that was developed for high-quality wafers results in a surface that has very good quality as x-ray mirror. Typical surface roughnesses are 0.1 nm and the flatness is better than a few nanometers over 25 mm. The plates are stacked such that the ribs make contact with the next plate and form a strong bond. This way stacks of a large number of concentric mirrors can be made. The porous structure, shown in Fig. 11, is almost as stiff as a solid silicon block. The x rays will reflect from the front side of the wafer toward the focus. A second stack is required to eliminate the severe coma caused by a single reflection. The current imaging resolution is 17 arcsec HEW over the full area and below 5 arcsec over selected regions.²⁸

Metallic and multilayer coatings can be applied to the surface inside the pore before stacking, provided that the coating is laid down in stripes, leaving uncoated silicon for the bonding.

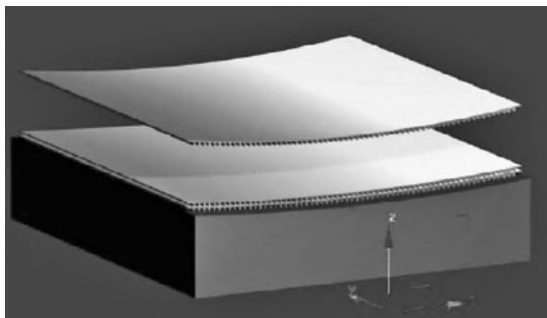


FIGURE 10 Silicon pore optics are produced by stacking silicon plates with ribs on the backside onto a concave mandrel.



FIGURE 11 A silicon pore optics produced by stacking ribbed silicon plates. The pore size is approximately $1 \times 1 \text{ mm}^2$, the width of the stack is 68 mm.

The resulting optic has a typical pore size of $1 \times 1 \text{ mm}^2$ and a typical mirror thickness of $170 \text{ }\mu\text{m}$. The mass of these optics is about 1/10 of nickel replicated optics and is comparable to foil optics. The resolution is potentially better. This technology is well suited for conical approximations to a Wolter telescope. In that geometry the resolution will be inversely proportional to the focal length for a given pore size, and therefore this optic is best suited for large focal lengths. A closer approximation to a Wolter-I optic would require bending the plates in two directions, leading to additional distortion which has to be balanced against the benefit of the improved design.

The resulting stacks can be mounted as smaller modules into a larger structure. Note that the combination of two stacks forms a lens and therefore the alignment of a so-called tandem is much less critical than the alignment of an individual stack.

This optic is the baseline technology for the XEUS/IXO mission concept and would allow a collecting area of more than 3 m^2 in space.

49.4 MICROMACHINED SILICON

Careful etching of silicon can result in surfaces that are sufficiently smooth for x-ray reflection. With this method a structure can be etched into silicon that forms a geometry that can be used for imaging.²⁹ To get good surface quality requires etching along silicon crystal planes, which limits the geometries that can be realized.

49.5 REFERENCES

1. S. W. Wilkins, A. W. Stevenson, K. A. Nugent, H. Chapman, and S. Steenstrup, "On the Concentration, Focusing and Collimation of X-Rays and Neutrons Using Microchannel Plates and Configurations of Holes," *Rev. Sci. Instrum.* **60**:1026–1036 (1989).
2. H. N. Chapman, K. A. Nugent, and S. W. Wilkins, "X-Ray Focusing Using Square Channel-Capillary Arrays," *Rev. Sci. Instrum.* **62**:1542–1561 (1991).
3. G. W. Fraser, J. E. Lees, J. F. Pearson, M. R. Sims, and K. Roxburgh, "X-Ray Focusing Using Microchannel Plates," *Proc. SPIE* **1546**: 41–52 (1991).

4. P. Kaaret and P. Geissbühler, "Lobster Eye X-Ray Optics Using Microchannels Plates," *Proc. SPIE* **1546**:82–90 (1992); P. Kaaret, P. Geissbühler, A. Chen, and E. Glavinias, "X-Ray Focusing Using Microchannel Plates," *Appl. Opt.* **31**:7339–7343 (1991).
5. G. W. Fraser, A. N. Brunton, J. E. Lees, J. F. Pearson, and W. B. Feller, "X-Ray Focusing Using Square-Pore Microchannel Plates First Observation of Cruxiform Image Structure," *NIM A* **324**:404–407 (1993).
6. G. W. Fraser, A. N. Brunton, J. E. Lees, and D. Lemberson, "Production of Quasi-Parallel X-Ray Beams Using Microchannel Plate X-Ray Lenses," *NIM A* **334**:579–588 (1993).
7. A. N. Brunton, G. W. Fraser, J. E. Lees, W. B. Feller, and P. L. White, "X-Ray Focusing with 11 μm Square Pore Microchannel Plates," in *X-Ray and Ultraviolet Sensors and Applications*, *Proc. SPIE* **2519** (1995).
8. I. C. E. Turcu, A. N. Brunton, G. W. Fraser, and J. E. Lees, "Microchannel Plate (MCP) Focusing Optics for a Repetitive Laser-Plasma Source," in *Applications of Laser Plasma Radiation II*, *Proc. SPIE* **2523** (1995).
9. A. G. Peele, K. A. Nugent, A. V. Rode, K. Gabel, M. C. Richardson, R. Strack and W. Siegmund, "X-Ray Focusing with Lobster-Eye Optics: A Comparison of Theory with Experiment," *Appl. Opt.* **35**:4420–4425 (1996).
10. A. N. Brunton, J. E. Lees, G. W. Fraser, and A. S. Tremsin, "MCP-Based X-Ray Collimators for Lithography of Semiconductor Devices," *Proc. SPIE* **2805**:212 (1996).
11. A. G. Peel and W. Zhang, "Lobster-Eye All-Sky Monitors: Comparison of One- and Two-Dimensional Designs," *Rev. Sci. Instrum.* **69**:2785–2793 (1998).
12. A. N. Brunton, A. P. Martin, G. W. Fraser, and W. B. Feller, "A Study of 8.5 μm Microchannel Plate X-Ray Optics," *NIM A* **431**:356–365 (1999).
13. M. W. Beijersbergen, M. Bavdaz, E. J. Buis, and D. H. Lumb, "Micro-Pore X-Ray Optics Developments and Application to an X-Ray Timing Mission," *Proc. SPIE Int. Soc. Opt. Eng.* **5488**:468 (2004).
14. M. Beijersbergen, S. Kraft, R. Gunther, et al., "Silicon Pore Optics: Novel Lightweight High-Resolution X-Ray Optics Developed for XEUS," *Proc. SPIE* **5488** (2004).
15. M. J. Collon, M. W. Beijersbergen, K. Wallace, M. Bavdaz, R. Fairbend, J. Séguy, E. Schyns, M. Krumrey, and M. Freyberg, "X-Ray Imaging Glass Micro-Pore Optics," *Proc. SPIE* **6688**:668812 (2007).
16. G. W. Fraser, A. N. Brunton, J. E. Lees, J. F. Pearson, R. Willingale, D. L. Emberson, W. B. Feller, M. Stedman, and J. Haycocks, "Development of Microchannel Plate MCP X-Ray Optics," *Proc. SPIE* **2011**:215–226 (1993).
17. J. R. P. Angel, "Lobster Eyes as X-Ray Telescopes," *Astrophys. J.* **233**:364–373 (1979).
18. T. J. Norton, P. F. Morrissey, J. P. Haas, L. J. Payne, J. Carbone, and R. A. Kimble, "Photon-Counting Intensified Random-Access Charge Injection Device," in S. Fineschi, B. E. Woodgate, and R. A. Kimble (eds.), *Ultraviolet and X-Ray Detection, Spectroscopy, and Polarimetry III*, *Proc. SPIE* **3765**:452 (1999).
19. G. W. Fraser, A. N. Brunton, N. P. Bannister, J. F. Pearson, M. Ward, T. J. Stevenson, D. J. Watson, et al., "LOBSTER-ISS: An Imaging X-Ray All-Sky Monitor for the International Space Station," *Proc. SPIE* **4497**:115 (2002).
20. A. P. Martin, A. N. Brunton, G. W. Fraser, A. D. Holland, A. Keay, J. Hill, N. Nelms, et al., "Imaging X-Ray Fluorescence Spectroscopy Using Microchannel Plate X-Ray Optics," *X-Ray Spectrometry* **28**:64–70 (1999).
21. G. J. Price, G. W. Fraser, J. F. Pearson, J. P. Nussey, I. B. Hutchinson, A. D. Holland, K. Turner, and D. Pullan, "Prototype Imaging X-Ray Fluorescence Spectrometer Based on Microchannel Plate Optics," *Rev. Sci. Instrum.* **75**:2314 (2004).
22. M. W. Beijersbergen, M. Bavdaz, A. J. Peacock, E. Tomaselli, G. W. Fraser, and A. N. Brunton, "Novel Micropore X-Ray Optics Produced with Microchannel Plate Technology," *Proc. SPIE Int. Soc. Opt. Eng.* **4012**:218 (2000).
23. G. J. Price, A. N. Brunton, M. W. Beijersbergen, G. W. Fraser, M. Bavdaz, and J. -P. Boutot, "X-Ray Focusing with Wolter Microchannel Plate Optics," *NIM A* **490**:276 (2002).
24. A. L. Mieremet and M. W. Beijersbergen, "Fundamental Spatial Resolution of an X-Ray Pore Optic," *Appl. Opt.* **44**:7098–7105 (2005).
25. A. Owens, M. Bavdaz, M. W. Beijersbergen, A. N. Brunton, G. W. Fraser, D. Martin, P. Nieminen, A. J. Peacock, and M. G. Pia, "HERMES: An Imaging X-Ray Fluorescence Spectrometer for the BepiColombo Mission to Mercury," *Proc. SPIE* **4506**:136–145 (2001).

26. R. Willingale, G. W. Fraser, and J. F. Pearson, "Optimization of Square Pore Optics for the X-Ray Spectrometer on Bepi-Columbo," *Proc. SPIE* **5900**:590012 (2005).
27. M. Bavdaz, D. H. Lumb, A. Peacock, and M. Beijersbergen, "MCP-Optics for X-Ray Timing, X-Ray Timing 2003: Rossi and Beyond," *AIP Conf. Proc.* **714**:443-446, held 3-5 November, 2003 in Cambridge, Ma. P. Kaaret, F. K. Lamb, and J. H. Swank, (ed.). Melville, NY: American Institute of Physics (2004).
28. M. J. Collon, R. Günther, M. Ackermann, E. J. Buis, G. Vacanti, M. W. Beijersbergen, M. Bavdaz, K. Wallace, M. Freyberg, and M. Krumrey, "Performance of Silicon Pore Optics," *Proc. SPIE* **7011**:70111E (2008).
29. Y. Ezoe, M. Koshiishi, M. Mita, K. Mitsuda, A. Hoshino, Y. Ishisaki, Z. Yang, T. Takano, and R. Maeda, "Micropore X-Ray Optics Using Anisotropic Wet Etching of (110) Silicon Wafers," *Appl. Opt.* **45**:8932-8938 (2006).

This page intentionally left blank

ADAPTIVE X-RAY OPTICS

Ali Khounsary

Argonne National Laboratory
Argonne, Illinois

50.1 INTRODUCTION

The field of adaptive optics (AO) has its origin in the second half of the twentieth century^{1,2} in connection with improving the image quality of celestial objects taken by Earth-based telescopes. Prior to this, and for some 300 years, telescopes had evolved significantly, from primitive constructs to sophisticated systems that incorporated the latest in design, optical material, polishing, and assembly.³ Further improvement in image quality required collecting more light and, more importantly, dealing with the blurring caused by atmospheric turbulence. These two requirements have led to the developments of active and adaptive optics, respectively.

To collect more light, telescopes with large primary mirrors several meters in diameter were proposed. To control optical aberrations due to mechanical and thermal effects in such large systems, *active optics*⁴ techniques were developed and implemented that allowed automatic adjustments via built-in corrective optical elements operating at fairly low temporal frequency ($\ll 1$ Hz).

To address the degrading effects of atmospheric turbulence and to obtain sharper images, *adaptive optics* was developed and implemented.⁵ Using a bright star or a beacon,⁶ wavefront distortions due to atmospheric turbulence are measured in real time, and this information is fed into a system of many high-speed (kHz range) actuators operating on a deformable mirror that adjusts its shape, as shown in Fig. 1, to correct for the atmospheric distortion to produce sharper images. A deformable mirror of this type is typically about 10 to 20 cm in diameter, located behind the telescope focus, and its axial range of deformation is typically on the order of a few microns.

Developments in adaptive optics in the visible or IR region of the spectrum have continued with military, communication, astronomy, energy, and medical applications (see Chap. 5). These efforts are now being extended to the x-ray region, notably in x-ray astronomy and synchrotron-based x-ray systems.

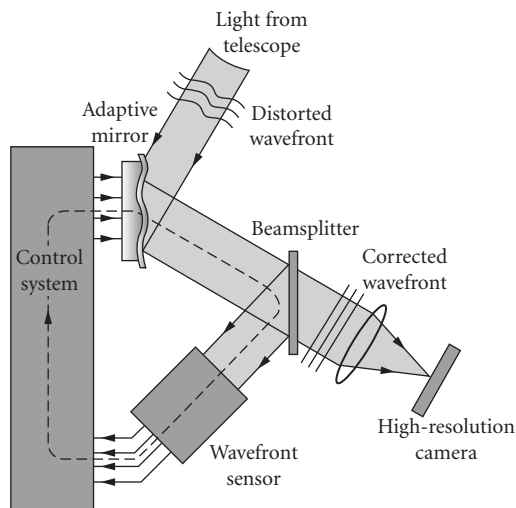


FIGURE 1 Light from the celestial object of interest and a nearby “guide star” passes through the telescope optics and propagates within the adaptive system shown. The light from the star is continuously measured by a high-speed wavefront sensor and analyzed to determine atmospheric distortion. This information is used to change the shape of the deformable mirror in order to cancel out the distortion due to atmosphere, which allows the high-resolution camera to obtain sharper images of the stars and celestial objects. (Courtesy of Claire Max, Center for Adaptive Optics, University of California at Santa Cruz, California.)

50.2 ADAPTIVE OPTICS IN X-RAY ASTRONOMY

Unlike optical astronomy, space-based hard x-ray astronomy relies on grazing incident optics to collect an image. This severely restricts the amount of light that can be collected because the available optical surface is limited. To address this problem, light-weight optics composed of a large number of nested confocal mirrors with very thin walls have been developed (see Chap. 47). Mirrors are individually produced by electroforming on a suitably shaped mandrel and then releasing them.^{7,8} For high-angular-resolution hard x-ray astronomy, telescopes with rigid optics are presently preferred, as they maintain their figure during the rigor of launch and beyond. The light collection area is severely restricted, however, due to weight considerations. Research is continuing in the use of adaptive optics techniques for in situ control and figure correction.^{9,10} For soft x-ray telescopes operating near normal-incident angles, on the other hand, adaptive optics is being developed to correct figure errors of the primary mirror in an effort to approach diffraction limit performance.¹¹

50.3 ACTIVE AND ADAPTIVE OPTICS FOR SYNCHROTRON- AND LAB-BASED X-RAY SOURCES

The terms active optics and adaptive optics have distinct meanings in the *optical telescope community*, denoting low-speed and high-speed correction systems, respectively. Elsewhere, this distinction is not

always observed. For example, within the *x-ray optics community*, the terms active, adaptive, bendable, and deformable are used interchangeably, yet only a few systems are truly adaptive in the sense of high-speed automatic sensing, feedback, and correction. In what follows, the term adaptive is used to denote optics that during operation undergo complex deformation or rigid body motions that require (manual or automatic) monitoring and active feedback, or optics that are of the self-correcting type, which have built-in self-adjustment mechanisms by design. Other systems are termed active optics.

In recent years, several groups of researchers have begun exploring adaptive optics technologies for application to synchrotron and lab-based x-ray optics. This trend is expected to accelerate because of its impact on beamline performance and throughput. More significantly, because of the relative ease with which adaptive technologies can be implemented at the so-called mezzo (sub-mm) level (using MEMS, micro-sensors, and related processing and technologies), an entirely new generation of x-ray optical systems could result.

The ultimate goal would be to develop a new class of enabling adaptive optical systems that can, in real time, sense and compensate for undesirable changes (a) in the optics itself, (e.g., due to beam heat load), (b) in the laboratory environment (e.g., thermal, mechanical), and (c) possibly in the x-ray source (temporal and spatial). As an example, it would be possible to focus and keep a nanofocused x-ray beam on a small sample for an extended period of time, a goal that might best be achieved not by expansive control of a long beamline but by active/adaptive control of the optics.

Major x-ray optical systems currently in use on synchrotron x-ray beamlines or in lab-based x-ray systems include monochromators, mirrors, multilayer-coated mirrors, zone plates, polycapillaries, and compound refractive lenses. Many are active optics, as they have provisions for angular and spatial movements and adjustments. They generally lack automatic sensing, feedback, and control because, unlike optical astronomy, which requires an automatic and relatively high-speed feedback system on the order of kHz, many laboratory and synchrotron x-ray optics rely on painstakingly slow operator feedback and adjustments. In many instances “sensing” is difficult, labor intensive, or complex, and not amenable to effective automation. Thus, of the three components of an adaptive optical system (sensor, optics, and control), the sensor remains most problematic. Currently, for example, when bending a mirror to focus an x-ray beam, intensive operator intervention, in the form of repeated knife-edge scans and bending adjustments, is necessary to arrive at the optimal mirror shape for finest focus. With an adaptive x-ray system, the mirror surface profile could be automatically measured (optically or directly by x rays) and the bending adjusted to arrive at the best surface profile.

Thus, incorporating adaptive techniques in synchrotron- and lab-based x-ray optics optical systems is highly desirable.

Progress in three cases is described in this chapter:

- Bendable optics: Controlled elastic bending of optical substrates into versatile shapes necessary to focus, collimate, or collect x rays.
- Figure correction: Controlled correction of long spatial wavelength figure errors (due to manufacturing, thermal load, mounting, gravity, etc.).
- Novel optics: Development of new types of optical systems, such as x-ray telescopes for both tabletop and synchrotron x-ray sources, where adaptive techniques form the basis of the design rather than an enhancing feature.

Bendable Optics

Because of the widespread need to collect and focus x rays on one hand, and the availability of higher-quality flat substrates on the other, a number of bendable optical systems (composed of bent mirrors, monochromators, or both) have been developed.^{12–17} An additional impetus is the difficulty and expense in manufacturing high-quality mirrors with elliptical or higher-order polynomial profiles.^{18–21} A variety of bending tools and techniques using electrical, mechanical, and thermal means have been developed, although most are mechanically bent by applying moments at or near the ends of the mirrors, as shown in Fig. 2.

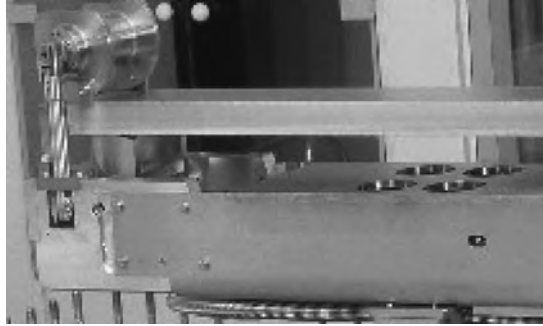


FIGURE 2 A mirror bender designed by IRELEC for ESRF. The system applies two controlled bending moments, by means of two electrical actuators, at each end of a mirror. The moments are controlled independently to be able to bend the optics into elliptical or cylindrical profiles. (*Courtesy of IRELEC, France.*)

The design principle underlying bendable optics is simple elastic beam theory. It is well known that a flat beam of uniform cross section can be bent to the arc of a circle of radius R given by $(1/R) = (M/EI)$, where M is the applied bending moment at its ends, E is the modulus of elasticity, and I is the beam's moment of inertia.²² This one-dimensional treatment is applicable to long mirror substrates where the length is much larger than other dimensions. Application of two equal moments at the ends produces a circular arc; two unequal moments produce an ellipse. More generally, this equation can be written as a function of location x along the length of the mirror and measured from an arbitrary origin as

$$\frac{1}{R(x)} = \frac{M(x)}{EI(x)}$$

where $R(x)$ is the *local* radius of curvature. Thus, a homogeneous elastic substrate may be bent into a range of longitudinal profiles by the appropriate application of moments along its length.^{14,18,23} Alternatively, because of its dependence on moment of inertia, even with the application of a constant moment M , the curvature $R(x)$ can be made to vary along the mirror length by using a substrate with a tailored nonuniform cross section (width or thickness). Thus, within limits, a flat mirror can be bent into one with an arbitrarily varying longitudinal surface profile, and indeed many such optics have been made and are in operation.^{20, 24–26}

With the few exceptions, these bendable systems do not employ adaptive technology, although most of the necessary apparatus, except for an automatic sensing system, are in place. An automatic sensing and feedback system could provide real-time enhanced dynamic bending options and automatic adjustments in response to environmental changes, and the capacity to correct long wavelength tangential slope errors in systems with multiple actuators.

Although bendable optical systems have successfully produced focused x-ray beams down to submicron size, it should be noted that they have some practical limitations.^{18,27,28} First, they are sensitive to the bender adjustment, especially when adaptive techniques are not used. Second, they are bulky and harder to cool in comparison to the prefigured alternative. Lastly, in practice, bent mirrors have not been able to match the precise figures possible with prefigured rigid optics to produce nanometer-size focal spots.^{29,30}

Adaptive X-Ray Optics

For the reasons described earlier, only a few truly adaptive x-ray optical systems that use actuators and robust sensing techniques, such as Shack-Hartmann wavefront sensing, have been developed to date. A few of the main systems are described next.

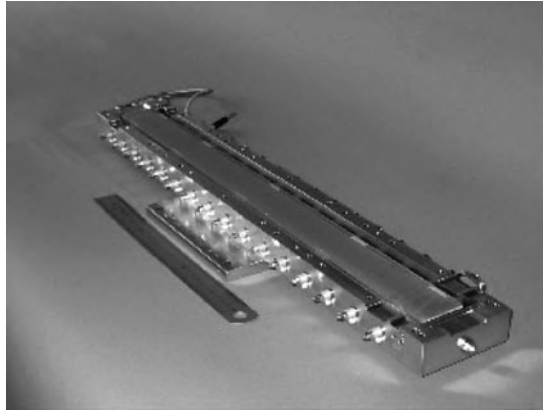


FIGURE 3 A bimorph mirror, 600 mm in length, composed of four bimorph plates each with four electrodes, installed at the GM/CA-CAT beamline at the Advanced Photon Source in the United States.³⁶ (Courtesy of SESO, France.)

Bimorph Mirrors One of the promising adaptive optical systems developed for x-ray applications is the bimorph mirror system. Piezoelectric bimorph mirrors³¹ were first proposed for adaptive control of laser³² and astronomical mirrors,³³ and prototype systems were built.^{34,35} A bimorph mirror consists of two faceplates with a sheet of active elements sandwiched symmetrically between them. The active elements are made of two piezoelectric plates glued together with their polarization vectors in the same direction perpendicular to the plate face. Each plate is coated with a thin metallic electrode at the glued surface. When a voltage is applied at the interface electrodes, one plate contracts and the other expands, bending the mirror assembly spherically. To change the radius, the applied voltage is changed. To change the mirror shape locally, a number of smaller embedded piezoelectric plates can be used to provide versatile shape control, as shown in Fig. 3.

In the early 1990s, the European Synchrotron Radiation Facility (ESRF) began exploring active and adaptive technologies in connection with thermally induced deformation in cooled x-ray mirrors for high-heat-load x-ray beamlines.^{37,38} A system composed of multiple discrete piezoelectric actuators and a Shack-Hartmann wavefront sensor³⁹ was built, as shown in Fig. 4.⁴⁰ It was recognized, however, that thermally induced deformations are rather smooth and slow varying along the mirror length, and their correction requires only a gentle bending for which a bimorph mirror with one or a small number of electrodes is most appropriate.⁴¹ For focusing x rays that require an elliptically bent mirror, a bimorph composed of a few electrodes is sufficient;⁴² short bimorph mirrors were fabricated and bent to an elliptical shape in a Kirkpatrick-Baez geometry⁴³ (see Chap. 44) to focus x-ray beams in two directions.⁴⁴ The design has been evolving over time and improvements are gradually being made in fabrication, control, feedback, and performance.^{45–48} Presently about 150 of these bimorph mirrors are in use on synchrotron x-ray beams worldwide, a number that is likely to grow as they are recognized as an economical approach to adaptive optics for synchrotrons and x-ray astronomy.⁴⁹

It is important to note that one of the severe limitations on the use of bimorph mirrors at synchrotron facilities is that the mirror temperature must remain below 60°C, and thus this technology is only suitable for monochromatic beams or other low thermal load beams.

Adaptive Optics for Thermal and Environmental Control and Correction As noted, one potential use of adaptive optics in synchrotron and perhaps lab-based x-ray systems is for the control and correction of surface deformations due to beam thermal load, gravity, or environmental factors. Thermally induced deformations are particularly important, and an adaptive optics system that allows cooling to take place while providing dynamic correction is highly desirable. For reference,

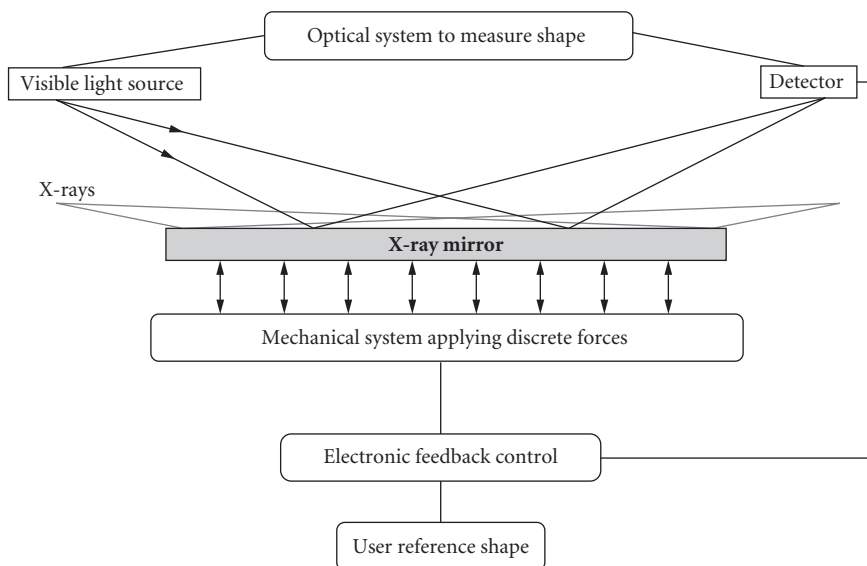


FIGURE 4 Schematic of a cooled adaptive mirror system for synchrotron x-ray applications. A Shack-Hartmann wavefront sensor is used to measure the thermally induced deformation of the optical surface and provide feedback to 22 (in two rows of 11) piezoelectric actuators that flex the 1-m-long mirror. Each actuator has a maximum stroke of 40 micrometers.⁴⁰ (Courtesy of Jean Susini, ESRF Grenoble, France.)

it should be noted that the third-generation synchrotron x-ray facilities that have come online since the early 1990s generate very small but powerful x-ray beams with low divergence⁵⁰ but with thermal loads from a few hundred to thousands of watts. Optical substrates subjected to high-heat-load x-ray beams typically deform into a convex shape. For reflective optics, this results in increased beam divergence; for refractive optics, it results in loss of x-ray beam intensity. These optical elements, therefore, must be cooled to remove the heat, and their adverse thermal deformations corrected, internally or externally. While most corrections require external means, under some circumstances it is possible to build into the design of the system some untraditional adaptive features such that the sensing, feedback, and correction are performed by the optics itself.⁵¹ In one design, two properly devised cooling blocks are optimally positioned on the long sides of a mirror close to the reflecting surface. When a high-heat-load x-ray beam strikes the mirror at grazing angles, the heat is deposited along a narrow longitudinal footprint on the optical surface. The mirror immediately deforms into a convex shape, but within several minutes a reverse thermal moment is automatically generated within the mirror body that reverses the initial curvature and largely flattens the optical surface.^{51,52}

In general, however, thermal distortions cannot be mitigated by internal reversal, and traditional adaptive techniques must be implemented to correct the deformations. One such system was developed at the ESRF.⁴⁰ In this system, shown in Fig. 4, a Shack-Hartmann sensor monitors the mirror surface to determine thermally induced deformations and provides feedback to a system of 22 piezoelectric actuators to correct the figure of the thermally deformed mirror.

X-Ray Monochromators As noted, x-ray monochromators (see Chaps. 30 and 39) installed on high-heat-load x-ray synchrotron beamlines can undergo thermal deformation that effectively reduces the intensity of the diffracted beam. Unlike mirrors, however, monochromators intercept the beam typically at large angles and absorb most of the white incident beam, and thus they are subjected to a substantially higher heat flux. They must therefore be efficiently cooled and their thermal

deformation managed. This thermal management issue is likely to become more acute as a result of the upgrades planned at synchrotron facilities. Presently most high-heat-load silicon monochromators, by necessity, employ cryogenic cooling around 125 K to exploit silicon's high thermal conductivity and nearly zero thermal expansion coefficients around this temperature. Thermal deformation is insignificant, requiring little or no corrections. However, the cooling systems are somewhat complex and water cooling is preferred, if possible, especially for lower-heat-load beamlines. A number of "adaptive" techniques to overcome undesirable thermal distortions were explored earlier. For example, in one design, a silicon monochromator with a thin diffracting surface was water-cooled with water jets on its back while pressurized helium was used on the front side of the diffracting surface to flatten its convex shape and increase diffracted intensity.⁵³ In another design, a silicon monochromator, consisting of a thin crystal faceplate with small cooling channels in it, was bonded to and supported by beryllium through which some 20 actuators were used to flatten the faceplate.⁵⁴

While many investigators have examined adaptive techniques for the control of thermal effects, some have examined active control of an optical system thermally. For example, to increase the throughput or resolution in a monochromator subjected to a divergent incident beam, heat can be applied to generate a thermal gradient across the face of the diffracting crystal such that the d-spacing variation matches incident angles of the divergent beam, i.e., $\lambda = 2d \sin \theta_B$ is a constant. This is useful when the divergences of a beam appreciably exceed the acceptance angles of the monochromator used. Gains of up to two orders of magnitude in intensity, relative to standard flat isothermal crystal systems, are expected.⁵⁵ A number of other schemes to use thermal gradients to actively control optics have also been suggested or implemented with varying degrees of success.^{56,57}

Novel Optics

Microchannel Plates (MCPs) Microchannel plates, also known as "micro pore optics" (MPOs), (see Chap. 49) were originally developed to provide high-resolution images in low-light conditions (e.g., night vision goggles) using image intensifiers. More recently, they have been used in x-ray detectors and have also been investigated for x-ray focusing⁵⁸ in a manner similar to the lobster eye vision proposed earlier.⁵⁹ A typical MCP is a planar glass, on the order of 1 mm thick, composed of thousands of small channels. To focus x rays using MCPs, these channels must be curved. This can be accomplished by spherically slumped or adaptively controlled bending of the plate so that incident rays on the channel walls of the MCP are reflected onto a screen producing an image of the object. Developments in this field continue through various design improvements, smoother walls, and more accurate channel alignment. Lobster-ISS (International Space Station) is one such system that is expected to be deployed in 2010 aboard the ISS.^{60,61} While lobster x-ray optics was originally intended for astronomical applications to view a broad swath of the sky, these compact x-ray focusing devices are being explored for lab-based sources and possible synchrotron use.⁶²

Adaptive Microstructured Optical Arrays (MOAs) Microstructured optical arrays are a new class of x-ray focusing optical systems under development that is conceptually similar to polycapillary and microchannel plane optics but differs in that x-rays are guided by a single (rather than multiple) grazing angle reflection from each MOA.^{63,64} In a sense, MOAs are a discretized version of a polycapillary optics, allowing considerable flexibility in the design of the channels and layout of the arrays. Figure 5 shows a sketch of a one-dimensional MOA system composed of two arrays focusing x rays from S to S'. The second array shown is bent. Bending one or both arrays provides a variable focal length. Adaptive control, for example, of the piezoelectric material coated on the support structure of the arrays allows bending, changing the dimensions of the channels, or correcting optical aberrations. Preliminary MOAs are being made by the deep silicon etching process.⁶⁵ MOAs hold considerable potential for x-ray optics in general and adaptive optics in particular. For example, it is estimated that an optimally designed MOA could provide two orders of magnitude more focused flux than the corresponding zone-plate, primarily due to the larger aperture afforded by an MOA.⁶⁵

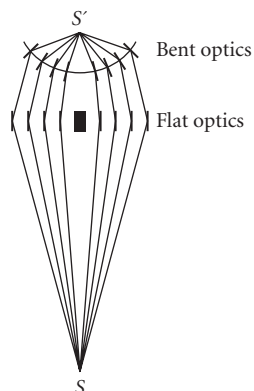


FIGURE 5 Schematic of a microstructured optical array system used to focus x rays from a source S to focus S' . This one-dimensional system is composed of two arrays, the second of which (the top one) is bent. (Courtesy of Alan Michette, Kings College, U.K.)

50.4 CONCLUSIONS

Adaptive x-ray optics, as a field, is in its early stages of development. Most x-ray optical systems in use today are passive because stability is valued more than versatility.

However, development of robust adaptive systems is likely to provide the means to achieve enhanced throughput, resolution, stability, and versatility. Concerted effort is needed to develop cost-effective and reliable systems that provide real improvement over the existing optical systems. Actuation mechanisms (thermal, piezoelectric, magnetic, electrostatic, motor driven) and sensing techniques suitable in a variety of x-ray environments must be developed. Development of novel x-optics based on adaptive techniques would enhance throughput and versatility of both synchrotron- and lab-based systems.

50.5 REFERENCES

1. H. W. Babcock, "The Possibility of Compensating Astronomical Seeing," *Pub. Astr. Soc. Pac.* **65**:229–236 (1953).
2. V. P. Linnik, "On the Possibility of Reducing the Influence of Atmospheric Seeing on the Image Quality of Stars," (1957) article translated and reprinted in F. Merkle, ed., *ICO-16 Satellite Conference on Active and Adaptive Optics*, Vol. 48 of *ESO Conference and Workshop Proc.*, 535–537, ESO, Garching, 1993.
3. R. N. Wilson, *Reflecting Telescope Optics I*, 2nd ed., Springer, Berlin, 2007.
4. R. K. Tyson, (ed.), *Adaptive Optics Engineering Handbook*, Marcel Dekker, Inc., New York, 2000.
5. J. M. Beckers, "Adaptive Optics for Astronomy: Principles, Performance, and Applications," *Ann. Rev. Astron. Astrophys.* **31**:13–62 (1993).
6. N. Hubin and L. Noethe, "Active Optics, Adaptive Optics, and Laser Guide Stars," *Science* **262**:1390–1394 (1993).
7. O. Citterio, G. Bonelli, G. Conti, E. Mattaini, and B. Sacco, "High Throughput Optics for X-Ray Astronomy," *IL Nuovo Cimento* **13**(2):375–389 (1990).

8. R. Shibata, Y. Ogasaka, K. Tamura, A. Furuzawa, Y. Haba, Y. Tawara, H. Kunieda, et al., "Hard X-Ray Mirrors by Multiplayer Replication: Development and Applications," *Proc. 8th Int. Conf. X-Ray Microscopy*, pp. 159–161, 2006.
9. M. Ulmer, Private Communication, 2008.
10. O. Cugat, S. Basrour, C. Divoux, P. Mounaix, and G. Reyne, "Deformable Magnetic Mirror for Adaptive Optics," *Sensors and Actuators A*, **89**:1–9 (2001).
11. S. Kitamoto, H. Takano, H. Saitoh, N. Yamamoto, T. Kohmura, K. Suga, H. Sekiguchi, Y. Ohkawa, J. Kanai, and S. Chiba, "Development of an Ultra-High Precision X-Ray Telescope, Stellar-Mass, Intermediate-Mass, and Supermassive Black Holes," S. Mineshige and K. Makishima (eds.) *Prog. of Theor. Phys. Suppl.* **155**:363–364 (2004).
12. W. Ehrenberg, "X-Ray Optics," *Nature* **160**:4062 (1947).
13. W. Ehrenberg, "X-Ray Optics: The Production of Converging Beams by Total Reflection," *J. Opt. Soc. Am.* **39**:741–746 (1949).
14. J. A. Howell and P. Horowitz, "Ellipsoidal and Bent Cylindrical Condensing Mirrors for Synchrotron Radiation," *Nucl. Instrum. Meth.* **125**:225–230 (1975).
15. *Proc. Workshop on X-Ray Instrumentation for Synchrotron Radiation*, SSRL Report No. 78/04, VII/36–37, 1978.
16. J. B. Leigh and G. Rosenbaum, "Synchrotron X-Ray Sources: A New Tool in Biological Structural and Kinetic Analysis," *Annu. Rev. Biophys. Bioeng.* **5**:239–270 (1976).
17. G. E. Ice and C. J. Sparks, "Conical Geometry for Sagittal Focusing as Applied to X-Rays from Synchrotrons," *Oak Ridge National Laboratory Report*, ORNL/TM-12327, 1993.
18. S. J. Chen, C. K. Kuan, S. Y. Perng, D. J. Wang, H. C. Ho, T. C. Tseng, Y. C. Lo, and C. T. Chen, "New Focusing Mirror System for Synchrotron Radiation Infrared Beamlines," *Opt. Eng.* **43**:3077–3082 (2004).
19. H. A. Padmore, M. R. Howells, S. Irick, T. Renner, R. Sandler, and Y.-M. Koo, "New Schemes for Producing High-Accuracy Elliptical X-Ray Mirrors by Elastic Bending," *Proc. SPIE* **2856**:145–156 (1996).
20. P. Eng, M. Newville, M. L. Rivers, and S. R. Sutton, "Dynamically Figured Kirkpatrick Baez X-Ray Microfocusing Optics," *Proc. SPIE* **3449**:145–156 (1998).
21. B. W. Adams and K. Attenkofer, "An Active-Optic X-Ray Fluorescence Analyzer with High Energy Resolution, Large Solid Angle Coverage, and a Large Tuning Range," *Rev. Sci. Instrum.* **79**:023–102 (2008).
22. W. C. Young, *Roark's Formula for Stress and Strain*, 6th ed., McGraw-Hill, New York, 1989.
23. M. R. Howells and D. Lunt, "Design for an Adjustable-Curvature, High-Power, X-Ray Mirror Based on Elastic Bending," *Opt. Eng.* **32**:1981–1989 (1993).
24. D. Turner and J. M. Bennett, "An Elliptical Reflector Formed by Bending a Cantilever," Imperial College (1971).
25. J. H. Underwood, "Generation of a Parallel X-Ray Beam and Its Use in Testing Collimators," *Space Sci. Instrum.* **3**:259–270 (1977).
26. G. E. Ice and C. J. Sparks, "A Simple Cantilevered Mirror for Focusing Synchrotron Radiation," *Nucl. Instrum. Meth. A* **266**:394–398 (1988).
27. D. Lunt, J. Bender, D. W. Ewing, and W. R. McKinney, "XUV Synchrotron Optical Components for the Advanced Light Source," *Proc. SPIE* **1740**:161–172 (1993).
28. M. R. Howells, D. Camble, R. M. Duarte, S. Irick, A. A. MacDowell, H. A. Padmore, T. R. Renner, et al., "Theory and Practice of Elliptically Bent X-Ray Mirrors," *Opt. Eng.* **39**:2748–2762 (2000).
29. W. Liu, G. E. Ice, J. Z. Tischler, A. Khounsary, C. Liu, L. Assoufid, and A. T. Macrander, "Short Focal Length Kirkpatrick-Baez Mirrors for a Hard X-Ray Nanoprobe," *Rev. Sci. Instrum.* **76**:113701 (2005).
30. H. Mimura, H. Yumoto, S. Matsuyama, Y. Sano, K. Yamamura, Y. Mori, and M. Yabashi, "Efficient Focusing of Hard X Rays to 25 nm by a Total Reflection Mirror," *Appl. Phys. Lett.* **90**:051903 (2007).
31. J. H. McElroy, P. E. Thompson, H. E. Walker, E. H. Johnson, D. J. Radecki, and R. S. Reynolds, "Laser Tuners Using Circular Piezoelectric Benders," *Appl. Opt.* **14**:1297–1302 (1972).
32. P. V. Nikolaev and A. V. Smirnov, "Model Wavefront Correctors," *Sov. J. Opt. Technol.* **54**:693–700 (1987).
33. F. Roddier, "A New Concept in Adaptive Optics: Curvature Sensing and Compensation," *Appl. Opt.* **27**:1223–1225 (1998).
34. F. Roddier and F. Forbes, "Curvature Sensing and Compensation," in *Adaptive Optics in Solar Observation*, O. Envold (ed.), p. 176, 1987.
35. A. V. Ikramov, S. V. Ramanov, I. M. Roshchupkin, A. G. Safronov, and A. O. Sulimov, "Bimorph Adaptive Mirror," *Sov. J. Quantum Electron.* **22**:163–166 (1992).

36. R. F. Fischetti, D. W. Yoder, S. Xu, S. Stepanov, O. Makarov, R. Benn, S. Corcoran, et al., "Optical Performance of the GM/CA-CAT Canted Undulator Beamlines for Protein Crystallography," *Synch. Rad. Instrum* **879**:754–757 (2007).
37. J. Susini, G. Marot, L. Zhang, R. Ravelet, and P. Jagourel, "Conceptual Design of an Adaptive X-Ray Mirror Prototype for the ESRF," *Rev. Sci. Instrum.* **63**:489–492 (2002).
38. J. P. Gaffard, R. Ravelet, and C. Boyer, "X-Ray Adaptive Mirror: Principle and State of the Art," *Proc. SPIE* **1739**:474–488 (1992).
39. J. Schwiegerling and D. R. Neal, "Historical Development of the Shack-Hartmann Wavefront Sensor," in *Robert Shannon and Roland Shack*, J. E. Harvey and R. B. Hooker (eds.), *SPIE* 132–139, 2005.
40. J. Susini, R. Baker, M. Krumrey, W. Schwegle, and A. Kvik, "Adaptive X-Ray Mirrors Prototype: First Results," *Rev. Sci. Instrum.* **66**:2048–2052 (1995).
41. J. P. Gaffard and P. Jagourel, "An Active Bimorph Structure for X-Ray Gratings and Mirrors," *Proc. SPIE* **2856**:197–206 (1997).
42. J. Susini, D. Laberge, and L. Zhang, "Compact Active/Adaptive X-Ray Mirror: Bimorph Piezoelectric Flexible Mirror," *Rev. Sci. Instrum.* **66**:2229–2231 (1995).
43. P. Kirkpatrick and A. V. Baez, "Formation of Optical Images by X-Rays," *J. Opt. Soc. Am.* **38**:766–774 (1948).
44. J. Susini, D. Laberge, and O. Hignette, "R & D Program on Bimorph Mirrors at the ESRF," *Proc. SPIE* **2856**:130–144 (1997).
45. R. Signorato, "R & D Program on Multi-Segmented Piezoelectric Bimorph Mirrors at the ESRF: Status Report," *Proc. SPIE* **3447**:20–31 (1998).
46. R. Signorato and T. Ishikawa, "R & D on Third Generation Multi-Segmented Piezoelectric Bimorph Mirror Substrates at SPring8," *Nucl. Instrum. Meth. A* **467–468**:271–274 (2001).
47. T. C. Tseng, S. -J. Chen, Z. C. Yen, S. Y. Perng, D. J. Wang, C. K. Kuan, J. R. Chen, and C. T. Chen, "Design and Fabrication of Aspherical Bimorph PZT Optics," *Nucl. Instrum. Meth. A* **467–468**:294–297 (2001).
48. R. Signorato, D. Haesermann, M. Somayazulu, and J. F. Carre, "Performance of an Adaptive μ -Focusing Kirkpatrick-Baez System for High Pressure Studies at the Advanced Photon Source," *Proc. SPIE* **5193**:112–123 (2004).
49. P. Doel, C. Atkins, S. Thompson, D. Brooks, J. Yao, C. Feldman, R. Willingale, T. Button, D. Zhang, and A. James, "Development of Piezoelectric Actuators for Active X-Ray Optics," *Proc. SPIE* **6705**:7050M (2007).
50. E. E. Koch, *Handbook on Synchrotron Radiation*, Vol. 1a, Elsevier Science Publishers, B. V. North Hollands, The Netherlands, 1983.
51. A. Khounsary and W. Yun, "On Optimal Contact Cooling of High-Heat-Load X-Ray Mirrors," *Rev. Sci. Instrum.* **67**:3354 (1996).
52. Y. Li, A. M. Khounsary, and S. Nair, "How and Why Side Cooling of High-Heat-Load Optics Works," *Proc. SPIE* **5533**:157 (2004).
53. L. E. Berman and M. Hart, "Adaptive Crystal for High Power Synchrotron Sources," *Nucl. Instrum. Meth. A* **302**:558–562 (1991).
54. D. Dezoret, R. Marmoret, A. K. Freund, A. Kvik, and R. Ravelet, "Design of an Adaptive Cooled First Crystal for an X-Ray Monochromator," *Proc. SPIE* **2279**:544–549 (1994).
55. G. S. Knapp and R. K. Smither, "High Resolution Monochromator Systems using Thermal Gradient Induced Variable Bragg Spacing," *Nucl. Instrum. Meth. A* **246**(1–3):365–367 (1986).
56. R. K. Smither, "Variable Focus Crystal Diffraction Lens," *Rev. Sci. Instrum.* **60**:2044–2047 (1990).
57. M. Popovici and W. B. Yelon, "On the Optical Design of Heat-Loaded Double Crystal Monochromators for Synchrotron Radiation," *J. Appl. Cryst.* **25**:471–476 (1992).
58. G. J. Price, A. N. Brunton, M. W. Beijersbergen, G. W. Fraser, M. Bavdaz, J. -P. Boutot, R. Fairbend, S. -O. Flyck, A. Peacock, and E. Tomaselli, "X-Ray Focusing with Wolter Microchannel Plate Optics," *Nucl. Instrum. Meth. A* **490**:276–289 (2002).
59. J. R. P. Angel, "Lobster Eyes as X-Ray Telescopes," *Astron. J.* **233**:364 (1979).
60. J. F. Pearson, N. P. Bannister, and G. W. Fraser, "Lobster-ISS: All-Sky X-Ray Imaging from the International Space Station," *Astron. Nachr./AN* **324**(1–2):168 (2003).
61. T. Roberts, "Future X-Ray Astronomy Missions," UK-XRA 2005, Leicester University, 2005.

-
62. J. Mutz, O. Bonnet, R. Fairbend, E. Schyns, and J. Seguy, "Micro-Pore Optics: From Planetary X-Rays to Industrial Market," *Proc. SPIE* **6479**:64790F.1–8 (2007).
 63. P. D. Prewett and A. G. Michette, "MOXI: A Novel Microfabricated Zoom Lens for X-Ray Imaging," *Proc. SPIE* **4145**:180–187 (2001).
 64. M. Y. Al Aioubi, P. D. Prewett, S. E. Huq, V. Djakov, and A. G. Michette, "A Novel MOEMS Based Adaptive Optics for X-Ray Focusing," *Microelec. Eng.* **83**:1321–1325 (2006).
 65. A. Michette, T. Button, C. Dunare, C. Feldman, M. Forkard, D. Hart, C. Mcfaul, et al., "Active Micro-Structured Arrays for X-Ray Optics," *Proc. SPIE* **6705**:670502.1–11 (2007).

This page intentionally left blank

THE SCHWARZSCHILD OBJECTIVE

Franco Cerrina

*Department of Electrical and Computer Engineering
University of Wisconsin
Madison, Wisconsin*

51.1 INTRODUCTION

The Schwarzschild objective is based on the use of two almost concentric spherical surfaces: a small convex mirror and a larger concave facing each other as shown in Fig. 1.

While this design is widely attributed to Schwarzschild, previous descriptions of the optical system had already been analyzed and published by Paul and Chretien. The design is simple and elegant, and well suited for optical systems with small field of view and high resolution. The Schwarzschild objective is an evolution of the Cassegrain telescope, where the primary and secondary are both nonspherical elements, providing good aberration correction and large field of view. Aspherical optics are, however, difficult and expensive to manufacture, and simpler designs are desirable. The Schwarzschild objective replaces the aspheres with spherical elements, as shown in Fig. 1.

The Schwarzschild objective has found its primary use in microscopy and astronomy at wavelengths where glass lenses are not suitable or where a truly achromatic optical is needed, like in microspectroscopy systems. The applications are thus mainly in the Infrared (FTIR microscopes) and the UV; recently, the Schwarzschild objective has been extended to the Extreme UV region (≈ 13 nm) for microscopy^{1,2-5} and for the development of advanced lithography.⁶ Some of the first X-ray images of the sky were also acquired using Schwarzschild objectives.²

In the Schwarzschild objective each spherical surface forms an aberrated image, and it is a simple exercise in third-order expansion to show that the aberrations can be made to compensate each other. This is because the surfaces have curvatures of different sign, and hence each aberration term is of the opposite sign. All the third-order Seidel aberrations are exactly zero on axis, and only some fifth-order coma is present. However, the focal plane of the Schwarzschild objective is not flat; because of the spherical symmetry, the focal plane forms a spherical surface. Thus, the Schwarzschild objective has excellent on-axis imaging properties, and these extend to a reasonably large field. It is ideally suited for scanning systems, where the field of view is very narrow.

The excellent imaging properties of the Schwarzschild objective are demonstrated by an analysis of the objective's modulation transfer function (MTF). In Fig. 2 one can see the excellent imaging property of the Schwarzschild objective, essentially diffraction limited, as well as the effect of central obscuration (increase of side wings). As shown in Fig. 3 the curve extends to the diffraction limit

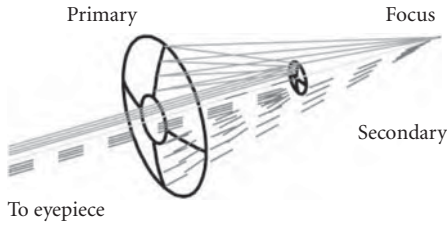


FIGURE 1 Layout of a Schwarzschild objective used as a microscope objective.

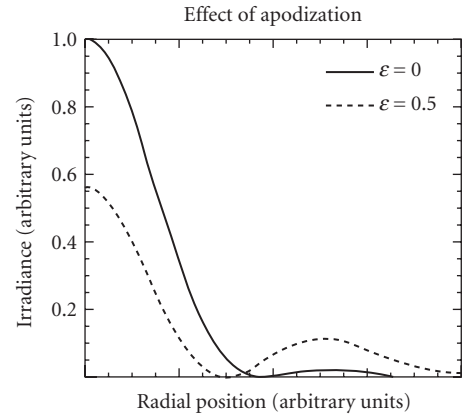


FIGURE 2 Point image of a Schwarzschild objective used as a microscope objective at a wavelength of 13 nm. ϵ refers to the amount of obstruction

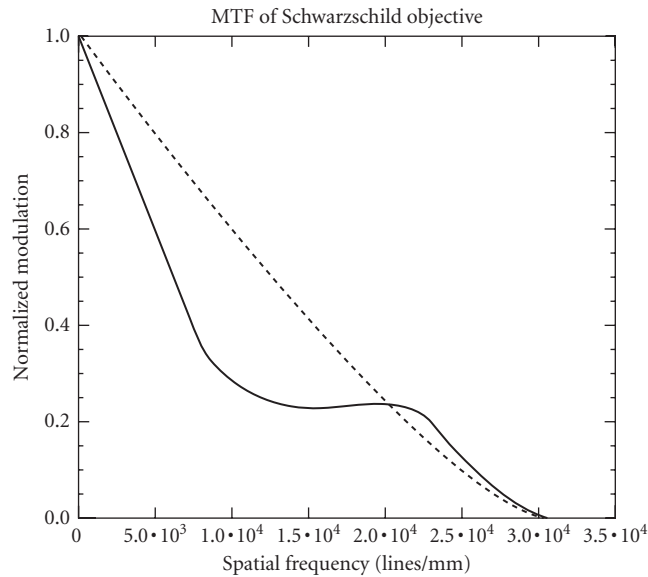


FIGURE 3 Modulation transfer function (MTF) of the previous Schwarzschild objective used as a microscope objective. Notice that in this figure the dotted line corresponds to $\epsilon = 0$ (no obstruction).

and remains very close to the ideal (diffraction limited) MTF curve even at wavelengths as short as 13 nm. However, in the midspatial frequency region, the MTF deviates considerably from the ideal. This behavior is typical of optical systems with central obstruction, and in general does not affect the ultimate resolution of the objective.

51.2 APPLICATIONS TO X-RAY DOMAIN

The high resolution of the microscope is achieved through the use of near normal-incidence optics; off-axis systems of comparable speed have unacceptable aberrations. This poses a big problem in the soft x-ray (EUV) region, where normal-incidence reflectivity is essentially negligible. The development of interference multilayer coatings for the soft x-ray region [now often called *Extreme UV* (EUV)] has in part solved this problem and made possible sophisticated optics.² (Also, see Chap. 41.) The combination of Schwarzschild objective and multilayer coatings, in its simplicity, is a very appealing design for the EUV region. Because the difference in optical path is so small, it is possible to achieve diffraction-limited performance even at wavelengths of 13 nm with good overall transmission. At wavelengths longer than about $2d = 40 \text{ \AA}$ (300 eV) it is possible to use interference filters to increase the reflectivity of surfaces in the soft x rays. These filters, often called “multilayers,” are, in effect, synthetic Bragg crystals formed by alternating layers with large optical contrast.² From another point of view, they are a $\lambda/4$ stack in the very short x-ray region. The possibility of using near normal optical surfaces clearly simplifies the design and allows greater freedom to the optical designer. Furthermore, the relatively high reflectivity of the multilayers (up to 60 percent) allows the design of multiple surface optics.

Several microscopes have been built at synchrotron facilities, and are operated successfully.³ Nonsynchrotron sources do not have enough brightness to deliver the flux required for practical experiments. The best resolution achieved with a Schwarzschild objective is of the order of 90 nm at the MAXIMUM microscope⁷ at the advanced light source. For a numerical aperture of 0.2 at the sample, the diffraction-limited resolution is approximately given by the Rayleigh criterion; at a wavelength of 13 nm, we have $\delta = \lambda/2 \text{ NA} = 32 \text{ nm}$. However, one is limited by the flux (i.e., by the finite brightness of the source) and by mounting and surface errors.³ Diffraction limit operation has not yet been demonstrated.

High resolution is particularly easy to achieve if only a small imaging field is required as in the case of a scanning microscope. The requirement of UHV conditions for the surface physics experiments forced the design of in situ alignment systems, by including piezodriven actuators in the mirror holders. During the alignment procedure, the use of an at-wavelength knife edge test made it possible to reach the ultimate resolution of approximately 900 $\text{\AA}$.⁴ Excellent reflectivities were achieved by using Mo-Si multilayers for the region below 100 eV, and Ru-B₄C for the region around 135 eV.

An example of a Schwarzschild objective is shown in Fig. 4. The two mirrors are coated with a Mo-Si multilayer for 92 eV operation. Notice the groove on the large concave mirror for stress-free clamping. These mirrors are then mounted in the casing shown in Fig. 5, which includes piezoelectric adjustments suitable for ultrahigh vacuum operation. Finally, Fig. 6 shows some images acquired using this objective in a scanning x-ray microscope.^{5,7}



FIGURE 4 Mirrors forming a Schwarzschild objective during assembly.

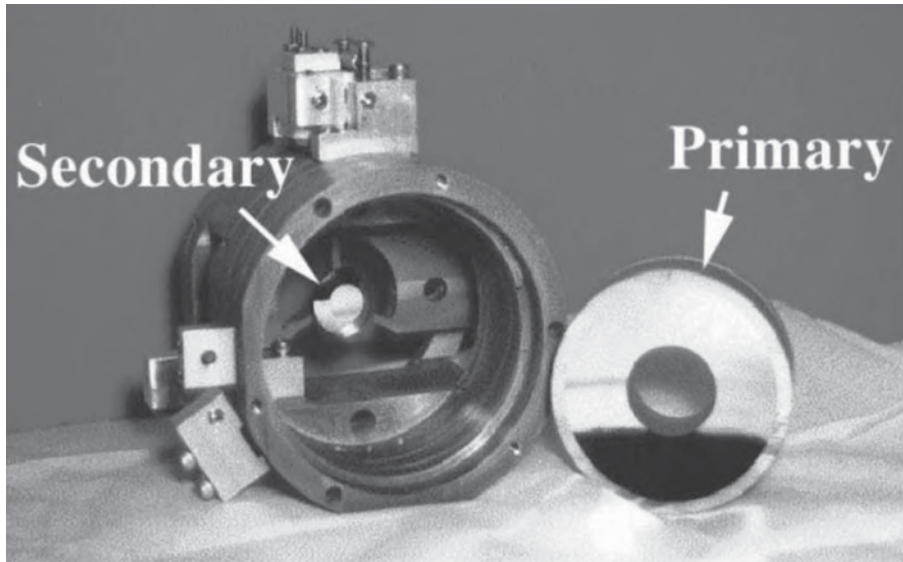


FIGURE 5 Schwarzschild objective in its mount with in situ adjustments.

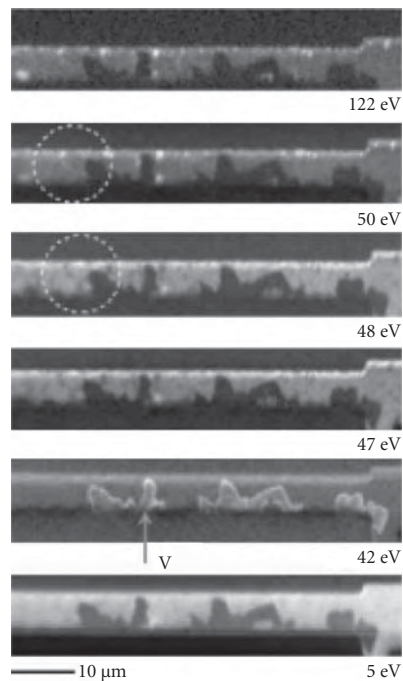


FIGURE 6 Images acquired from an Al-Cu interconnect wire in an integrated circuit after failure using the Schwarzschild objective as part of a scanning photoemission microscope.⁵ The images correspond to different electron kinetic energy and show chemical contrast.

51.3 REFERENCES

1. I. Lovas, W. Santy, E. Spiller, R. Tibbetts, and J. Wilczynski, *SPIE High Resolution X-Ray Optics* **316**:90–97 (1981).
2. D. Attwood, *Phys. Today* **45**:24 (1992).
3. F. Cerrina, *J. Electr. Spectr. and Rel. Phenom.* **76**:9–19 (1995).
4. A. K. Ray-Chaudhuri, W. Ng, S. Liang, and F. Cerrina, *Nucl. Instr. and Methods in Physics* **A347**:364–371 (1992).
5. H. H. Solak, G. F. Lorusso, S. Singh-Gasson, and F. Cerrina, *Appl. Phys. Lett.* **74**(1):22 (Jan. 1999).
6. C. W. Gwyn, R. Stulen, D. Sweeney, and D. Attwood, *Journ. Vac. Sci. and Techn.* **16**:3142–3149 (1998).
7. W. Ng, A. K. Ray-Chaudhuri, S. Liang, S. Singh, H. Solak, F. Cerrina, G. Margaritondo, et. al. *Synchr. Rad. News* **7**:25–29 (Mar./Apr. 1994).

This page intentionally left blank

SINGLE CAPILLARIES

Donald H. Bilderback and Sterling W. Cornaby

*Cornell High Energy Synchrotron Source
School of Applied and Engineering Physics
Cornell University
Ithaca, New York*

52.1 BACKGROUND

Monocapillary x-ray optics can be used to increase the x-ray flux per square micrometer onto a small sample while also controlling the divergence of the x-ray beam. These optics efficiently collect and transport x rays of all energies up to a cutoff energy that is dependent on the capillary material and shape. The past decade has seen the rapid development of elliptically figured monocapillary optics that are designed to condense an x-ray beam and produce a highly demagnified image of the x-ray source (tube or synchrotron) at the sample position. Monocapillary optics are being used in a wide variety of applications such as x-ray diffraction, x-ray fluorescence, small angle x-ray scattering, confocal x-ray microscopy, and so on.

52.2 DESIGN PARAMETERS

Monocapillary optics relies on total external reflection of the x rays from the internal surface of the glass tube to transport x rays. To keep the x rays from being absorbed in the wall of the capillary, the angle of incidence must be kept below the critical angle, which is typically less than 4 milliradians (0.23°). Glass materials that have been used to fabricate capillary optics are borosilicate (Pyrex), lead-based, and silica glasses. The composition of these typical glasses is shown in Table 1.

The critical angle, θ_c , is equal to $(2\delta)^{1/2}$ where δ is the refractive index decrement of the material at the energy of the x-ray photon. For borosilicate glass (Corning 7740) with a density of 2.23 g/cm^3 , θ_c is approximately $(3.8 \times 10^{-2})/E$ radians where E is the x-ray energy in keV.

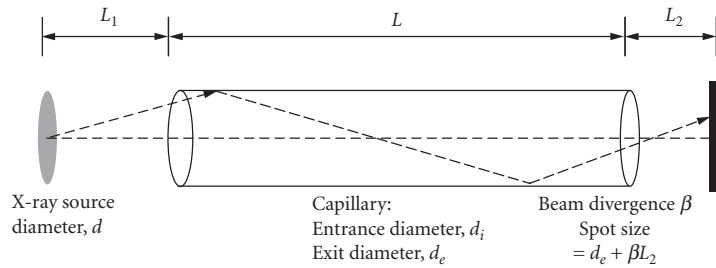
The simplest form of a monocapillary optic is a hollow straight glass tube, as shown in Fig. 1.

Straight capillary tubes were first used in the 1920s when Jentsch and Nahring demonstrated that x rays¹ can be guided down the length of the tube by multiple reflections of x rays at constant angle from the inner glass surface. The intensity of the beam is proportional to the solid angle subtended by the capillary entrance² or to $(d_i/L_1)^2$. The ideal intensity gain over a pinhole at the exit is proportional to $[(L + L_1)/L_1]^2$ provided that the glancing angle at entrance of the capillary is less

TABLE 1 Composition of Typical Glass Starting Materials and Their Basic Physical Properties

Glass Code	Type	Density (g/cm ³)	Softening Point (°C)	Weight Percent (%)							
				SiO ₂	Na ₂ O	K ₂ O	CaO	MgO	PbO	B ₂ O ₃	Al ₂ O ₃
0080	Soda lime	2.47	696	73.6	16	0.6	5.2	3.6	0	0	1.0
7050	Borosilicate	2.25	703	67.3	4.6	1.0	0	0.2	0	24.6	1.7
7740	Borosilicate	2.23	820	80.5	3.8	0.4	0	0	0	12.9	2.2
7900	96% Silica	2.18	1500	96.3	0.2	0.2	0	0	0	2.9	0.4
7910	99% Silica	2.18	1500	99.5	0	0	0	0	0	0	0
8870	High lead	4.28	580	35	0	7.2	0	0	58	0	0

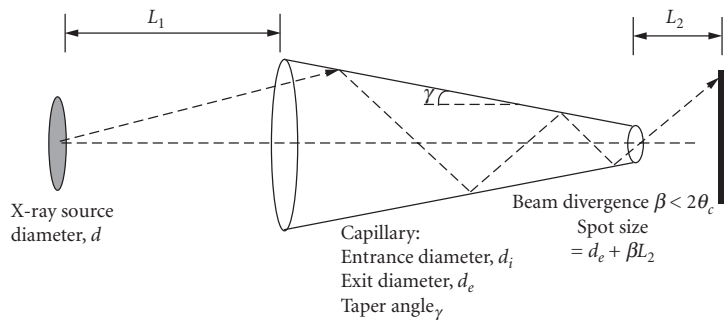
Source: Corning Glass Works, Corning, NY.

**FIGURE 1** Design parameters of a straight capillary tube geometry (diameter of capillary entrance = diameter of exit) and its divergence, β .

than or equal to θ_c . These properties have been used by several investigators to create small beams of x rays that ranged from 10 to 200 μm in diameter.^{3–10}

Condensing capillary optics are figured so that the x-ray beam is compressed with each reflection, producing a gain in intensity. These optics produce a smaller, more intense beam at the expense of divergence. Stern et al.¹¹ were the first to quantitatively describe how x rays would propagate down a linearly tapered or optic. The design parameters for this optic are shown in Fig. 2 and are primarily driven by the following parameters:

1. Divergence of the exit beam, and spot size
2. Distance from the source, source dimensions, and source beam divergence

**FIGURE 2** Design parameters for a condensing capillary optic including its entrance capillary diameter, d_i , its exit diameter, d_e , and its tapering angle, γ .

The acceptance angle, at a given x-ray energy, of the capillary is $2\theta_c$. Upon reflection, the angle of incidence increases by 2γ and only photons for which the angle of incidence is less than the critical angle will emerge from the capillary. The emerging beam has a divergence which is generally less than $2\theta_c$ and the sample must be placed close to the exit end of the capillary to preserve the small spot size. The beam is smallest at the tip and the sample should be positioned no further away than 20 to 100 times the exit diameter.¹² The design of condensing capillary optics relies on the use of ray tracing computer models that allow the capillary shape, figure errors, surface roughness, materials, and source parameters to be included.^{13–15} Advanced condensing capillary designs are being investigated that include using elliptical and hybrid combinations of capillary shapes.^{16,17}

The present state-of-the-art in single channel capillaries are single-bounce capillary optics, which produce a well-defined focus. A single hollow capillary tube is shaped into a parabolic or elliptical curve, (a conic section of eccentricity less than 1) rotated about an axis of symmetry. With the well-controlled shape, the x rays need only one bounce from the inner surface to be directed to the capillary optic's focus. For synchrotron applications, the source is typically located many meters away from the optic and the incident radiation has a low divergence.

The key parameters in the design of these optics are:

1. The working distance (distance F in the diagram). This sets the demagnification that determines the spot size. The spot size is approximately $d^*(L_2/2 + F)/L_1$.
2. The divergence of the x-ray beam incident upon the sample. The full divergence of the optic is $\theta_D = (\text{tip inner diameter})/F$. For total external reflection, $\theta_D < 4\theta_c$.

The elliptically shaped optics were first developed at the University of Melbourne, Australia, and have demonstrated the ability to produce a focused spot of 40 to 50 μm with monochromatic light, and a gain factor of 700 (the increase in flux through a small $5 \times 5 \mu\text{m}$ aperture with the optic) with a divergence angle of approximately 6 milliradians.^{18,19}

Single-bounce monocapillary optics have been subsequently used for a number of years at Cornell High Energy Synchrotron Source (CHESS). They have been made in an array of sizes, with focal lengths ranging from 20 to 150 mm, divergences from 2 to 10 mrad, and have produced spot sizes between 5 and 50 μm , with gains in intensity ranging from 10 to 1000. The single-bounce optics available at CHESS generally have inner base diameter sizes ranging from 80 μm to 1.2 mm, inner tip diameter sizes ranging from 40 μm to 1 mm, and tube lengths ranging from 40 to 150 mm. The real spot size and divergence of the x rays in the image plane depends on the size of the source, the shape of the glass tube, and the slope errors introduced during the manufacturing (Fig. 3).^{20,21}

Elliptically shaped single-bounce capillary x-ray optics are focusing optics and not imaging optics. The optic's focus is an extremely distorted image of the source since the rays are effectively smeared over 720° of rotation.^{20,21} Additionally, the optic's image is smeared due to the optic's range of magnifications; the rays closest to exit tip have the strongest focusing and those furthest from the

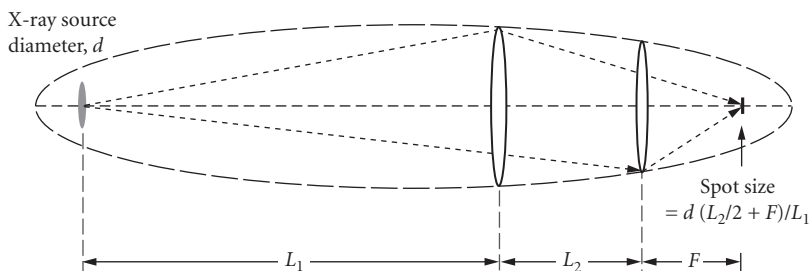


FIGURE 3 Focusing ellipse for single-bounce capillary. L_1 is the distance from the x-ray source to the capillary entrance. L_2 is the length of the capillary and F is distance from the tip of the capillary to the focus.

tip have the weakest focusing (or smallest angular deflection). Grazing incident optics, however, can produce images with a double-bounce Wolter mirror design, with one bounce from an ellipsoid surface and a second bounce from a hyperboloid surface.^{22,23}

52.3 FABRICATION

Capillary optics are typically fabricated by heating the glass tube and then pulling the glass out of the furnace in a controlled way. The shape is varied by changing the rate at which the glass is fed into and pulled out of the furnace and by the effects of surface tension. A schematic of a typical puller is shown in Fig. 4.

This puller, located at the Cornell High Energy Synchrotron Source (CHESS), controls the furnace motion, glass tension, and the rate at which the glass enters and leaves the furnace.²¹ This puller has been used to fabricate capillary optics which are 4 to 20 cm in length starting with outer tubing diameter up to 3 mm. The furnace limits the peak temperature to about 900°C, so that low temperature glasses, such as borosilicate glass, soda lime glass, etc., can be pulled with the present equipment.

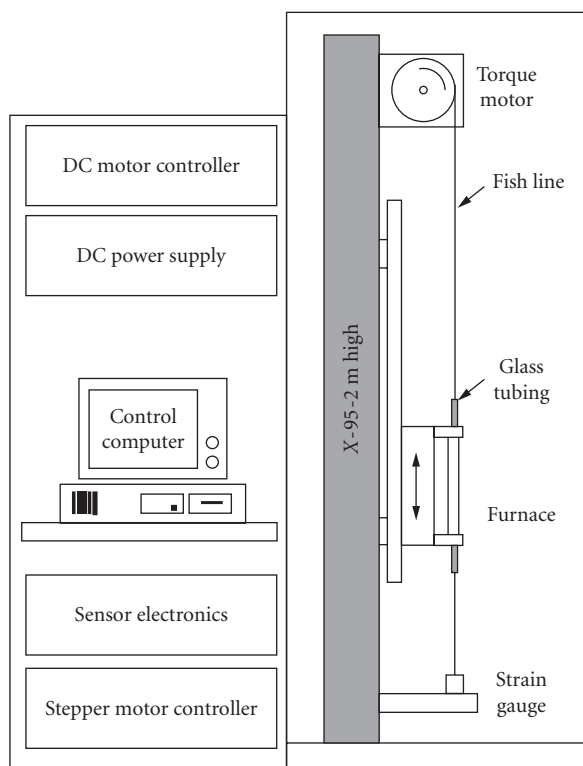


FIGURE 4 A glass tube is suspended in an electric furnace from a piece of fish line that is attached to a strain gauge at the bottom. The torque motor keeps a constant tension as the glass yields during drawing. The furnace is programmed to move based on the amount of glass yielding to make the desired elliptical, parabolic, etc., shape. (See also color insert.)

The fabrication of high efficiency capillary optics requires three factors: (1) a diameter versus length profile that approaches an ideal shape, (2) a concentric capillary bore, and (3) glass that is smooth on an atomic scale. Other glass pullers have been fabricated for capillary drawing.^{21,24,25}

A second method has been developed to fabricate capillary optics from metals. In this technique, the capillary optic is replicated from an ultrasmooth mandrel through sputtering, vacuum evaporation, or other coating techniques, followed by electroforming.²⁶ The mandrel is expendable and is formed by the precision etching of a glass or metal wire into the shape of the desired bore of the capillary optic. This technique has been successfully used to fabricate gold and copper paraboloidal imaging optics that produce a focused spot size of less than 10 μm in diameter with a collimated synchrotron radiation source. Advantages of this technique include wide latitude in the selection of materials compromising the optics, good control of the figure and straightness of the part, and high thermal conductivity.

52.4 APPLICATIONS OF SINGLE-BOUNCE CAPILLARY OPTICS

Applications that have used the single-bounce monocapillary optics are very diverse. They include high pressure powder diffraction, high resolution microdiffraction (μXRD), micro-x-ray fluorescence (μXRF), confocal x-ray fluorescence, microprotein crystallography, Laue protein crystallography, microsmall angle x-ray scattering (μSAXS), x-ray absorption fine-structure (μXAFS), and x-ray absorption near edge structure (μXANES).^{19,21,27–32} The small micrometer-sized x-ray beams have been used on a very large array of samples, including proteins, synthetic and natural fibers, polymers, foils, paintings from antiquity, solid-state devices, biological tissues, and so on. At CHESS, the single-bounce capillary optics have been used to focus beams with bandwidths ranging from 0.01 to 30 percent (they are achromatic), produced spot sizes as small as 5 μm , and total flux throughput ranging from 10^9 to 10^{13} photons/second, as seen in Fig. 5.

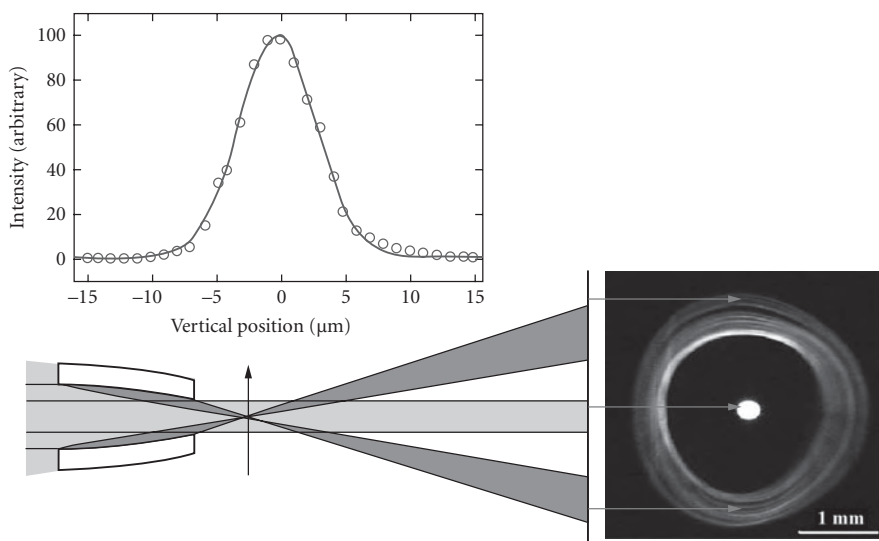


FIGURE 5 Upper panel: Profile of intensity versus 5 μm vertical pinhole position at focus of a 9 milliradians capillary producing a spot size of 5 μm FWHM at a distance of 20 mm beyond the tip of the capillary.²¹ Lower panel: The far-field image shows the direct beam (center dot) passing through the capillary and the once-reflected beam forming the outer ring of intensity. The structure in the ring is due to slope-error imperfections arising from the pulling process. (See also color insert.)

The highest flux density achieved with these optics is 3×10^{12} photons/second in a $10 \mu\text{m}$ spot size ($\sim 4 \times 10^{10}$ photons/second/ μm^2) at the advance photon source (APS).³² The uses of the small x-ray beams produced with these optics, in conjunction with synchrotron sources, are still in infancy. Also, they are just beginning to be used in conjunction with microfocusing x-ray tubes and synchrotron radiation sources for functions such as condensers in x-ray microscopes.³³

52.5 APPLICATIONS OF CONDENSING CAPILLARY OPTICS

Condensing capillary optics are used to produce small spots of x rays whose diameter ranges from 1 to about $25 \mu\text{m}$ with x-ray tubes and from 0.1 to $10 \mu\text{m}$ with synchrotron sources. A major application area is in determining the composition and structure of heterogeneous samples through x-ray fluorescence and diffraction. In these techniques, capillaries are used to increase the flux density of the x-ray beam onto the sample. By scanning a sample through the beam and recording the resulting x-ray fluorescence spectra, one can obtain the spatial distribution of major, minor, and trace elements contained within the sample. The diffracted radiation can be used to identify the crystallographic structure of particular regions of the sample. Diffraction data has been obtained with x-ray beams from 0.8 to $0.05 \mu\text{m}$ in diameter.^{17,34–36} Yamamoto developed an integrated spectrometer that he used for the simultaneous measurement of local strain and trace metal contaminants in integrated circuits.^{17,35,37–39} York has developed and implemented multiple capillary-based microdiffractometers for defect analysis in magnetic disk read heads.³⁴ Capillary-based microbeams have been used to study the impact of crystallographic structure on the sensitivity of IC interconnects to the effects of electromigration.^{35,37–39} Other laboratory based microfluorescence instruments have been developed to produce x-ray microbeams that range from $200 \mu\text{m}$ to approximately $5 \mu\text{m}$ in diameter.^{7–10,16,38–41}

52.6 CONCLUSIONS

Single-bounce monocapillary optics are still in the developmental stage. They are highly useable for making micrometer-sized x-ray beams, but there is still further room for improvement. The most serious limitations with single-bounce optics arises from slope errors and lack of centerline straightness during the time of manufacture. Presently, slope errors are at the $50 \mu\text{rad}$ level and centerline errors are generally $1 \mu\text{m}$ or less for the Cornell puller. Slope errors on the order $10 \mu\text{rad}$ or less will be required to take full advantage of third generation synchrotron sources. One advantage these grazing incidence devices have over many other competing microbeam optics (e.g., zone plates, refractive lenses, etc.) is that they are achromatic; they work over a wide energy range without changing the position or the spot size and they are very easy to add or align on a beam line when x-ray microbeam capability is required.

52.7 ACKNOWLEDGMENTS

CHESS is supported by the National Science Foundation and NIH-NIGMS via NSF grant DMR-0225180.

52.8 REFERENCES

1. F. Jentzch and E. Nhring, "Die Fortleitung von Licht—und Rntgenstrahlen durch Rhren," *Zeitschr. F. Techn. Phys.* **12**:185 (1931).
2. W. T. Vetterling and R. V. Pound, "Measurements on an X-Ray Light Pipe at 5.9 and 14.4 keV," *J. Opt. Soc. Am.* **66**(10):1048 (1976).
3. D. Mosher and S. Staphanakis, "X-Ray Light Pipes," *Appl. Phys. Lett.* **29**:105 (1976).

4. P. S. Chung and R. H. Pantell, "Properties of X-Ray Guide Tubes," *Electron. Lett.* **13**:527 (1977).
5. H. Nakazawa, "X-Ray Guide Tube for Diffraction Experiments," *J. Appl. Crystallog.* **16**:239 (1983).
6. A. Rindby, "Application of Fiber Technique in the X-Ray Region," *Nucl. Instrum. Methods A* **249**:536 (1986).
7. P. Engström, S. Larsson, A. Rindby, and B. Stocklassa, "A 200 μm X-Ray Microbeam Spectrometer," *Nucl. Inst. Meth. Phys. Res.* **B36**:222 (1989).
8. D. A. Carpenter, "An Improved Laboratory X-Ray Source for Microfluorescence Analysis," *X-Ray Spectrometry* **18**:253 (1989).
9. H. Fukumoto, Y. Kobayashi, M. Kurahashi, and A. Kawase, "Development of a High Spatial Resolution X-Ray Fluorescence Spectrometer and its Application to Quantitative Analysis of Biological Systems," *Advances in X-Ray Analysis* **35**:1285 (1992).
10. S. Shimomura and H. Nokazawa, "Scanning X-Ray Analytical Microscope Using X-Ray Guide Tube," *Advances in X-Ray Analysis* **35**:1289 (1992).
11. E. A. Stern, Z. Kalman, A. Lewis, and K. Lieberman, "Simple Method for Focussing X Rays Using Tapered Capillaries," *Appl. Opt.* **27**(24):5135 (1988).
12. D. H. Bilderback, D. J. Theil, R. Pahl, and K. E. Brister, "X-Ray Applications with Glass-Capillary Optics," *J. Synchrotron Rad.* **1**:37 (1994).
13. L. Vincze, K. Janssens, A. Rindby, and F. Adams, "Detailed Ray-Tracing Code for Capillary Optics," *X-Ray Spectrometry* **24**:27 (1995).
14. D. X. Balaic and K. A. Nugent, "The X-Ray Optics of Tapered Capillaries," *Appl. Opt.* **34**:7263 (1995).
15. D. J. Thiel, "Ray-Tracing Analysis of Capillary Concentrators for Macromolecular Crystallography," *J. Synchrotron Rad.* **5**:820 (1998).
16. A. Attaelmanan, P. Voglis, A. Rindby, S. Larsson, and P. Engström, "Improved Capillary Optics Applied to Microbeam X-Ray Fluorescence: Resolution and Sensitivity," *Rev. Sci. Instrum.* **66**(1):24 (1995).
17. N. Yamamoto, "A Micro-Fluorescent/Diffracted X-Ray Spectrometer with a Micro-X-Ray Beam Formed by a Fine Glass Capillary," *Rev. Sci. Instrum.* **69**(9):3051 (1996).
18. D. X. Balaic, K. A. Nugent, Z. Barnea, R. Garrett, and S. W. Wilkins, "Focussing of X-Rays by Total Reflection from a Paraboloidally-Tapered Glass Capillary," *J. Synchrotron Rad.* **2**:296 (1995).
19. D. X. Balaic, Z. Barnea, K. A. Nugent, R. Garrett, J. N. Varghese, and S. W. Wilkins, "Protein Crystal Diffraction Patterns Using Capillary-Focused Synchrotron X-Ray Beam," *J. Synchrotron Rad.* **3**:289 (1996).
20. R. Huang and D. H. Bilderback, "Single-Bounce Monocapillaries for Focusing Synchrotron Radiation: Modeling, Measurements and Theoretical Limits," *J. Synchrotron Rad.* **13**:74–84 (2006).
21. S. Cornaby, "The Handbook of X-Ray Single-Bounce Monocapillary Optics, Including Design of the Optics and Synchrotron Applications," Dissertation, Cornell University, May 2008.
22. H. Wolter, "Spiegelsysteme Streifenden Einfalls Als Abbildende Optiken für Röntgenstrahlen," *Annalen der Physik* **445**(1–2):94–114 (1952).
23. H. Takano, S. Aoki, M. Kumegawa, N. Watanabe, T. Ohhigashi, T. Aota, K. Yamamoto, et al., "X-Ray Scattering Microscope with a Wolter Mirror," *Rev. Sci. Instrum.* **73**(7):2629–2633 (2002).
24. A. Rindby, "Production of Capillaries," *Proc. First International Developers Workshop on Glass Capillary Optics for X-Ray Microbeam Applications*, Don Bilderback and Per Engstrom, eds., Cornell University, 157 (October 18–19, 1996).
25. M. Haller, J. Heck, S. Kneip, A. Knöchel, F. Lechtenberg, and M. Radtke, "Stretching Tapered Capillaries from Bubbles," *Proc. First International Developers Workshop on Glass Capillary Optics for X-Ray Microbeam Applications*, Don Bilderback and Per Engstrom, eds., Cornell University, 165 (October 18–20, 1996).
26. G. Hirsch, "Tapered Capillary Optics," US Patent—5,772,903 (1998).
27. A. A. Sirenko, A. Kazimirov, S. Cornaby, D. H. Bilderback, B. Neubert, P. Bruckner, F. Scholz, V. Shneidman, and A. Ougazzaden, "Microbeam High Angular Resolution X-Ray Diffraction in InGaN/GaN Selective-Area-Grown Ridge Structures," *Appl. Phys. Lett.* **89**:181926-1 to 181926-3 (2006).
28. K. Limburg, R. Huang, and D. Bilderback, "Fish Otolith Trace Element Maps: New Approaches with Synchrotron Microbeam X-Ray Fluorescence," *X-Ray Spectrometry* **36**:336–342 (2007).
29. A. R. Woll, J. Mass, C. Bisulca, R. Huang, D. H. Bilderback, S. Gruner, and N. Gao, "Development of Confocal X-Ray Fluorescence (XRF) Microscopy at the Cornell High Energy Synchrotron Source," *Appl. Phys. a-Mater* **83**(2):235–238 (2006).

30. J. S. Lamb, S. Cornaby, K. Andresen, L. Kwok, H. Y. Park, X. Y. Qiu, D. M. Smilgies, D. H. Bilderback, and L. Pollack, "Focusing Capillary Optics for use in Solution Small-angle X-Ray Scattering," *J. Appl. Crystallog.* **40**:193–195 (2007).
31. S. Cornaby, T. Szebenyi, R. Huang, and D. H. Bilderback, "Design of Single-Bounce Monocapillary X-Ray Optics," *Advances in X-Ray Analysis* **50**:194–200 (2006).
32. R. A Barrea, R. Huang, S. Cornaby, D. Bilderback, and T. Irving, "High Flux Hard X-Ray Microprobe Using a Double Focused Undulator Beam and a Single-Bounced Monocapillary," *J. Synchrotron Rad.* **16**:76–82 (2009).
33. X. Zeng, F. Duewer, M. Feser, C. Huang, A. Lyon, A. Tkachuk, and W. Yun, "Ellipsoidal and Parabolic Glass Capillaries as Condensers for X-Ray Microscopes," *App. Opt.* **47**(13):2376–2381 (2008).
34. B. R. York, "Chemical Analysis Using X-Ray Microbeam Diffraction (XRMD) with Tapered Capillary Optics," *Microbeam Analysis*, Proc. 28th Annual MAS Meeting, 11 (1994).
35. N. Yamamoto, Y. Homma, S. Sakata, and Y. Hosokawa, "X-Ray Spectrometer with a Submicron X-Ray Beam for ULSI Microanalysis," *Mat. Res. Soc. Symp. Proc.* **338**:209 (1994).
36. D. H. Bilderback, S. A. Hoffman, and D. J. Theil, "Nanometer Spatial Resolution Achieved in Hard X-Ray Imaging and Laue Diffraction Experiments," *Science* **263**:201–203 (1994).
37. N. Yamamoto and Y. Hosokawa, "Development of an Innovative 5 $\mu\text{m}\phi$ Focussed X-Ray Beam Energy Dispersive Spectrometer and its Applications," *Jpn. J. Appl. Phys.* **27**(11):L2203 (1988).
38. N. Yamamoto and S. Sakata, "Applications of a Micro X-Ray Spectrometer and a Computer Simulator for Stress Analyses in Al Interconnects," *Jpn. J. Appl. Phys.* **28**(11):L2065 (1989).
39. N. Yamamoto and S. Sakata, "Strain Analysis in Fine Al Interconnections by X-Ray Diffraction Spectrometry Using Micro X-Ray Beam," *Jpn. J. Appl. Phys.* **34**:L664 (1995).
40. P. -C. Wang, G. S. Cargill, I. C. Noyan, and C. -K. Hu, "Electromigration-Induced Stress in Aluminum Conductor Lines measured by X-Ray Microdiffraction," *Appl. Phys. Lett.* **72**:1296 (1998).
41. D. A. Carpenter, A. Gorin, and J. T. Shor, "Analysis of Heterogeneous Materials with X-Ray Microfluorescence and Microdiffraction," *Advances in X-Ray Analysis* **38**:557 (1995).

POLYCAPILLARY X-RAY OPTICS

Carolyn MacDonald

*University at Albany
Albany, New York*

Walter Gibson*

*X-Ray Optical Systems, Inc.
East Greenbush, New York*

53.1 INTRODUCTION

Polycapillary optics are arrays of a large number of small hollow glass tubes.^{1,2} X rays are guided down these curved and tapered tubes by multiple reflections in a manner analogous to the way fiber optics guide light. Like micropore and multifoil optics (see Chaps. 48 and 49), they differ from single bore capillaries in that the focusing or collecting effects come from the overlap of the beams from hundreds of thousands of channels, rather than from the action within a single tube. A cross section of a polycapillary fiber is shown in Fig. 1. As for single bore capillaries, x rays can be transmitted down a curved hollow tube as long as the tube is small enough, and bent gently enough, to keep the angles of incidence less than the critical angle for total reflection θ_c . The critical angle for borosilicate glass is approximately

$$\theta_c \approx \frac{30 \text{ keV}}{E} \text{ mrad} \quad (1)$$

which is approximately 30 mrad (1.7°) for 1-keV-photons and 1.5 mrad (0.086°) for 20-keV photons. The angles are somewhat larger for leaded glass. As shown in Fig. 2, the angle of incidence for a ray near one edge (toward the center of curvature) increases with tube diameter. The requirement that the incident angles remain less than the critical angle necessitates the use of tiny tubes. However, mechanical limitations prohibit the manufacture of capillary fibers with outer diameters smaller than about 300 μm . For this reason, polycapillary fibers, which have tube diameters that are much smaller than the fiber diameter, are employed. Typical channel sizes are between 2 and 50 μm , although some research has been performed with channel sizes down to submicron sizes.³

While large area optics can be produced by stringing thousands of such fibers through lithographically produced metal grids to produce a multifiber lens, as shown in Fig. 3, most commercial optics are one-piece, monolithic optics as sketched in Fig. 4.

Systems involving the use of a large number of capillary channels were first suggested by M. A. Kumahov and his collaborators in 1986.⁴ The development and study of polycapillary optics and applications^{2,5-13} have been pursued since 1990. Polycapillary optics are well-suited for broadband or divergent radiation. They have been used as focusing collectors for x-ray astronomy, to produce large area

*In memoriam Distinguished Professor Emeritus Walter M. Gibson.

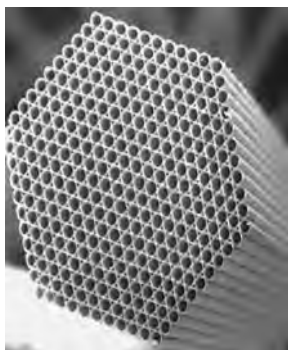


FIGURE 1 Cross-sectional scanning electron micrograph of a polycapillary fiber with 0.55-mm outer diameter and 50- μm -diameter channels.

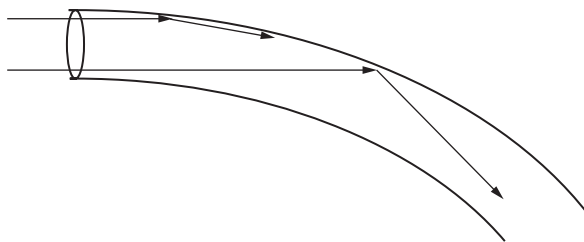


FIGURE 2 X rays traveling in a bent capillary tube. The ray entering at the bottom (closest to the center of curvature) strikes at a large angle. (*Adapted from Ref. 8.*)



FIGURE 3 Multifiber collimating lens constructed from over a thousand individual polycapillary fibers strung through a metal grid. The lens is 10 cm long with an output of 20×20 mm. The fibers are parallel at the output end (shown) and at the input end point to a common focal spot.



FIGURE 4 Sketch of the interior channels of a monolithic polycapillary optic. Monolithic optics can be focusing, as shown, or collimating, as in Fig. 3. (*Sketch from Ref. 14.*)

collimated beams for wafer analysis, and to provide small focused beams for protein crystallography with low power x-ray sources. They are also being developed for a number of medical applications, including the removal of Compton scattering with the resultant improvement in contrast and resolution in mammography,^{9,11} the production of monochromatic parallel beams for high-contrast imaging in a clinical setting,¹⁵ and the detection and localization of radioactive tracers in microscintigraphy.^{16–19}

53.2 SIMULATIONS AND DEFECT ANALYSIS

The realization of numerous applications has been advanced by the development of simulation analyses which allow for increasingly accurate assessment of optics defects. These computer codes, like Shadow²⁰ (see Chap. 35), are generally based on Monte Carlo simulations of geometrical optics trajectories and provide essential information on performance, design, and potential applications of capillary optics.²¹ Some simulations also allow for the roughness^{22,23} and waviness^{24,25} of the capillary walls, as well as channel blockage and profile error to be taken into account. Optics performance over a range of energy from 10 to 80 keV can often be matched with one or two fitting parameters. For submicron channels, wave effects become significant.

Detailed measurements of polycapillary fibers, including transmission, absorption, and exit divergence, have been performed as a function of length, bend radius, x-ray source position, x-ray source geometry, and x-ray energy.^{9,12,26–28} Computer-automated systems for fiber measurement have repeatability within 1 percent.²⁹ Very good agreement is found between simulation and experimental results for a wide range of geometries.

Profile Error

Bending the channels increases the x-ray incidence angles, as shown in Fig. 2. Because the critical angle θ_c is inversely proportional to the x-ray photon energy, bending decreases the x-ray transmission down the channels most significantly at higher-photon energies. Lenses can be designed to have curvatures which deliberately discriminate against high-photon energies, for example, to reject higher-order harmonics from monochromators. Experimental data taken on a nominally straight fiber is shown in Fig. 5. The transmission drops off above 45 keV. The model cannot fit the entire spectra with bending alone. A bending radius smaller than 100 m would be necessary to fit the mid-range energies, but underestimates the high-energy transmission.

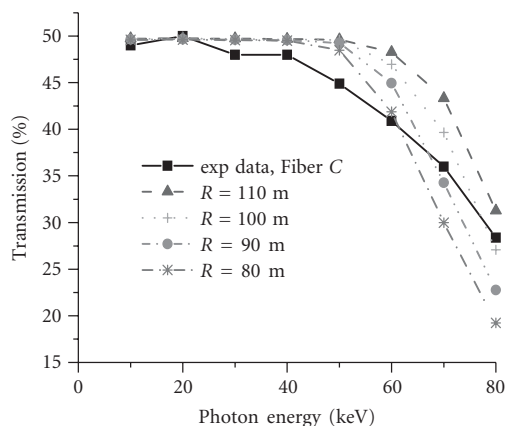


FIGURE 5 Transmission spectra of a nearly straight fiber simulated with different bending curvatures alone and compared with experimental data. (From Ref. 25.)

Waviness

Midrange spatial frequency slope errors, i.e., surface oscillations with wavelengths shorter than the capillary length and longer than the wavelength of the roughness, are often called waviness. The detailed shape of the channel walls is unknown, but waviness is modeled as a random tilt of the glass wall. The tilt angles are assumed to have a Gaussian distribution with width σ .²⁵ For high quality glass and photon energies less than 200 keV, σ is much smaller than the critical angle θ_c . Consideration is taken in the simulation of the fact that the surface tilt angle will affect the probability of x-ray impact on that surface.²⁴ The effect of waviness on fiber transmission is shown in Fig. 6. Waviness is primarily responsible for reduction of transmission at midrange energies, and additionally, for an increased reduction in transmission as the source is moved away from the fiber axis. A simulation fit including waviness and bending for a single 0.5-mm diameter fiber with 10 μm channels is shown in Fig. 7.³⁰ Most borosilicate and lead glass optics have simulation fitting parameters which give a Gaussian width for the waviness of 0.12 to 0.15 mrad. This is in agreement with the slope-error data of the Cornell group.³¹

Roughness

Roughness is treated as reducing the reflectivity of the glass (see Chap. 46). Roughness only slightly decreases the specular reflectivity at low angles, and so has almost no impact on the transmission spectra, but becomes increasingly important under circumstances in which the angle and number of reflections increase. Surface roughness must be considered to model the effects of moving the source away from the focal point.

Blockage

Another defect that is seen occasionally in borosilicate glass optics, and more prevalently in lead glass fibers,^{27,32} is a drop in transmission at low energies, as shown in Fig. 8. Reasonable agreement is obtained over the whole range of photon energies by assuming that a glass layer of fixed thickness

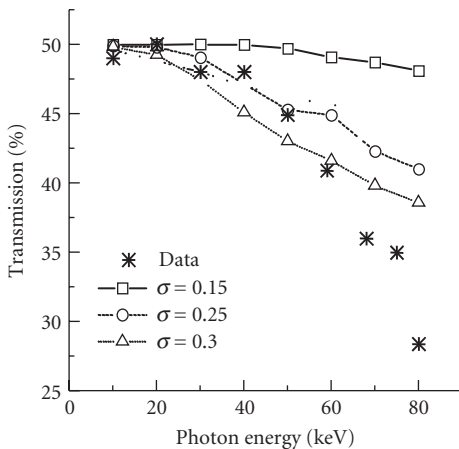


FIGURE 6 Simulations of transmission spectra for a fiber with waviness values from 0.15 to 0.3 mrad compared with the experimental data. The simulations do not include the roughness or bending. (From Ref. 24.)

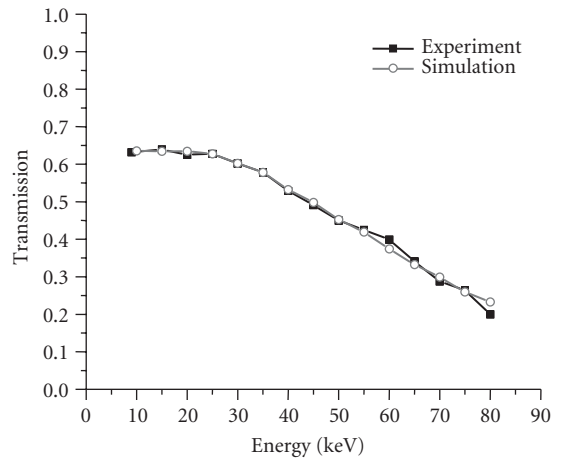


FIGURE 7 Transmission of a single fiber of the type used for a multifiber lens, compared to a simulation with fitting parameters waviness = 0.15 mrad and unintentional central axis bending radius $R = 120$ m. (From Ref. 30.)

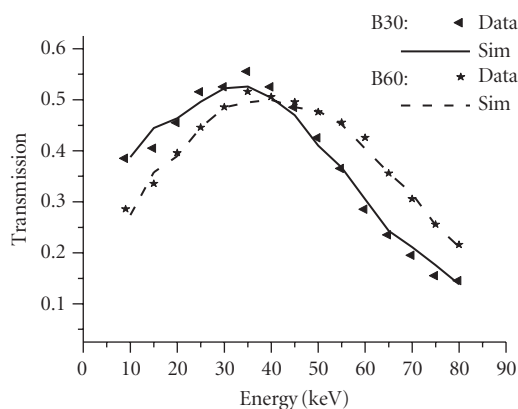


FIGURE 8 Transmission of two similar lead glass fibers, 30 and 60 mm in length. The simulation fits include 17 and 33 μm of glass layer, respectively, or 0.55 μm of blockage per mm of length. (Figure from Ref. 27.)

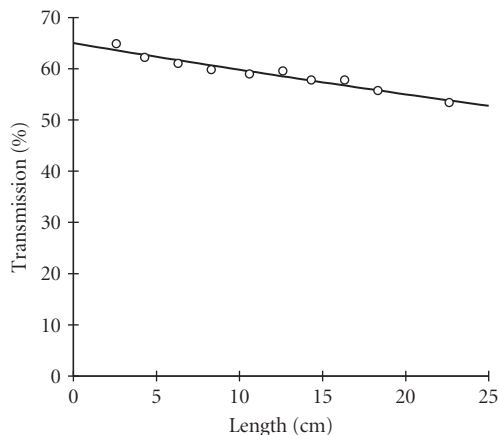


FIGURE 9 Measured transmission at 8 keV, as a function of length, for borosilicate polycapillary fibers, with 17 μm channels. Solid line is an exponential fit with decay length = 120 cm. (Adapted from Ref. 9.)

blocks the channels. The increase in required layer thickness with fiber length is consistent with a stochastic random model of glass inclusions. This random probability of glass inclusions would cause the transmission to drop exponentially with optic length, as shown in Fig. 9.

53.3 RADIATION RESISTANCE

Because the x-ray optical properties of materials depend on total electron density, the optical constants are insensitive to changes in electronic state. Color centers that form rapidly in glass during exposure to intense radiation are not indicative of a change in the x-ray transmission of a polycapillary optic. Thin fibers exposed to intense beams undergo reversible deformation. However, rigid optics, if annealed in situ at 100°C, were shown to withstand in excess of 2 MJ/cm² of white-beam bending-magnet radiation without measurable change in performance at 8 keV.³³

53.4 ALIGNMENT AND MEASUREMENT

Standard techniques have been developed for aligning and characterizing polycapillary optics. A typical setup is shown in Fig. 10.³⁴ Depending on the source geometry and desired beam placement, either the source location or the optic position is translated perpendicular to the optic axis in small steps, producing a measurement of intensity versus relative source position, as shown in Fig. 11.³⁵ The plot is symmetric and Gaussian, which indicates good alignment of the source, optic and detector. In order to determine the focal distance of the optic, source scans are performed at different distances from the source. At the focal distance, the ratio of the width of the scan curve to the optic-to-source distance, called the source scan angle, should be the smallest, as shown in Fig. 12.⁹

Transmission is the ratio of the number of photons passing along the channels to the number incident on the front face of the optic. Transmission with respect to the source-optic distance is also shown in Fig. 12. The highest transmission and the lowest source angle occurred at the focal distance.

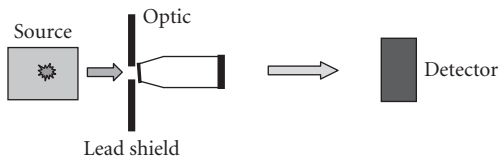


FIGURE 10 Set up for optics measurement.

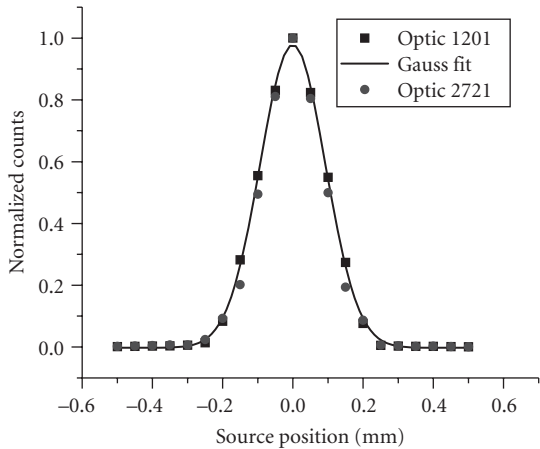


FIGURE 11 Source scan plot at 17.5 keV for two different optics with input focal lengths ranging from 48 to 56 mm. (From Ref. 34.)

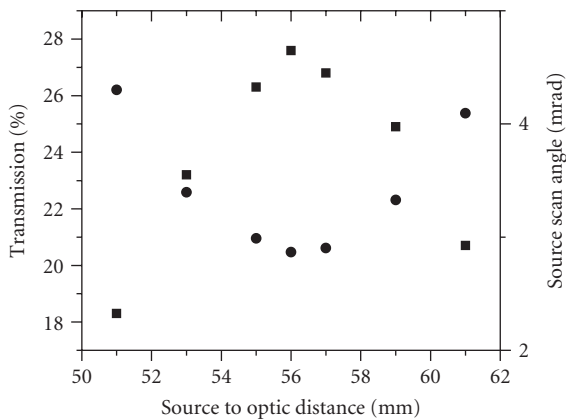


FIGURE 12 Transmission (■) and source angle (●) with respect to the source optic. The transmission is the highest at the designed focal distance of 56 mm. (From Ref. 34.)

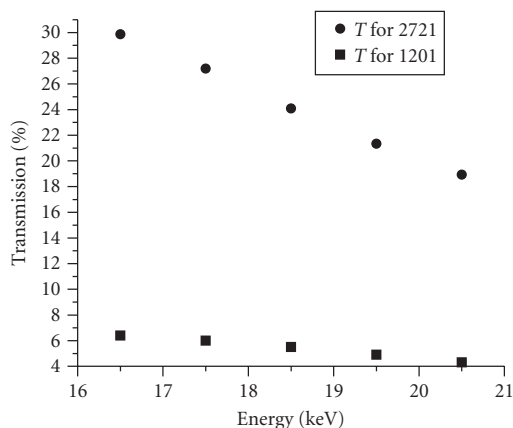


FIGURE 13 Transmission with respect to energy for optic 1201 (■) and 2721 (●).

Transmission with respect to energy can then be measured, as shown in Fig. 13.³⁴ At 17.5 keV, the transmission is 6 percent for optic 1201 and 27 percent for optic 2721 with the source at the focal distance. The transmission of optic 2721 was about 4.5 times larger than that of optic 1201 because the optics were designed with different radii of curvature R of the outer channel and different channel size c , resulting in different maximum incident angle.² The maximum incident angle for a parallel input beam is about $\theta_i = \sqrt{2c/R}$.² For optic 1201, $c = 10 \mu\text{m}$ and $R \approx 1.5 \text{ m}$, giving a maximum incident angle of 3.9 mrad, equal to the critical angle for 8 keV, so the transmission is poor for the measured energies. Instead, for optic 2721 $c = 7 \mu\text{m}$ and $R \approx 2.3 \text{ m}$, giving a maximum incident angle of 2.3 mrad, equal to the critical angle for 13-keV photons. A channel halfway from the center to the outer edge had a maximum incident angle of 1.6 mrad, equal to the critical angle for 18 keV. Thus, the overall optics performance should drop off starting in this range of 13–18 keV. Measured and calculated intensity after the optic is shown in Fig. 14 for a low power source. The measured values were

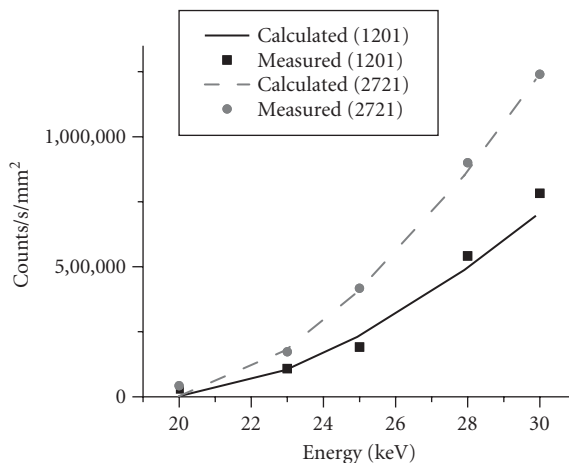


FIGURE 14 Comparison of calculated and measured intensity after optic 1201(■) and 2721(●) for a source current of 0.1 mA. (From Ref. 34.)

in good agreement with calculated values. The counts per unit time and unit area after optic 2721 are about 1.7 times larger than that after optic 1201. This factor is less than $T_{2721}/T_{1201} = 4.5$. That is because the two optics have the different focal distances and different areas.

53.5 COLLIMATION

As shown in Fig. 15, the output from a multifiber polycapillary collimating optic has both global divergence α and local divergence β . Even if the fibers are parallel ($\alpha = 0$), the output divergence β is not zero, but is determined by the critical angle and therefore the x-ray energy.

The exit divergence from capillary optics is measured by rotating a high-quality crystal in the beam and measuring the angular width of a Bragg peak. Since the Darwin width and mosaicity of the crystal are typically much smaller than the exit divergence from the optic, the measurement yields the divergence directly. The result at 8 keV, for a 5-mm diameter monolithic collimating optic, is a full width at half maximum of 3 mrad, which is approximately given by the critical angle for total reflection.³⁶

The divergence of the beam, and therefore the angular resolution of a diffraction measurement using polycapillary optics, does not depend on the source size, unlike the case for pinhole collimation. Thus larger, higher-power sources may be used without adversely affecting the resolution of the measurement. The maximum useful x-ray source spot size is limited by the acceptance area of the collimating lens, which is about 1 mm wide and 15 mm long for a multifiber lens, well matched for a typical rotating anode. Conversely, the divergence does not decrease for smaller sources because waviness increases the angle of reflection for x-ray photons and thus the average angle at which they exit the fiber. For example, for the case of a small source with a local divergence of 2.4 mrad, a simulation at 8 keV with no channel wall defects produces a divergence less than the critical angle, but for a simulation including a typical waviness of 0.15 mrad, the divergence grows to 3.9 mrad, which matches the measured value.²⁴

A comparison of measured and simulated transmission for a collimating lens is shown in Fig. 16. The output varies by less than 3 percent across the face of the optic, as shown in Fig. 17. The transmission of a 3-cm square multifiber collimating lens developed for astrophysical applications is shown as a function of photon energy in Fig. 18. The “gain” obtained from any optic ultimately should be a figure of merit defined by a particular application. For a diffraction measurement which requires 2D collimation, e.g., strain measurements, the intensity ratio delivered within a given output divergence angle is a reasonable figure.

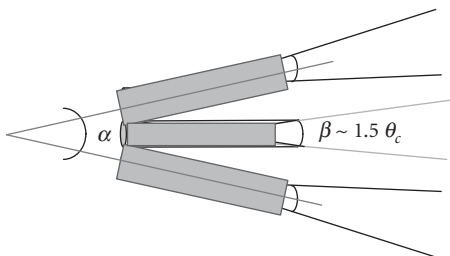


FIGURE 15 The output divergence from a multifiber collimating optic is characterized by global and local divergence. (From Ref. 2.)

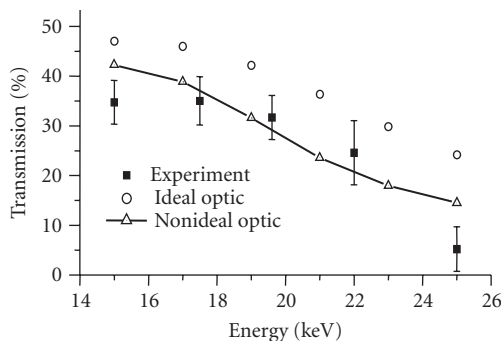


FIGURE 16 Comparison of experimental and simulated transmission as a function of energy for a multifiber lens with a 250 mm input focal length. The nonideal simulation includes a waviness of 0.15 mrad and an unintentional bend with a radius of 125 m for the center fibers. (From Ref. 15.)

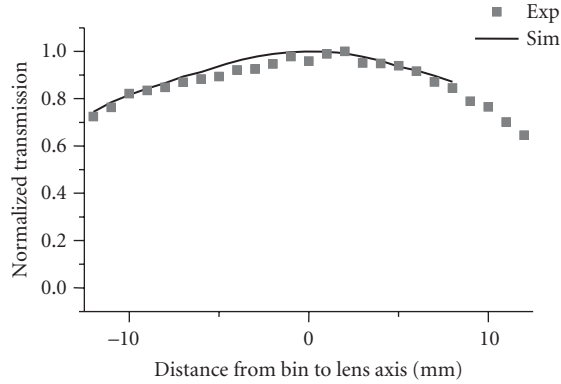


FIGURE 17 Measured and simulated local transmission of the lens in Fig. 16, at 8 keV. The uniformity scan was carried out by scanning a 5×6 mm lead aperture across the output beam. (From Ref. 30.)

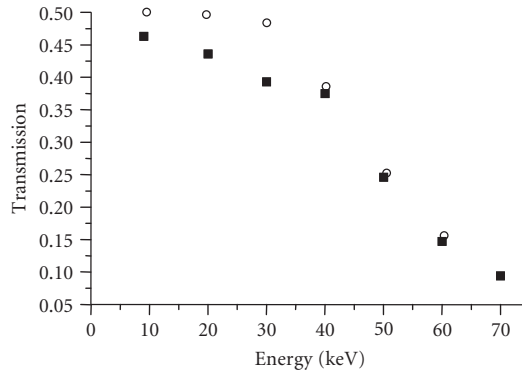


FIGURE 18 Simulated⁶ (circles) and measured⁷ (squares) transmission of a 3-cm square multifiber collimating lens with 2-m focal length.

53.6 FOCUSING

Polycapillary lenses can be used to collect broadband divergent radiation and redirect it toward a focal spot. The first Kumakhov capillary lens was built to demonstrate focusing. It had channel sizes of $300 \mu\text{m}$ and was about 1 m long.³⁷ Smaller channel sizes allow for tighter bending and shorter lenses.

Focusing the beam increases the intensity on a small sample, compared to pinhole collimation. The intensity gain depends on the spot size produced by the optic, which is determined by the capillary channel size c , output focal length f , and x-ray critical angle θ_c ,

$$d_{\text{spot}} \approx \sqrt{c^2 + (1.3 \cdot f_{\text{out}} \cdot \theta_c)^2} \quad (2)$$

The factor 1.3 is an experimentally determined parameter that arises from the fact that most of the beam has a divergence less than the maximum possible divergence of $2\theta_c$ produced by reflection.

TABLE 1 Results for Monolithic Focusing Optic in a Synchrotron Beam³⁹

X-Ray Energy (keV)	Spot Size (mm)	Transmission (%)	Measured Gain 350- μ m Pinhole	Calculated Gain 350- μ m Pinhole	Calculated Gain 90- μ m Pinhole	Calculated Gain 10- μ m Pinhole
6	0.09	36	78	81	645	911
8	0.08	49	96	110	933	1359
10	0.09	39	83	87	624	842
12	0.09	39	74	87	654	903
white	0.17	42	11	89	243	266

The critical angle θ_c at 20 keV is 1.5 mrad. A lens with $c = 3.4 \mu\text{m}$ and $f_{\text{out}} = 9 \text{ mm}$ has a predicted spot size of $18 \mu\text{m}$. An intensity distribution measurement, made by scanning a small pinhole, gave a FWHM of $21 \mu\text{m}$.³⁸

For the case of a very small sample with diameter σ , the gain, relative to pinhole collimation, is given by

$$\text{Gain} = \frac{P_{\text{lens}}}{P_{\text{pinhole}}} = \left(\frac{d_{\text{spot}}^2}{f_{\text{in}}^2} \right) T_{\text{Lens}} \left(\frac{L_{\text{source-sample}}}{\sigma} \right)^2 \quad (3)$$

For the slightly convergent lens used to produce lysozyme diffraction data,³⁶ the computed gain for a 0.5-mm pinhole is 124. The measured intensity gain is 110. The intensity gain for a more highly convergent lens with an output focal spot of $21 \mu\text{m}$ is 2400 at a source to sample distance of 100 mm. Because of the divergence from each channel, lenses with smaller focal lengths have smaller spot sizes, as do measurements at higher photon energies.

Polycapillary optics can also be used to focus parallel beam radiation for astrophysical or synchrotron applications. A 5-mm diameter lens with a 17-mm focal length was measured using knife-edge scans in a synchrotron beam.³⁹ The measured spot size, transmission, and gain are shown in Table 1 along with calculated gains.

53.7 APPLICATIONS

Energy Filtering

The dependence of the critical angle for reflection on photon energy results in an energy-dependent transmission, which can be varied with lens parameters, as shown in Figs. 16 and 18. Thus, capillary optics can be used as a low pass filter to remove high-energy bremsstrahlung photons above a given energy threshold. With this low-pass filter, high-anode voltages can be used to increase the intensity of the characteristic lines without increasing the high-energy background. An example of the effect of a monolithic optic designed to pass 8-keV Cu K_α radiation is shown in Fig. 19.³⁶ The lens reduces the Cu K_β 9 keV peak and suppresses the high-energy bremsstrahlung. For low-resolution diffraction applications, the energy filtration provided by the optic allowed the monochromator to be replaced with a simple filter. Albertini has proposed using polycapillary optics as a low-pass-filter for energy-dispersive x-ray diffractometry and reflectometry.⁴⁰

X-Ray Fluorescence

Polycapillary optics are widely used in x-ray fluorescence (XRF). Details of x-ray fluorescence analysis are described in Chap. 29. Focusing the beam from a divergent source can result in a large intensity increase compared to pin hole collimation. This allows for flexible system development. Formica designed a system for *in situ* thin-film deposition.⁴¹ Nikitina,⁴² Buzanich,⁴³ and Vittiglio⁴⁴

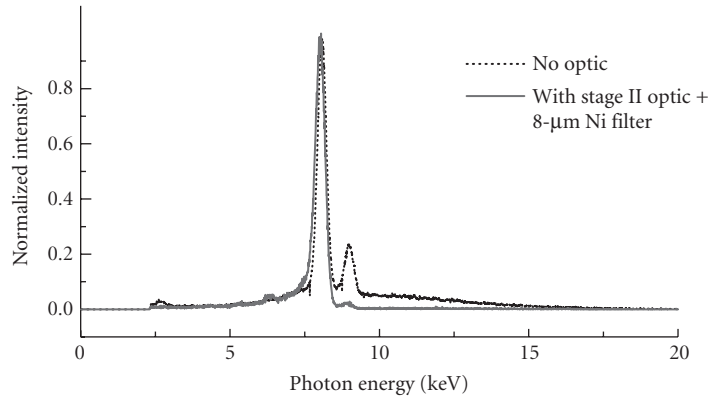


FIGURE 19 Spectrum of a copper tube source with and without a slightly focusing optic. The optic suppresses the high-energy bremsstrahlung. The nickel filter suppresses the $K\beta$ peak. (From Ref. 36.)

have demonstrated portable systems for materials and archaeometric analysis. Luo has developed a system for combinatorial materials studies.⁴⁵ Feldkamp has developed a system consistent with microtomography.⁴⁶ Kanngiesser provided detailed analysis of the beam shape and spectrum from focusing polycapillary optics to allow for accurate quantitative analysis.⁴⁷ Langer showed that polycapillary focussing makes possible the analysis of Kossel patterns from the fluorescent excitation using a tube source.⁴⁸ Instead of using the focussing optic on the excitation side, Smit uses a focusing optic to collect the fluorescence radiation in a synchrotron beam system,⁴⁹ Zitnik in a proton beam system,⁵⁰ and Alberi for particle induced x-ray emission (PIXE).⁵¹ With a conventional x-ray source, two optics can be used to both excite and collect the fluorescent radiation, as shown in Fig. 20.⁵² This provides the double benefit of enhanced signal intensity and three-dimensional spatial resolution. The three-dimensional resolution arises from the overlap of the cone of irradiation of the first lens and the cone of collection of the second. Sun uses a similar arrangement for energy-dispersive micro-x-ray diffraction.⁵³ Alternatively, Yang uses a collimating optic to increase the excitation intensity for total reflection x-ray fluorescence,⁵⁴ and Sun uses a reversed collimating optic in combination with a torroidal mirror for synchrotron radiation EXAFS.⁵⁵

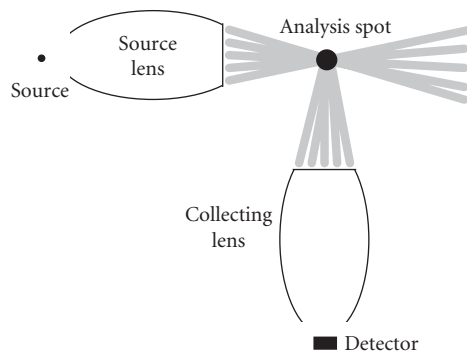


FIGURE 20 Sketch of microfluorescence experiment, showing that overlap of irradiation and collection volumes yields three-dimensional spatial resolution. (From Ref. 2.)

Single Crystal Diffraction

Collimating Significant reduction in data collection times for single crystal diffraction can be achieved with collimating polycapillary optics. The divergence from the optic, for example, 0.19° at 8 keV, is less than the ω crystal oscillations typically employed to increase the density of reflections captured in a single image in protein crystallography.⁵⁶ Using a low-power x-ray source which allows close access to the beam spot, the x-ray intensity obtained with a 20 W x-ray source was comparable to that achievable with a 3.5-kW rotating anode source with pinhole collimation. Chicken egg white lysozyme data taken with a 20-W source and a collimating lens in 20 minutes per frame produced an R -factor (variance between the measured and model structure factors) of 5.1 percent and resolution of 1.6 Å, as good as equivalent rotating anode data taken with the same or longer exposure time. The data shown in Fig. 21 was taken in 10 minutes with the collimating optic and 20-W source. Gubarev reported high-quality data extending to 1.7 Å with a 47-W system with greater flux than for a rotating anode using graded multilayer mirrors.⁵⁷

Focused For focused beam diffraction, the volume of reciprocal space that is accessed in a single measurement is greatly increased compared to parallel beam geometries. Additionally, the small irradiation spot on the sample reduces angular broadening due to sample size effects and allows the detector to be placed close enough to the sample that a small imaging detector will intercept diffracted beams at high 2θ angles. The reduction in angular broadening due to the small source size is significant because the diffraction spot is not isotropically broadened by the convergence of the focused beam.⁵⁸

Figure 22 displays a sketch of the diffraction condition for a single crystal with a monochromatic convergent beam. Diffraction conditions are satisfied for the two incident beam directions, $\mathbf{k}_0$ and $\mathbf{k}_1$, when they make the same angle with the reciprocal lattice vector, $\mathbf{G}$. Thus, changing from $\mathbf{k}_0$ to $\mathbf{k}_1$ rotates the diffraction triangle of $\mathbf{k}_0$, $\mathbf{G}$, and $\mathbf{k}_1$ about the vector $\mathbf{G}$ by an angle ϕ . This results in the diffracted beam $\mathbf{k}_1$ moving to trace out a tangential line on the detector. The resultant streak, shown in Fig. 23 has length

$$x = 2z \tan\left(\frac{\phi}{2}\right) = 2 \frac{D}{\cos(2\theta)} \tan\left(\frac{\phi}{2}\right) \quad (1)$$

The maximum value of ϕ is the convergence angle. There is no broadening in the radial direction. The “straight through” beam shown as the vector D will vary within the range of incoming beam directions for different reciprocal lattice vectors.

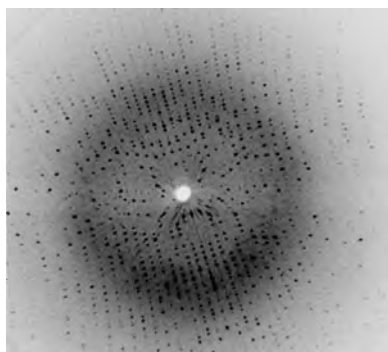


FIGURE 21 Lysozyme diffraction image taken with a 20-W microfocus source and collimating polycapillary optic in 10 minutes. (From Ref. 56.)

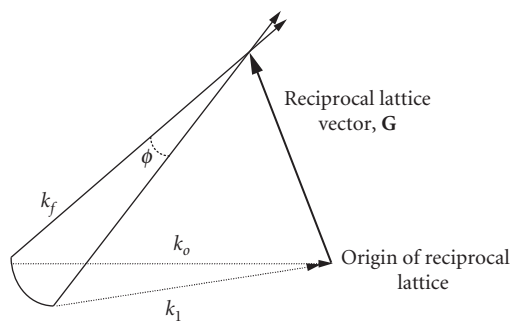


FIGURE 22 Ewald sphere description of focused beam diffraction on a single crystal.

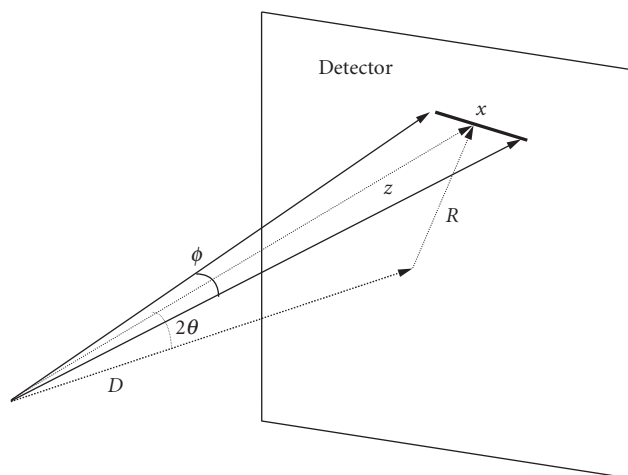


FIGURE 23 Diffraction streak due to beam focusing.

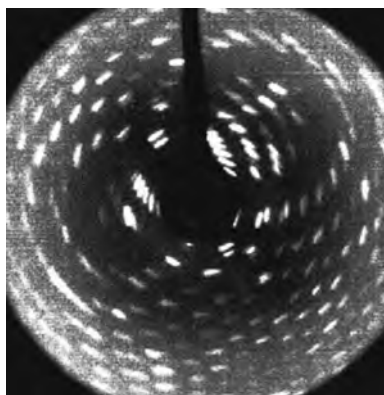


FIGURE 24 Lysozyme diffraction image taken in 5 min with a 1-mA tube source and a focusing polycapillary optic. (From Ref. 14.)

The effects of the one-dimensional streaking are shown in Fig. 24 for a single crystal egg-white lysozyme diffraction pattern taken with a 2.1° focusing angle.¹⁴ Serious overlap problems were not encountered except in low index directions, which are of less interest for structure determination. However, so long as the streaked diffraction spots are narrow compared to their separation, they can, in principle, be analyzed. For protein crystals with unit cell dimensions $< 200 \text{ \AA}$, such as lysozyme, this does not preclude structural determinations. However, for cell dimensions $> 200 \text{ \AA}$, the diffraction spots are not completely separated. Patterns with smaller convergence angles can be analyzed with conventional software and give good results.^{56,59} Patterns from crystals with smaller, less complex, unit cells will have lower-diffraction spot density and therefore less potential overlap than for protein crystals.

A direct comparison was made of data quality and collection time on a rotating anode system using a single egg-white lysozyme crystal, with and without polycapillary optics.⁵⁶ The direct beam intensity gain and the diffracted beam signal gain were both a factor of 20 for a lens with a 0.3° convergence. There was no degradation in data quality. Li reported intensity and resolution improvement with the use of a slightly focusing polycapillary optic for protein crystallography compared to double focusing mirrors.⁶⁰

Powder Diffraction

Collimating Reductions in data collection time can also be obtained for powder diffraction with collimating polycapillary optics. In addition, the parallel beam geometry provides insensitivity to sample preparation, shape, position, and transparency. Clapp reported significant reduction in sensitivity to sample preparation.⁶¹ Chen reported reduction in sensitivity to sample shifts and acceptability for whole pattern fitting.⁶² Secondly, the symmetric beam profile and enhanced flux gives much improved particle statistics and measurement statistics.

The nearly Gaussian peaks produced by the polycapillary collimating optics give more precise peak localization. For example, a pinhole with a divergence of 0.2 mrad produces peak localization errors approximately equal to the divergence, while the system with the polycapillary optic on the same sample produced peak localization that was detector limited to 0.04° .⁶³ Misture reported a significant intensity gain, reduction in peak shape and peak localization errors (also see Chap. 28).⁶⁴ The constant peak width and resolution throughout the diffraction space facilitates very high precision residual stress and texture analysis and reciprocal space mapping. The peak shape is ideally suited to phase identification and full pattern analysis of phase content using wavelet transforms.

In addition, polycapillary optics can be used as high-resolution soller slits. A schematic for an experiment comparing conventional and polycapillary soller slits is shown in Fig. 25. Replacing the soller slit with a polycapillary straight angular filter provided approximately a factor of two gain in signal-to-noise ratio for the identification of the ξ phase on a galvanized steel sheet.⁶⁵ Leoni performed detailed analysis of the beam shape and correction algorithms for a system similar to that of Fig. 25.⁶⁶

Focusing For focusing polycapillary optics, the peak resolution for powder diffraction continues to be much smaller than the beam divergence, even for highly convergent beams, and agrees well with a simple geometrical model.^{63,67}

Scatter Rejection in Imaging

The earliest announcements of the discovery of x rays included a radiograph of a human hand.⁶⁸ One hundred years after its first use, simple radiography is still the most common medical imaging modality. The relative simplicity of the required apparatus and the speed with which the image is

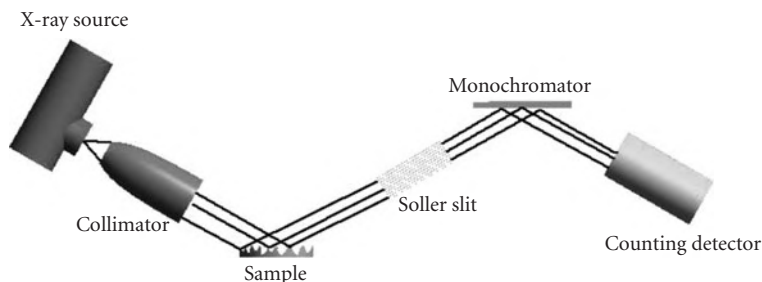


FIGURE 25 Schematic of diffraction setup showing polycapillary collimator and normal soller slit configuration. (From Ref. 65.)

obtained (with the resultant low cost) make x-ray imaging an extremely important diagnostic tool. However, Compton scattering can result in substantial image degradation. In a conventional medical imaging system, scatter is partially removed by inserting a grid with lead ribbons parallel to the incoming beam. Alternatively, scatter can be removed by inserting a polycapillary optic between the patient and the detector. Because capillary optics have an angular acceptance that is limited by the very small critical angle, scattered photons are not transported down optics channels, but are largely absorbed by the glass walls of the capillary optic. The scatter transmission of a polycapillary optic, measured by moving the source off axis to an angle much larger than the critical angle, was less than 1 percent at 20 keV,⁹ resulting in a contrast enhancement for a Lucite phantom of a factor of 1.4 compared to a conventional antiscatter grid, with a maximum enhancement of more than a factor of 2 in regions of low contrast or high scatter.¹¹ Measurements at 40 keV showed a contrast enhancement of more than a factor of 4.⁶⁹ Lead glass optics performed well at higher energies.⁷⁰

While not true imaging optics, polycapillary fibers can transmit an image in the same manner as a coherent fiber bundle. Polycapillary optics can be used to magnify and demagnify images. Magnification can improve image resolution, particularly if detector limited. Magnification is conventionally performed by increasing the air gap between the patient and the detector, but this results in a loss of resolution due to the geometrical blurring from the finite source size, as shown in Fig. 26.

However, magnification using polycapillary optics, even with a large focal spot, would not be subject to a loss in resolution, as shown in Fig. 26. Measured modulation transfer functions (MTF) for computed radiography image plate detectors using a magnifying polycapillary optic are compared to air gap magnification in Fig. 27.¹¹ The limiting MTF was increased by the polycapillary magnification factor of 1.8. The resolution was not degraded by the capillary structure, which was on a smaller scale (20- μm channel size) than the desired resolution. The optic yields an MTF increase at all spatial frequencies, which may be diagnostically more significant than the increase in limiting MTF. Magnifying capillary optics provide simultaneous contrast enhancement and resolution increase. Efforts have been made to increase the size of the optics. A transmission image of a multioptic jig is shown in Fig. 28.⁷¹

A technical difficulty with many direct x-ray detectors is that they are difficult to manufacture with large enough areas for medical imaging. With the use of x-ray optics, demagnification could be used to match the image size to a CCD, CID,⁷² or other digital detectors. A common digital imaging modality is

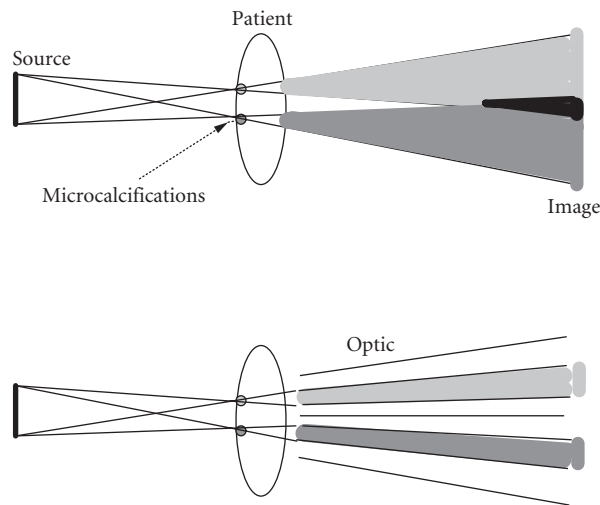


FIGURE 26 Air gap magnification (top) showing degradation of image due to finite source size. Magnification with a long tapered polycapillary optic (bottom) showing no increased blurring after the exit plane of the patient.

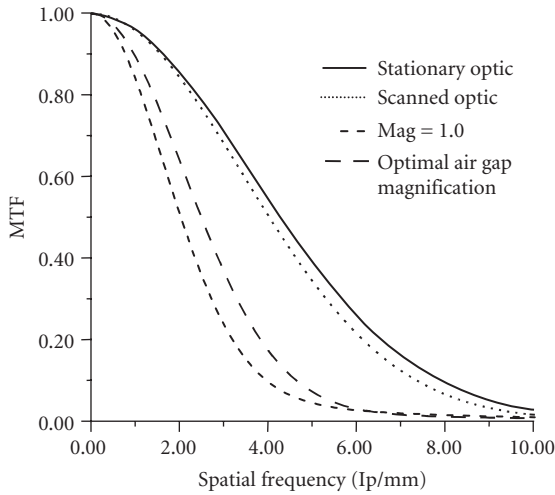


FIGURE 27 Modulation transfer function of a computed radiography plate with no optic, with airgap magnification, and with a polycapillary optic. (From Ref. 11.)

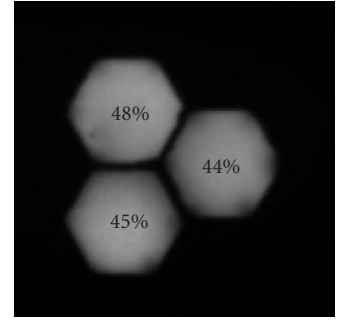


FIGURE 28 Transmission image of triad of monolithic optics. (From Ref. 71.)

the use of a fused fiber optic following a phosphor screen to demagnify the resulting visible light image for recording with CCD technology. The use of capillary x-ray optics to demagnify the x-ray image in connection with direct x-ray sensitive detectors may be more efficient and would avoid the loss of resolution in the optical conversion process.

Monochromatic Imaging

In conventional radiography, subject contrast arises from relatively low differences in absorption coefficients between different tissue types. The already low subject contrast is further reduced in a conventional system by averaging over relatively large-energy bandwidths. Synchrotron measurements using monochromatic beams have demonstrated higher contrast, but synchrotrons are not clinically available. Using monochromator crystals with a conventional source without an optic is not practical because the low intensity of the diffracted beam will not allow imaging *in vivo* before motion blur occurs. Polycapillary collimating optics can allow sufficient diffracted beam intensity to make clinical monochromatic imaging possible without a synchrotron.¹⁵ Measurements at 17.5 keV showed subject contrast enhancement of a factor of 2, in agreement with theoretical calculations. This contrast enhancement is in addition to that expected from the reduction of scattered radiation.

The output divergence of the collimating optic affects the resolution and is an important parameter, especially for low-energy and high-resolution modalities. Good angular resolution was achieved even with a large spot source. Resolution was measured by recording a knife-edge shadow with a restimable phosphor-computed-radiography image plate. For a silicon crystal, the width results primarily from the energy spread of the incident radiation. The image in Fig. 29 shows the resolved Mo $K\alpha$ energy doublet.³⁴ The divergence causes the fiber structure to blur and the field to become more uniform as the beam propagates away from the optic, as shown in Fig. 30.¹⁵

In addition to the monochromatization, the use of a uniform parallel beam would eliminate the variations in resolution and magnification of objects on the entry and exit side of the patient. A parallel, tunable, monochromatic source would facilitate the use of a number of innovative imaging techniques such as refraction contrast,⁷³ K-edge tuning,⁷⁴ dual energy,⁷⁵ and phase imaging. Refraction contrast is observed for highly monochromatic beams when a second crystal is placed after the specimen. Gradients

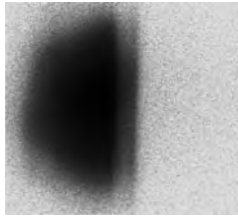


FIGURE 29 Knife-edge image. The extra shadow is due to the $K\alpha$ doublet. The image plate is 550 mm from the knife-edge.

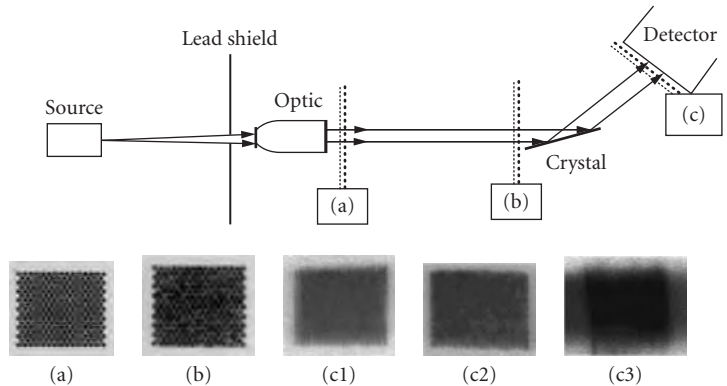


FIGURE 30 X-ray images taken at the three indicated locations along the beam. The fiber structure is clearly visible in (a) and blurs as the beam propagates. Image (c1) is for silicon, (c2) for mica, and (c3) for graphite.

in the index of refraction deflect the beam at the edges of features. The deflected beams are not diffracted by the second crystal, resulting in greatly enhanced contrast in the vicinity of feature edges. K-edge tuning is the selection of a photon energy just above the absorption edge for an element that is preferentially concentrated in, e.g., cancerous tissues. Dual-energy imaging uses the tunability of the source to make a subtraction image taken at energies just above and below the critical edge. Monochromatization with the Bragg angle at 45° produces a polarized beam, as shown in Fig. 31.⁷⁶

Scintigraphy

Unlike external beam radiography, collimators are required for nuclear imaging (see Chap. 32). In addition, detectors with energy discrimination are required because only a small fraction of the photons from the radioactive source are transmitted through the collimator. The majority of the radiation is scattered by the surrounding tissue, to create a broad source with approximately the same total

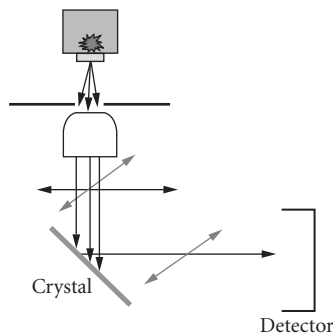


FIGURE 31 Set up for polarization by diffraction at 90° . After diffraction by the crystal, the beam is polarized in the out-of-plane direction.

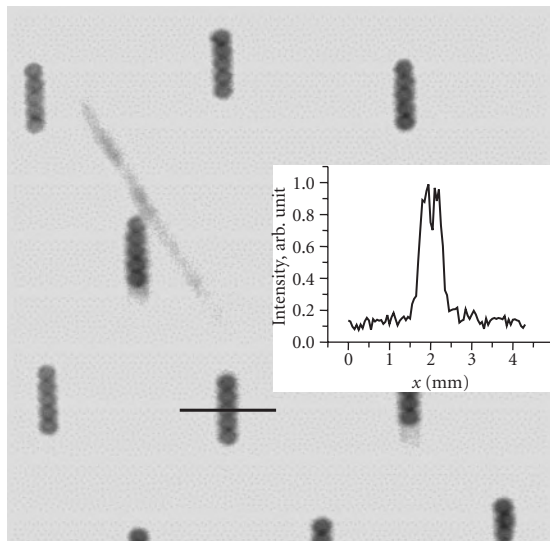


FIGURE 32 Images taken with set of brachytherapy seeds show five beads per seed and a ring-like structure. The source to optic distance was $z = 18$ mm.

intensity. Since Compton scattering causes the photons to lose energy, energy discrimination can be used to distinguish between rays which have traveled directly to the detector from those which have been scattered by the tissue. However, energy-sensitive detectors tend to have relatively poor spatial resolution compared to radiographic detectors. Testing has been performed of a different mechanism to discriminate against scatter and provide high resolution. In this study, a polycapillary optic was used as a high resolution collimator, paired with a radiographic detector with a large number of pixels.^{16–19}

Images taken with a parallel 20-mm-thick, 30-mm-wide lead glass polycapillary parallel hole collimator and a computed radiography plate demonstrated the ability to image features of individual ^{125}I ion exchange beads within brachytherapy seeds even in the presence of 30 mm of tissue-equivalent scatter material, as shown in Fig. 32. Calculations using simple geometrical models were in good agreement with measured signal-to-background ratios and count rates. The calculations and measurements showed that for highly heterogeneous radiation distributions high signal-to-background ratios could be achieved by using the high resolution to “dilute” the scatter background without energy discrimination. In addition, the high resolution collimator/radiographic detector system is significantly more compact than typical γ camera systems. Because energy sensitivity is no longer required, a wide range of radiographic detectors with high-resolution could be employed. The polycapillary collimator could be either parallel or converging.

Therapy

Conventional x-ray radiation therapy is currently performed with high energy x-ray or gamma radiation. The patient is exposed to multiple parallel beams, produced with slit collimators, with an intersection at the tumor. High-energy photons are chosen to minimize the absorbed skin dose relative to the dose at the tumor, although energies as low as 100 keV are employed in orthovoltage modalities. The choice of high energies to reduce skin dose is necessary because, in an unfocused beam, the intensity is necessarily higher near the point of entry than at the tumor site. Use of focusing optics at energies less than 100 keV might substantially increase the tumor dose relative to the skin dose.

Neutron beams can also be focussed by polycapillary optics (see Chaps. 63 and 64).^{77–79} A focused neutron beam could be used in boron neutron capture therapy (BNCT), based on the selective delivery of a boronated pharmaceutical to cancerous tissue followed by irradiation with thermal neutrons.⁸⁰ Perrung⁸¹ has described a BNCT system using capillary optics. Mayer has performed an experimental study to determine the radiation dose introduced by focusing a cold neutron beam through a tissue equivalent mass.⁸² With the use of radiochromic images, it was found that neutron scattering within the material did not significantly alter the focal image. As a result, the deposition of dose peaks quickly at the depth of the focal point. This procedure could be useful in treating near-surface regions such as ocular melanomas.

53.8 SUMMARY

Polycapillary x-ray optics are a powerful control technology for x-ray beams. Using polycapillary optics to collimate the output from a point source provides much higher intensity than pinhole collimation, particularly if two dimensional collimation is required. Focusing the beam yields even higher intensity gains.

Polycapillary optics can also be used to perform low-pass spectral filtering, which allows the use of increased source voltage. Further, the optics also remove the connection between source size and resolution, which allows the use of increased source current. Pairing a polycapillary optic with a diffracting crystal results in efficient production of monochromatic beams for imaging and analysis.

Tapered or straight polycapillary optics can also be used as angular filters, to replace soller slits in diffraction systems, to reduce scatter fraction in imaging thick objects, and to provide resolution in microscintigraphy.

53.9 ACKNOWLEDGMENTS

The authors are grateful for useful discussions and data from a large number of collaborators, including Carmen Abreu, David Aloisi, Simon Bates, Ayhan Bingolbali, David Bittel, Cari, Dan Carter, Heather Chen, Patrick Conlon, Greg Downing, Ning Gao, David Gibson, Mikhail Gubarev, Joseph Ho, Frank Hoffman, Huapeng Huang, Huimin Hu, Abrar Hussein, Chris Jezewski, Kardiawarman, John Kimball, Ira Klotzko, David Kruger, Danhong Li, Dip Mahato, Kevin Matney, David Mildner, Johanna Mitchell, Charles Mistretta, Robin Moresi, Noor Mail, Scott Owens Rohrbach, Wally Peppler, Sushil Padiyar, Igor Ponomarev, Bimal Rath, Christine Russell, Robert Schmitz, Francisca Sugiro, Suparmi, Christi Trufus-Feinberg, Johannes Ullrich, Hui Wang, Lei Wang, Russel Youngman, Brian York, Qi Fan Xiao, and Wei Zhou and for grant support from the Department of Commerce, NASA, NIH, and the Breast Cancer Research Program.

53.10 REFERENCES

1. V. A. Arkadev, M. A. Kumakhov, et al., *Sov. Physics. Usp.* **32**:3, March 1989.
2. C. A. MacDonald, "Applications and Measurements of Polycapillary X-Ray Optics," *J. X-Ray Science and Tech.* **6**:32–47, 1996.
3. A. Bjeoumikhov, S. Bjeoumikhova, H. Riesemeier, M. Radtke, and R. Wedell, "Propagation of Synchrotron Radiation through Nanocapillary Structures," *Phys. Lett. A* **366**(4–5):283–288, July 2, 2007.
4. V. A. Arkd'ev, A. I. Kolomitsev, M. A. Kumakhov, I.Yu. Ponorave, I. A. Khodееv, Yu. P. Chertov, and I. M. Shakparonov, "Wide-Band X-Ray Optics with a Large Angular Aperture," *Sov. Phys. Usp.* **32**(3):271, March 1989.

5. I. L. Klotzko, Q. F. Xiao, D. M. Gibson, R. G. Downing, W. M. Gibson, Karnaukhov A., and C. J. Jezewski, "Investigation of Glass Polycapillary Collimator for Use in Proximity Based X-Ray Lithography," *Proc. SPIE* **2523**:175–182, 1995.
6. C. H. Russell, W. M. Gibson, M. V. Gubarev, F. A. Hofmann, M. K. Joy, C. A. MacDonald, L. Wang, Qi-Fan Xiao, and R. Youngman, "Application of Polycapillary Optics for Hard X-Ray Astronomy," in *Grazing Incidence and Multilayer X-Ray Optical Systems*, R. B. Hoover and A. B. C. Walker II, eds., *Proc. SPIE* **3113**:369–377, 1997.
7. C. H. Russell, M. Gubarev, J. Kolodziejczak, M. Joy, C. A. MacDonald, and W. M. Gibson, "Polycapillary X-Ray Optics for X-Ray Astronomy," in *Advances in X-Ray Analysis*, 43, *Proc. 48th Denver X-Ray Conference*, 1999.
8. C. C. Abreu and C. A. MacDonald, "Beam Collimation, Focusing, Filtering, and Imaging with Polycapillary X-Ray and Neutron Optics," *Phys. Med.* **XIII**(3):79–89, 1997.
9. C. C. Abreu, D. G. Kruger, C. A. MacDonald, C. A. Mistretta, W. W. Peppler, and Q. F. Xiao, "Measurements of Capillary X-Ray Optics with Potential for Use in Mammographic Imaging," *Med. Phys.* **22**(11)(Pt. 1):1793–1801, November 1995.
10. W. M. Gibson, C. A. MacDonald, and M. S. Kumakhov, "The Kumakhov Lens: A New X-Ray and Neutron Optics with Potential for Medical Applications," in *Technology Requirements for Biomedical Imaging*, S. K. Mun, ed., I.E.E.E. Press, New York, #2580, 1991.
11. D. G. Kruger, C. C. Abreu, E. G. Hendee, A. Kocharian, W. W. Peppler, C. A. Mistretta, and C. A. MacDonald, "Imaging Characteristics of X-Ray Capillary Optics in Mammography," *Med. Phys.* **23**(2):187–196, February 1996.
12. C. A. MacDonald, C. C. Abreu, S. Budkov, et al., "Quantitative Measurements of the Performance of Capillary X-Ray Optics," in *Multilayer and Grazing Incidence X-Ray/EUV Optics II*, R. B. Hoover and A. Walker, eds. *Proc. SPIE*, SPIE, Bellingham, Wash., 2011, 1993.
13. W. M. Gibson and C. A. MacDonald, "Polycapillary Kumakhov Optics: A Status Report," in *X-Ray and UV Detectors*, *Proc. SPIE*, Bellingham, Wash., 2278, 1994.
14. S. M. Owens, F. A. Hoffman, C. A. MacDonald, and W. M. Gibson, "Microdiffraction Using Collimating and Convergent Beam Polycapillary Optics," *Advances in X-Ray Analysis*, vol. 41, *Proc. 46th Annual Denver X-Ray Conference*, Steamboat Springs, Colorado, August 4–8, 1997.
15. F. R. Sugiro, D. Li, and C. A. MacDonald, "Beam Collimation with Polycapillary X-Ray Optics for High Contrast High Resolution Monochromatic Imaging," *Med. Phys.* **31**:3288, 2004.
16. C. A. MacDonald, N. Mail, W. M. Gibson, S. M. Jorgensen, and E. L. Ritman, "Micro Gamma Camera Optics with High Sensitivity and Resolution," in M. J. Flynn, ed., *Physics of Medical Imaging*, *Proc. SPIE* **5745**:1–6, 2005.
17. S. M. Jorgensen, M. S. Chmelik, D. R. Eaker, C. A. MacDonald, and E. L. A. Ritman, "Polycapillary X-Ray Optics-Based Integrated Micro-SPECT/CT Scanner," in *Developments in X-Ray Tomography IV*, U. Bonse, ed., *Proc. SPIE* **5535**:36–42, 2004.
18. W. M. Gibson, C. A. MacDonald, and N. Mail, "Potential for Radioscinotography with Polycapillary Optics," in A. M. Khounsary, C. A. MacDonald, eds., *Advances in Laboratory-Based X-Ray Sources and Optics III*, *Proc. SPIE* **4781**:104–111, 2002.
19. N. Mail, C. MacDonald, and W. M. Gibson, "Microscintigraphy with High-Resolution Collimators and Radiographic Imaging Detectors," *Med. Phys.* **39**(2):645–655, 2009.
20. A. Liu, "The X-Ray Distribution after a Focussing Polycapillary—A Shadow Simulation," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **243**(1):223–226, January, 2006.
21. D. Hampai, G. Cappuccio, G. Cibin, S. B. Dabagov, and V. Sessa, "Modeling of X-Ray Transport through Polycapillary Optics," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *Proc. 10th International Symposium on Radiation Physics—ISRP* **10**, **580**(1):85–89, September 21, 2007.
22. D. Bittel and J. Kimball, *J. Appl. Phys.* **74**(2):877–883, 1993.
23. J. Harvey, chap. 11, In *Handbook of Optics*, Volume II, M. Bass, ed., McGraw-Hill, New York, 1996.
24. H. Wang, W. Lei, W. M. Gibson, and C. A. MacDonald, "Simulation Study of Polycapillary X-Ray Optics," in *X-Ray Optics, Instruments, and Missions*, R. B. Hoover and A. B. C. Walker II, ed., *Proc. SPIE* **3444**:643–651, 1998.
25. L. Wang, B. K. Rath, W. M. Gibson, J. C. Kimball, and C. A. MacDonald, *J. Appl. Phys.* **80**(7):3628–3638, 1996.

26. J. B. Ullrich, V. Kovantsev, and C. A. MacDonald, *J. Appl. Phys.* **74**(10):5933–5939, 1993.
27. Suparmi, Cari, W. Lei, H. Wang, W. M. Gibson, and C. A. MacDonald, *J. Appl. Phys.* **90**:5363–5368, 2001.
28. C. A. MacDonald and W. M. Gibson, “Applications and Advances in Polycapillary Optics,” *X-Ray Spectrometry* **32** (3):258–268, 2003.
29. B. Rath, R. Youngman, and C. A. MacDonald, *Rev. Sci. Instrum.* **65**:3393–3398, 1994.
30. F. R. Sugiuro, S. D. Padiyar, and C. A. MacDonald, “Characterization of Pre- and Post- Patient X-Ray Polycapillary Optics for Mammographic Imaging” in C. A. MacDonald and A. M. Khounsary, eds., *Advances in Laboratory-Based X-Ray Sources and Optics, Proc. SPIE* **4144**:204, 215, 2000.
31. D. Bilderbeck and E. Fontes, *AIP Conference Proceedings* **417**:147–155, 1997.
32. Cari, C. A. MacDonald, W. M. Gibson, C. D. Alexander, M. K. Joy, C. H. Russell, and Z. W. Chen, “Characterization of a Long Focal Length Polycapillary Optic for High Energy X Rays,” in *Advances in Laboratory-Based X-Ray Sources and Optics*, C.A. MacDonald and Ali M. Khounsary, eds., *Proc. SPIE* **4144**:183–192, 2000.
33. B. K. Rath, W. M. Gibson, L. Wang, B. E. Homan, and C. A. MacDonald, “Measurement and Analysis of Radiation Effects in Polycapillary X-Ray Optics,” *J. Appl. Phys.* **83**(12):7424–7435, June 15, 1998.
34. D. Li, F. R. Sugiuro, and C. A. MacDonald, “Source-Optic Optimization for Compact Monochromatic Imaging”, in *X-Ray Sources and Optics*, C. A. MacDonald, A. T. Macrander, T. Ishikawa, C. Morawe, J. L. Wood, eds., *Proc. SPIE* **5537**:105–114, 2004.
35. D. Li, N. Mail and C. A. MacDonald, “A Comparison of Doubly Curved Crystal and Polycapillary Optics for Monochromatic Beam Production from a Clinical Source,” in M. J. Flynn, ed., *Physics of Medical Imaging, Proc. SPIE* **5745**:754–763, 2005.
36. S. M. Owens, J. B. Ullrich, I. Y. Ponomarev, D. C. Carter, R. C. Sisk, J. X. Ho, and W. M. Gibson, “Polycapillary X-Ray Optics for Macromolecular Crystallography,” in R. B. Hoover and F. P. Doty, eds., *Hard X-Ray/ Gamma-Ray and Neutron Optics, Sensors, and Applications, Proc. SPIE*, 2859, 1996.
37. W. M. Gibson and M. A. Kumakhov, *Yearbook of Science & Technology*. McGraw-Hill, New York, 488–490, 1993.
38. F. A. Hoffman, N. Gao, S. M. Owens, W. M. Gibson, C. A. MacDonald, and S. M. Lee, “Polycapillary Optics for In-Situ Process Diagnostics,” in *In Situ Process Diagnostics and Intelligent Materials Processing, Materials Research Society Proc.*, P. A. Rosenthal, W. M. Duncan, J. A. Woollam, eds., **502**:133–138, 1998.
39. F. A. Hofmann, C. A. Freinberg-Trufas, S. M. Owens, S. D. Padiyar, and C. A. MacDonald, “Focusing of Synchrotron Radiation with Polycapillary Optics,” *Beam Interactions with Materials and Atoms: Nuclear Instruments and Methods B*, **133**:145–150, 1997.
40. V. Rossi Albertini, B. Paci, A. Generosi, S. B. Dabagov, O. Mikhin, and M. A. Kumakhov, “On the Use of Polycapillary Structures to Improve Laboratory Energy-Dispersive X-Ray Diffractometry and Reflectometry,” *Spectrochimica Acta Part B: Atomic Spectroscopy* **62**(11):1203–1207, November 2007.
41. S. P. Formica and S. M. Lee, “X-Ray Fluorescence System for Thin Film Composition Analysis during Deposition,” *Thin Solid Films* **491**(1–2):71–77, November 22, 2005.
42. S. V. Nikitina, A. S. Shcherbakov, and N. S. Ibraimov, *Rev. Sci. Instrum.* **70**:2950 1999.
43. G. Buzanich, P. Wobrauschek, C. Strelai, A. Markowicz, D. Węgrzynek, E. Chinea-Cano, and S. Bamford, “A Portable Micro-X-Ray Fluorescence Spectrometer with Polycapillary Optics and Vacuum Chamber for Archaeometric and Other Applications,” *Spectrochimica Acta Part B: Atomic Spectroscopy* **62**(11):1252–1256, November 2007.
44. G. Vittiglio, S. Bichlmeier, P. Klinger, et al., “A Compact [mu]-XRF Spectrometer for (in situ) Analyses of Cultural Heritage and Forensic Materials,” *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **213**, 5th Topical Meeting on Industrial Radiation and Radioisotope Measurement Applications **213**:693–698, January 2004.
45. Z. Luo, B. Geng, J. Bao, C. Liu, W. Liu, C. Gao, Z. Liu, and X. Ding, “High-Throughput X-Ray Characterization System for Combinatorial Materials Studies,” *Rev. Sci. Instrum.* **76**:095105, 2005.
46. J. M. Feldkamp, C. G. Schroer, J. Patommel, B. Lengeler, T. F. Gunzler, M. Schweitzer, C. Stenzel, M. Dieckmann, and W. H. Schroeder, “Compact X-Ray Microtomography System for Element Mapping and Absorption Imaging,” *Rev. Sci. Instrum.* **78**:073702, 2007.
47. B. Kanngie[ss]er, N. Kemf, and W. Malzer, “Spectral and Lateral Resolved Characterisation of X-Ray Microbeams,” *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **198**(3–4):230–237, December 2002.

48. E. Langer, S. Dabritz, W. Hauße, and M. Haschke, "Advances in X-Ray Excitation of Kossel Patterns by a Focusing Polycapillary Lens," *Applied Surface Science, 13th Applied Surface Analysis Workshop—AOFA 13* **252**(1):240–244, September 30, 2005.
49. Z. Smit, K. Janssens, K. Proost, and I. Langus, "Confocal [mu]-XRF Depth Analysis of Paint Layers," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, Proc. Sixteenth International Conference on Ion Beam Analysis*, **219,220**:35–40, June 2004.
50. M. Zitnik, P. Pelicon, N. Grlj, A. G. Karydas, D. Sokaras, R. Schutz, and B. Kanngiesser, *Appl. Phys. Lett.* **93**:094104 (2008).
51. R. Alberti, A. Bjeoumikhov, N. Grassi, C. Guazzoni, T. Klatka, A. Longoni, and A. Quattrone, "Use of Silicon Drift Detectors for the Detection of Medium-Light Elements in PIXE," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, Accelerators in Applied Research and Technology - Proceedings of the 9th European Conference on Accelerators in Applied Research and Technology* **266**(10):2296–2300, May 2008.
52. W. M. Gibson and M. A. Kumakhov, in *X-Ray Detector Physics and Applications, Proc. SPIE* **1736**:172, 1992.
53. T. Sun, M. Zhang, X. Ding, Z. Liu, X. Lin, and H. Liu, "Characterization of a Polycapillary X-Ray Lens for Application in Confocal Three-Dimensional Energy-Dispersive Micro X-Ray Diffraction Experiments," *J. Appl. Cryst.* **40**:1169–1173, 2007.
54. J. Yang, D. Zhao, Q. Xu, and X. Ding, "Development, Application of Glancing Incident X-Ray Fluorescence Spectrometry Using Parallel Polycapillary X-Ray Lens," *Appl. Surface Science*, In Press, Accepted Manuscript, Available online October 14, 2008.
55. T. Sun, Y. Xie, Z. Liu, T. Liu, T. Hu, and X. Ding, "Application of a Combined System of Polycapillary X-Ray Lens and Toroidal Mirror in Micro-X-Ray-Absorption Fine-Structure Facility," *Appl. Phys.* **99**:094907, 2006.
56. F. A. Hofmann, W. M. Gibson, C. A. MacDonald, D. A. Carter, J. X. Ho, and J. R. Ruble, "Polycapillary Optic—Source Combinations for Protein Crystallography," *J. Appl. Crystallogr.* **34**:330–335, 2001.
57. M. Gubarev, E. Ciszak, I. Ponomarev, W. Gibson, and M. Joy, "A Compact X-Ray System for Macromolecular Crystallography," *Rev. Sci. Instrum.* **71**:3900, 2000.
58. C. A. MacDonald, S. M. Owens, and W. M. Gibson, "Polycapillary X-Ray Optics for Microdiffraction," *J. Appl. Crystallogr.* **32**:160–167, 1999.
59. J. X. Ho, E. H. Snell, C. R. Sisk, J. R. Ruble, D. C. Carter, S. M. Owens, and W. M. Gibson, "Stationary Crystal Diffraction with a Monochromatic Convergent X-Ray Beam Source and Application for Macromolecular Crystal Data Collection," *Acta Cryst.* **D54**:200–214, 1998.
60. P.-W. Li and R.-C. Bi, "Applications of Polycapillary X-Ray Optics in Protein Crystallography," *J. Appl. Cryst.* **31**:806–811, 1998.
61. R. A. Clapp and M. Haller, "Parallel Beam Methods in Powder Diffraction and Texture in the Laboratory," *Adv. X-Ray Anal.* **43**:135–140 (2000).
62. X. Chen, S. Bates, and K. R. Morris, "Quantifying Amorphous Content of Lactose Using Parallel Beam X-Ray Powder Diffraction and Whole Pattern Fitting," *J. Pharma. Biomed. Ana.* **26**(1):63–72, August 2001.
63. W. Zhou, D. N. Mahato, C. A. MacDonald, "Analysis of Powder X-Ray Diffraction Resolution Using Collimating and Focusing Polycapillary Optics," *Thin Solid Films*, accepted.
64. S. T. Mixture and M. Haller, "Application of Polycapillary Optics for Parallel Beam Powder Diffraction," *Adv. X-Ray Anal.* **43**:248–253, 2000.
65. W. M. Gibson, H. Huang, J. Nicolich, P. Klein, and C. A. MacDonald, "Polycapillary Optics for Angular Filtering of X Rays in Two Dimensions," in *Proc. 50th 2001 Denver X-Ray Conference, Advances in X-Ray Analysis* **45**:F-58, 2002.
66. M. Leoni, U. Welzel, and P. Scardi, "Polycapillary Optics for Materials Science Studies: Instrumental Effects and Their Correction," *J. Res. NIST* **109**(1):27–48, 2004.
67. A. Bingölbali, W. Zhou, D. N. Mahato, and C. A. MacDonald, "Focused Beam Powder Diffraction with Polycapillary and Curved Crystal Optics," in *Advances in X-Ray/EUV Optics and Components III*, Proc. SPIE, A. M. Khounsary, C. Morawe, S. Goto, eds., SPIE, Bellingham Wash., 7077, 2008.
68. W. C. Röntgen, "On a New Form of Radiation," *Nature*, **53**, 274–276, January 23, 1896, English Translation from *Sitzungsberichte der Würzburger Physik-med. Gesellschaft*, 1895.
69. Cari, Suparmi, W. M. Gibson, and C. A. MacDonald, "Contrast Enhancement Measurements Using Polycapillary X-Ray Optics at 20–40 keV," in L.E. Antonuk, M.J. Yaffe, eds., *Medical Imaging 2001:Physics of Medical Imaging, Proc. SPIE* **4320**:163–170, 2001.

70. Suparmi, Cari, Lei Wang, Hui Wang, W. M. Gibson, and C. A. MacDonald, "Measurement and Analysis of Lead glass Capillary Optic Performance for Hard X Ray Applications," *J. Appl. Phys.* **90**(10):5363–5368, 2001.
71. R. E. Ross, C. D. Bradford, and W. W. Pepler, "Optimization of X-Ray Capillary Optics for Mammography," *Med. Imaging*, 2002.
72. R. Wentink, J. Carbone, D. Aloisi, W. M. Gibson, C. A. MacDonald, Q. E. Hanley, R. E. Fields, and M. B. Denton, "Charge Injection Device (CID) Technology: An Imaging Solution for Photon and Particle Imaging Applications," *Proc. SPIE* **2279**, 1994.
73. V. A. Somenkov, A. K. Tklich, and S. S. Shilstein, "X-Ray Refraction Radiography of Biological Objects," *Zhurnal Tekhnicheskoy Fiziki* **11**:197, 1991.
74. F. E. Carroll, "Generation of Soft X-Rays by Using the Free Electron Laser as a Proposed Means of Diagnosing and Treating Breast Cancer," *Lasers Surg. Med.* **11**:72–78, 1991.
75. P. C. Johns, D. J. Drost, M. J. Yaffe, and A. Fenster, "Dual-energy Mammography: Initial Experimental Results," *Med. Phys.* **12**:297–304, May/June 1985.
76. R. Schmitz, A. Bingölbali, A. Hussain, and C. A. MacDonald, "Development of Polarized and Monochromatic X-Ray Beams from Tube Sources in Advances," in A. M. Khounsary, C. Morawe, S. Goto, eds., *X-Ray/EUV Optics and Components III*, *Proc. SPIE* **7077**, 2008.
77. H. Chen, R. G. Downing, D. F. R. Mildner, W. M. Gibson, M. A. Kumakhov, I. Yu Ponomarev, and M. V. Gubarev, "Guiding and Focusing Neutron Beams using Capillary Optics," *Nature* **357**:391, 1992.
78. Q. F. Xiao, H. Chen, V. A. Sharov, D. F. R. Mildner, R. G. Downing, N. Gao, and D. M. Gibson, "Neutron Focusing Optic for Submillimeter Materials Analysis," *Rev. Sci. Instrum.* **65**:3399–3402, 1994.
79. D. F. R. Mildner, H. H. Chen-Mayer, and R. G. Downing, "Characteristic of a Polycapillary Neutron Focusing Lens," *Proceed. of Intl. Symp. on Adv. in Neutron Optics and Related Research Facilities*, March 19–21 (1996), *J. Phys. Soc. Jpn.*
80. R. F. Barth, A. H. Soloway, and R. G. Fairchild, "Boron Neutron Capture Therapy for Cancer," *Sci. Am.* **263**(4):100–103, 1990.
81. A. J. Peurrung, "Capillary Optics for Neutron Capture Therapy," *Med. Phys. Apr.* **23**(4):487–494, 1996.
82. R. Mayer, J. Welsh, and H. Chen-Mayer, "Focused Neutron Beam Dose Deposition Profiles in Tissue Equivalent Materials: A Pilot Study of BNCT," presented at the *5th International Conference on Neutron Techniques*, Crete, Greece, June 9–15, 1996.

This page intentionally left blank

SUBPART

5.4

X-RAY SOURCES

This page intentionally left blank

X-RAY TUBE SOURCES

Susanne M. Lee

*GE Global Research
Nikayuna, New York*

Carolyn MacDonald

*University at Albany
Albany, New York*

54.1 INTRODUCTION

An x-ray measurement system includes a source, optics (at least pinhole optics or apertures), and a detector. To optimize the system, the requirements of the application and the properties of the available sources and detectors must be considered. This chapter discusses the properties of the most common type of source, x-ray tubes, and how they affect optics optimization. Even though tube sources have been around for more than 100 years, recent advances have been made in tube design and performance. These include the method by which the electrons are generated, submicron focal spot sizes, sources with multiple targets, and the ability to direct the x-ray beam to a desired location.

The most common x-ray tubes generate x rays by accelerating electrons toward a target anode. The electron beam impact on the target generates x rays, as shown in Fig. 1. The mechanisms for x-ray generation and the resulting spectra are discussed in Sec. 54.2. Electron generation in the cathode is discussed in Sec. 54.3. The design of the anode, considerations in the choice of anode material, and limitations in the electron spot size and shape on the anode are discussed in Sec. 54.4. General optimization considerations are discussed in Sec. 54.5.

X-ray tube sources can be unipolar or bipolar, depending on how the electron-accelerating potentials are distributed between the source anode and cathode. In unipolar sources, such as the one shown in Fig. 1, the anode can be grounded, and the accelerating potential created by keeping the cathode at a large negative voltage. Alternatively, the electron source can be grounded with the anode at positive voltage. One advantage of anode grounding is the elimination of high-voltage standoffs, allowing much smaller distances between the x-ray generation point on the anode and the x-ray source exit window. This is beneficial for many x-ray optics, since short input optic distances allow the optic to collect from large source solid angles. In bipolar circuits, the anode is usually held at half the full potential between the cathode and anode; the cathode is then held at the same negative potential. By splitting the potential, smaller high-voltage standoffs and lighter electrical cables can be used, which increase the portability of such sources. Since portability is important in many nondestructive testing applications, industrial x-ray tubes are frequently bipolar.

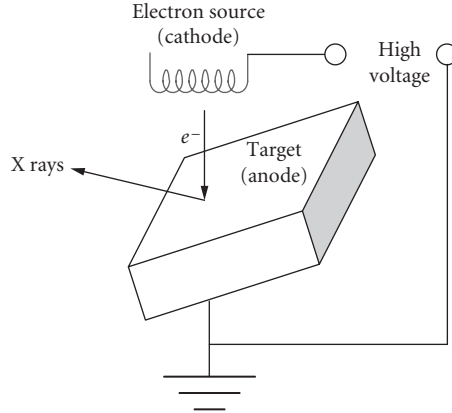


FIGURE 1 Schematic of a common unipolar anode-grounded electron impact source. Electrons emitted from the cathode, usually inside a vacuum vessel, are accelerated toward a metal anode by an accelerating voltage, which typically ranges from 10,000 to 500,000 V.

54.2 SPECTRA

Two basic processes convert the kinetic energy of the electrons into x rays, bremsstrahlung radiation, and characteristic emission.

Bremsstrahlung Continuous Radiation

The high-energy electrons emitted from the cathode impact the target and are decelerated by their interaction with the nuclei and electrons of the target material, converting the kinetic energy of the impacting electrons into electromagnetic energy. The x rays generated by this mechanism, known as braking radiation or bremsstrahlung, have a broad energy spectrum, with the high-energy cut-off determined by the maximum kinetic energy of the incident electrons, $K_{\max} = eV_{\text{tube}}$, where e is the electron charge and V_{tube} is the accelerating voltage. The efficiency of converting incident electron kinetic energy into continuum x-ray radiation energy is approximately¹

$$\eta = \frac{P_{\text{brems}}}{P_{\text{electrical}}} = \frac{1}{2} C Z_{\text{target}} (K_{\max})^{\alpha} \cong \frac{C}{2} Z_{\text{target}} (eV_{\text{tube}}) \quad (1)$$

where P_{brems} is the total bremsstrahlung radiation power, $P_{\text{electrical}}$ is the electrical power of the tube, $P_{\text{electrical}} = I_{\text{tube}} V_{\text{tube}}$, $\alpha \cong 1$, C is Kramer's constant, $C \cong 2.2 \times 10^{-9}/eV$, $K_{\max}$ is the maximum kinetic energy of the electrons, $K_{\max} \sim eV_{\text{tube}}$, and Z_{target} is the atomic number of the target material. The efficiency of bremsstrahlung production is low, even for a high Z material at high voltage. For example, for a tungsten target with an electron accelerating potential of 100 kV, only 0.8% of the impacting electron beam energy is converted into x rays. Most of the energy in the incident electron beam is converted to heat.

The total bremsstrahlung power from Eq. (1) is

$$P_{\text{brem}} = \eta P_{\text{electrical}} \cong \frac{C}{2e} Z_{\text{target}} (eV_{\text{tube}})^2 I_{\text{tube}} \quad (2)$$

where I_{tube} is the tube current. The emitted power depends roughly on the square of the electron-accelerating voltage.

For thin targets, the bremsstrahlung energy emission spectrum is usually regarded as flat, independent of x-ray energy, up to the maximum energy, where the intensity vanishes.² The differential bremsstrahlung energy intensity, $Ed\sigma$, between E and $E + dE$ is then a constant, i.e.,

$$Ed\sigma \sim (\text{const}) Z_{\text{target}} dE \quad (3)$$

where $d\sigma$ is the differential bremsstrahlung cross section. For thicker target materials, the impacting electrons undergo multiple scattering. The bremsstrahlung spectrum then for thicker targets becomes a superposition of thin-target bremsstrahlung spectra from thin layers at increasing depths, with successively lower cutoff energies due to the electron energy losses as the electrons penetrate into the material. At moderate tube voltages, up to approximately 100 kV, the shape of this spectrum, shown in Fig. 2, is roughly linear in x-ray energy.^{3,4}

While the x-ray spectrum produced by the electrons is roughly triangular for thick targets, if the target is thick enough and the electron penetration depth deep enough, the x rays can be absorbed by the target as they exit it, which results in a preferential decrease in the low energy x-ray intensity actually emitted by the tube. A similar loss in low energy photons is produced by intentional filtration and by absorption due to the source and detector windows and any air path the x rays traverse. As a result, while the Kramer approximation reproduces the general shape of the Bremsstrahlung from a real x-ray tube, it does not model the detailed spectrum well.

Accurate bremsstrahlung cross sections are difficult to calculate from first principles. A number of cross sections have been proposed with varying degrees of success (see Nakamori et al. for an excellent review of the literature).⁵ Birch and Marshall,⁶ proposed a more precise semi-empirical model where

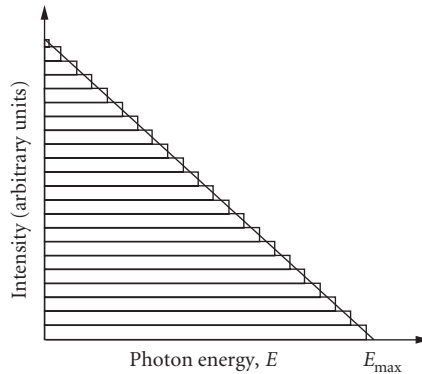


FIGURE 2 Triangular shape of bremsstrahlung intensity from a thick target, showing the superposition of flat spectra from thin-target bremsstrahlung with decreasing kinetic energy as the electrons penetrate into the material.

the bremsstrahlung cross section $d\sigma$ was modified to include a polynomial energy dependence and a relativistic correction factor,

$$\begin{aligned}
 Ed\sigma &= \rho N_A \left(1 + \frac{K_e}{m_0 c^2}\right) Q(E) dE \\
 &= \rho N_A \left(1 + \frac{K_e}{m_0 c^2}\right) \left[\left(\frac{Z_{\text{target}}}{K_e}\right)^2 \left[0.0503 - 0.94597 \left(\frac{E}{K_e}\right) + 0.1553 \left(\frac{E}{K_e}\right)^2 \right. \right. \\
 &\quad \left. \left. + 1.1632 \left(\frac{E}{K_e}\right)^3 - 0.6818 \left(\frac{E}{K_e}\right)^4 \right] \right] dE
 \end{aligned} \tag{4}$$

where N_A is Avagadro's number, ρ is the target material density, and $m_0 c^2$ is the electron rest mass energy. The polynomial coefficients were determined by fitting their calculated bremsstrahlung intensity to measured spectra.

Birch and Marshall also included x-ray attenuation in the target. In their model, the number of bremsstrahlung x rays having energy between E and $E + dE$ is given by⁶

$$N(E)dE = \frac{1}{A} \int_E^{eV_{\text{tube}}} d\sigma \left(\frac{1}{\rho} \frac{dK_e}{dx} \right)^{-1} \left[e^{-\frac{\mu(E)(eV_{\text{tube}}^2 - K_e^2) \cos \gamma}{\rho C_{\text{TW}} \sin \alpha}} \right] dK_e \tag{5}$$

where A is an empirical correction factor, $d\sigma$ is the differential bremsstrahlung cross section for x-ray emission with x-ray energy between E and $E + dE$, K_e is the electron kinetic energy, x is the electron depth into the target, $\mu(E)$ is the linear x-ray absorption coefficient in the target material at the x-ray energy E , γ is the angle the incident electron makes with the target, C_{TW} is the empirically determined Thomson-Whiddington constant, and α is the angle the exiting x ray makes with the target surface. The expression in brackets $[1/\rho(dK_e/dx)^{-1}]$ is the inverse of the electron stopping power of the target and tends to be a slowly varying function of the electron energy, such as $(a + b e^{-cK_e})$, where a , b , and c are estimated from experimental data, usually with a least squares fit. The exponential factor in Eq. (5) accounts for the absorption of generated x rays in the target material. Both the electron stopping power and x-ray absorption in the target have been tabulated as a function of electron energy.⁷ Note that Eq. (5) gives the number of photons per energy interval and must therefore be multiplied by the photon energy E and integrated to obtain the total bremsstrahlung power.

Characteristic Radiation

If the kinetic energy of the electron impacting the target exceeds the electron binding energy of the target material, bound target electrons can be ejected. Then either a free electron, or one bound in an upper shell of the target atom, can fill the vacant lower energy state, in the process emitting a photon. If the vacant lower energy state is an inner shell, the emitted photons will be in the x-ray regime for most common target materials. Because these emitted x rays have an energy that depends on the core electronic levels of the target material, they are known as characteristic x rays. The lines are enumerated according to the vacancy core level, with K levels having principal quantum number $n = 1$, L levels $n = 2$, and M levels $n = 3$. The α transitions have $\Delta n = 1$; the β transitions $\Delta n = 2$. The energy of the core atomic level with quantum number n is given by

$$E_n = -E_0 \frac{(Z - \zeta)^2}{n^2} \tag{6}$$

where E_0 is the Rydberg energy,

$$E_0 = \frac{me^4}{8\epsilon_0^2 h^2} = 13.6 \text{ eV} \quad (7)$$

Z is the atomic number of the anode material, and ζ is a screening constant that describes the reduction in effective nuclear charge experienced by outer shell electrons. The screening constant increases with n ; for K-level electrons, $\zeta \sim 3$; for L levels, $\zeta \sim 12$ for elements with moderate Z . For copper, with $Z = 29$, the energy of a K-shell electron is -9 keV. The $1/n^2$ dependence in Eq. (6) results in bunching of the upper levels so that the L energy is just -1 keV. When an electron from the L shell fills a vacancy in the K shell, a copper $K\alpha$ photon is emitted with energy 8.0 keV. The $K\beta$ line energy, which is due to filling the K shell vacancy with the slightly higher-energy M-shell electron, is 8.9 keV. The levels, and hence emission lines, are further split according to the total angular momentum quantum number J , resulting in doublet and multiplet lines.

Characteristic emission requires overcoming the inner-shell electron binding energy of the target atoms. (For low Z targets the binding energy of the electron in the solid is nearly equal to the ionization energy of a gaseous atom, although relaxation effects can be significant for tungsten.) No emission will be seen unless the tube voltage imparts to the impacting electron a kinetic energy greater than the electron binding energy in the target. For example, to produce the K-line emission from a copper target, the tube voltage must be larger than 9 kV. Above this minimum voltage $V_{\min}$ the characteristic photon emission rate increases with tube current I_{tube} and tube voltage V_{tube} roughly as

$$F = A \frac{I_{\text{tube}}}{e} \left(\frac{V_{\text{tube}}}{V_{\min}} - 1 \right)^p \quad (8)$$

where A is a constant determined by the target material, eV_{tube} is the maximum electron impact energy, and the exponent p ranges from 1.2 at high tube voltages to 1.7 at low tube voltages. At high tube voltages the electron penetration into the target is greater, leading to greater x-ray reabsorption in the target. The coefficient A gives the ratio of the number of characteristic photons emitted per incident electron when the tube voltage is twice the minimum voltage needed to excite the characteristic line, and is typically around 1 photon per 1000 to 6000 electrons. An example of a measurement and fit to a single value of A is shown in Fig. 3.⁸ This measurement was performed with a small pinhole to avoid flooding the detector. Measurements can also be made with filters of known

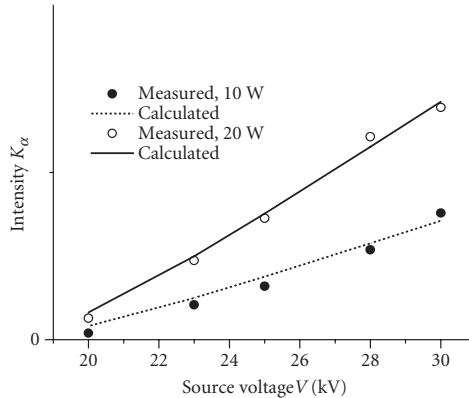


FIGURE 3 Intensity of Mo $K\alpha$ doublet versus anode voltage at 20 W ($\circ$) and at 10 W ($\bullet$). The solid and dashed lines are the calculated intensities.

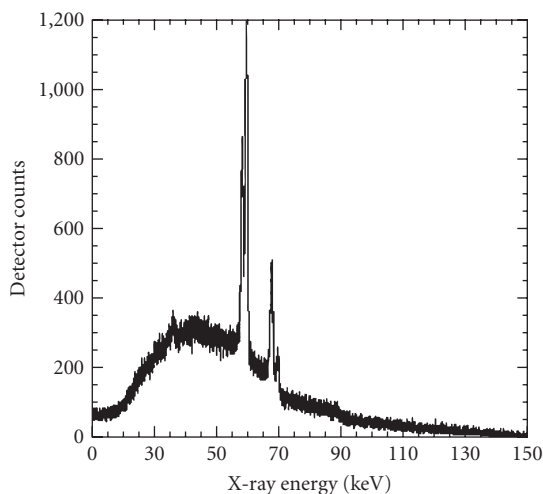


FIGURE 4 Source spectrum from a tungsten anode operated at 150 kV. The characteristic lines are superimposed on the bremsstrahlung, which has a maximum energy of 150 keV, according to Eq. (2). At low energies the spectrum differs significantly from the triangular approximation due to x-ray absorption in the target, air, windows of the x-ray tube, and the detector.

transmission. In either case, care must be taken so that errors in pinhole alignment, pinhole size, or filter thickness and composition do not affect the result, as the measured count rate must be orders of magnitude smaller than the source emission rate for most detectors.

The width of the characteristic emission lines is quite small; for Cu $K\alpha$ it is 4 eV and for Mo $K\alpha$ it is 8 eV.⁹ As a result, while the characteristic emission rate is typically similar in magnitude to the bremsstrahlung rate, the characteristic radiation is much brighter, and the diffracted intensity from, e.g., a narrow bandwidth monochromator, will be very much higher for the characteristic radiation.

Chapter 36 contains tables of the characteristic lines and corresponding energies for the elements commonly used in x-ray analysis. *The International Tables for X-Ray Crystallography*⁴ contain data on the efficiency with which various characteristic x-ray energies are produced as a function of incident electron beam energies. A typical spectrum from a tungsten anode, where both the characteristic lines and bremsstrahlung continuum can be seen, is shown in Fig. 4.

Spectral Selection

The Effect of Anode Material on Spectrum A consequence of the Z dependence in Eq. (8) is that increasing the atomic number of the anode slowly increases the emission energy of a particular emission line; e.g., the $K\alpha$ energy for W ($Z = 74$) is 59.3 keV compared to 8.0 keV for Cu ($Z = 29$). For that reason, for applications where monochromatic beams are desired, the anode material is chosen for the required x-ray energy. For example in mammography (described in Chap. 31), the anode material is chosen to provide a more intense beam in the 20 to 30 keV range, requiring selection of Mo, Rh, or Ag targets. In x-ray fluorescence (XRF) applications (described in Chap. 29), the anode material is chosen to be slightly higher in atomic number than the element to be sampled, since this will maximize the fluorescence emission intensity. In x-ray diffraction, when the sample is weakly diffracting, a Cu rather than Mo or W tube is frequently chosen because, at the lower energies, the cross section for diffraction is higher. In monochromatic x-ray diffraction, where an intense characteristic line is desired with a low intensity continuum, increasing the atomic number will have

a deleterious effect on the ratio of characteristic to continuum emission, since, according to Eq. (1), the continuum intensity will increase. Thus, the highest resolution x-ray diffraction measurements performed with laboratory sources use sources with lower atomic number targets.

In other applications where it is more important that a high total x-ray power is generated, such as white-beam diffraction or some imaging applications, the increase in bremsstrahlung continuum intensity with atomic number is a positive effect, and higher Z sources, like W, are common.

The Effect of Tube Voltage on Spectrum Many tubes use pulsed rather than continuous operation, and even in continuous operation the tube voltage may not be stable, so the tube voltage is usually specified in terms of peak voltage, kVp.

The tube voltage is also used to shape the spectrum. According to Eqs. (2) and (8), the total unfiltered bremsstrahlung power increases approximately as the square of the tube voltage, while the total characteristic intensity increases somewhat slower, especially for high accelerating voltages. The resultant ratio of the power in the characteristic line to the total bremsstrahlung power is highest at a tube voltage corresponding to

$$eV_{\text{tube}} = \frac{\alpha + 1}{\alpha + 1 - p} E_{\text{char}} \equiv \frac{2}{2 - p} E_{\text{char}} \quad (9)$$

For low voltages ($p \sim 1.7$), the electron energy eV_{tube} should be roughly six to seven times the characteristic line energy, or about 45 to 55 kVp for Cu $K\alpha$ radiation. For higher voltages ($p \sim 1.2$), the tube voltage need only be approximately 2.5 times the electron binding energy, or about 150 kVp for W $K\alpha$ radiation.

However, the ratio of the characteristic photon emission rate to the bremsstrahlung photon emission increases monotonically with tube voltage. The relative importance of the characteristic versus bremsstrahlung intensities depends on the experiment—in many imaging applications, the image reconstruction process is simplified considerably if the tube voltage does not excite the characteristic energies. In monochromatic x-ray diffraction applications, the characteristic lines are usually preferred; the monochromator, x-ray optics, and type of x-ray detector (photon counting or energy-integrating detector) can impact the characteristic to bremsstrahlung ratio. One consideration with x-ray optics is whether they will pass undesirable high-energy components, e.g., integral multiples of the characteristic energy, which may affect the choice of tube voltage. In addition, at a fixed current, increasing the tube voltage increases the power that must be dissipated in the anode, and thus generally the source size.

Filtering For some applications, such as low-resolution diffraction, a sufficiently monochromatic spectrum can be obtained by employing a simple absorption filter with a K-edge between the $K\alpha$ and $K\beta$ emission lines of the anode material. For example, Ni, with $Z = 28$ is used to absorb the Cu ($Z = 29$) $K\beta$ emission in many laboratory diffractometers. The filter also absorbs the low-energy bremsstrahlung, but the tube voltage should be set only slightly above the characteristic line voltage to avoid creating too much high-energy bremsstrahlung. Aluminum filters are also commonly used in medical imaging systems to remove the low-energy photons that would otherwise be absorbed in the patient as unnecessary dose.

Because the filter preferentially transmits high-energy photons, the bremsstrahlung power output rises more rapidly with tube voltage than the V^2 factor for thick targets of Eq. (2), to more than V^3 under heavy filtration.¹⁰

Multi-Energy Imaging Some materials have similar x-ray absorption properties for a limited range of energies, yet different properties for other energies. By imaging these objects with multiple spectral ranges, the materials can be differentiated from each other. This is known as multi-energy imaging and can be accomplished in several different ways. One way is to exploit the tube-voltage dependence of Eqs. (2) and (5) to produce bremsstrahlung spectra with different x-ray energy distributions. Operating the x-ray source at two different tube voltages will produce x-ray beams with bremsstrahlung spectra having different maximum intensity energies and different high-energy

cutoffs. This technique is known as kVp-switching. Unfortunately, the differences in tube voltage needed to provide clear image differences can be on the order of 100 kV, which is a large voltage difference to switch quickly and stably.

An alternative approach to kVp-switching is to have multiple target materials present in one x-ray source housing. The electron beam is directed toward each target material sequentially, creating images made with different energy spectra. This technique is beginning to see more use in imaging applications such as mammography, where it has been difficult to distinguish between tumors and the surrounding soft tissue due to similar absorption properties. The different energy images together with contrast agents allow other features, such as the vascularization of the tumors, to be seen more readily. One issue with using different target materials is the physical displacement of the targets from each other, which causes the x rays from each target to intersect the object to be imaged at different angles; the different viewing angles then have to be compensated for in the image processing algorithms unless the target material is moved between images.

54.3 CATHODE DESIGN AND GEOMETRY

Hot Filament Sources

Typically, a helical wire filament is heated in vacuum to slightly below the melting point of the filament material ($\sim 1000^\circ\text{C}$ or higher), allowing electrons with high kinetic energy to escape. A grid at a slightly positive potential with respect to the filament is usually placed near the filament to focus and accelerate the electrons in the direction of the target. Because of the high operating temperature of this filament type, substantial warm-up times, often on the order of a half hour, are required before stable filament current is achieved. In addition, the high temperature results in substantial evaporation from the filament, eroding and reducing the filament lifetime. As the filament wire diameter decreases, so does the electron density that can be extracted from the material. This has the negative effect of causing the electron beam intensity to be nonuniform throughout its lifetime. Higher electron extraction density requires hotter temperatures, which results in shorter lifetimes. However, wire filaments are the most physically robust of all electron emitters and can withstand the high-voltage arcs that sometimes occur inside the tube without significant filament degradation.

An alternative type of hot filament electron emitter is the dispenser cathode, which consists of a high electron-emissivity thin film on top of an integral resistor. Such integration reduces the power consumption of the device to around 1/2 W, while providing excellent thermal stability and fast warm-up times, as short as a few seconds, even though the operating temperatures are still on the order of 1000°C . The current densities obtained from dispenser cathodes, $\sim 8\text{ A/cm}^2$, are about four times higher than traditional hot filaments can provide, with similar lifetimes. In addition, since the architecture of dispenser cathodes is different than traditional hot filament cathodes, more uniform current densities are obtained over the cathode lifetime. However, the thin-film nature of these emitters makes them more susceptible to arc damage.

Cold-Cathode Field Emission Sources

Field emission sources, consisting of carbon nanotube (CNT) or small diameter nanorods of high-electron-emissivity materials, are just becoming available commercially. Currently, CNT sources are the more common. Since the diameter of the CNTs used in x-ray sources is so small, a few to tens of nanometers in diameter, very high electric field enhancements occur around the tips of the CNTs, reducing the need to apply high voltages to the whole device. In theory, voltages as low as a few tens of volts applied to a CNT device should be able to extract a sufficient electron beam density for x-ray source operation. In practice, hundreds to thousands of volts are used for the extraction process

on unheated CNT devices, although it is expected that as CNT device–manufacturing processes mature, the operating voltages will decrease. Since the operating temperature is so low, essentially no evaporation should occur, greatly increasing the “filament” lifetime. However, CNT adherence issues arise under arcing in the x-ray tube.

Inverse Geometry Sources

In conventional imaging, the source is well separated from the object and the detector, which is usually about the same size as the illuminated part of the object and in close proximity to it. In inverse geometry imaging applications, the object to be imaged is placed very close to the x-ray source and the detector is far away from both object and source. The small source-object distance requires either a prohibitively large x-ray beam or one that is scanned over the object. The latter approach is the most common for inverse-geometry sources; the electron beam is rastered across a target anode, which is approximately as large as the object being imaged. Rastering is usually accomplished by magnetically and/or electrically steering the electron beam across the target. The detector, which is typically placed about a meter away, is very small, roughly the size of a single pixel detector. Since the x-ray beam scans a large space, this type of source typically does not lend itself well to x-ray optic applications.

Electron Beam Steering and Multispot (Array) Sources

Even in conventional imaging, with the object and detector close to each other and the source farther away, being able to steer the electron beam—usually magnetically—to different positions on the target has the advantage that images from different viewing angles can be acquired simply and quickly. These images can then be combined, for example, to reconstruct a stereoscopic image of the object. Alternatively, multiple electron sources can be present inside a single x-ray source housing to accomplish the same type of imaging. Cold-cathode field emission devices particularly lend themselves to such multispot arrangements since they can be deposited in arrays with a huge number of addressable emitters.

54.4 EFFECT OF ANODE MATERIAL, GEOMETRY, AND SOURCE SIZE ON INTENSITY AND BRIGHTNESS

For electron impact sources, the two fundamental limiting factors on source brightness are the diameter of the electron beam on the target and the maximum power density the target can tolerate without melting, since most of the electron beam power is dissipated in the target as heat. This power density, determined by the electrical power, $I_{\text{tube}} V_{\text{tube}}$, limits the total flux and hence brightness that can be produced by the source, according to Eqs. (10) and (8). The brightness of the beam transmitted by an x-ray optic is limited, as stated in Liouville’s theorem,⁶ to this maximum source brightness.

Transmission Sources

In transmission sources, the electron beam is perpendicular to the target, which is thick enough to stop the electrons, but thin enough for the x rays to traverse the target thickness and exit the tube with minimal intensity losses. In reflection sources, the terms anode and target are used interchangeably, but in transmission targets they are not, since the anode tends to be positioned close to the cathode to provide the greatest electron acceleration, while the target (at ground potential)

is farther away, with electron focusing elements in between. Frequently the target material also acts as the source x-ray window, enabling focal-spot-to-optic distances of a few tens of microns, or the thickness of the transmission target. Another advantage of transmission sources, especially for many types of x-ray optics, is their small focal spot sizes, which can be as small as a few tenths of a micron with appropriate electron redirection and focusing optics. However, transmission source powers tend to be low due to limits on heat dissipation in the targets.

Target Material

For higher power sources metals with good thermal conductivity and high melting temperatures, such as tungsten and copper, are generally chosen. In addition, according to Eq. (10), the total source x-ray intensity increases with atomic number, so the most powerful sources have the largest atomic number targets, such as W or Re.

Stationary Anodes

To attain the maximum source brightness in a reflection source, where the electron beam makes an oblique angle with the anode target material and the target is stationary, direct anode cooling is necessary to prevent target melting. In this case, the target material has to be thick enough to separate the cooling medium from the vacuum required for the electron beam operation. Jets of water are frequently sprayed onto the backside of these thick targets to directly cool them. No matter how the anode is cooled, the greater the x-ray source power, the larger the x-ray beam diameter required to prevent melting. Müller⁷ has shown that the diameter (not, as one might expect, the area) increases linearly with power. For sealed-tube copper x-ray sources, the maximum power is typically 5 kW/mm of x-ray beam diameter. The x-ray beam focal spot size in stationary anode tubes can range from microns to as large as 6 mm or more, for very high voltage tubes (above 500 kVp). Sources with submillimeter diameter focal spots are referred to as micro-focus sources and contain electron-focusing elements that aid in attaining the small spot diameter. Additionally, many microfocus sources also contain electron redirection elements to keep the electron beam precisely on the same target spot. A shift in the x-ray source position of a few hundred microns is not significant when the focal spot is several millimeters in diameter; however, a similar shift of a micron diameter beam can eliminate all x-ray intensity through an optic with a small acceptance angle. Source manufacturers address this issue by using magnetic fields to correct the electron beam direction. These magnetic fields can be generated with simple external permanent magnets or with internal electromagnets connected into a feedback loop that monitors the beam position and adjusts the electromagnet current to re-center the electron beam.

Rotating Anodes

Alternatively, in reflection sources, the heat load can be reduced by rotating the target so rapidly that the electron beam dwell time on the target is too short to thermally overload the target. With advances in bearing lubricants, very fast rotations are possible. Maximum powers of roughly 150 kW/mm of x-ray beam diameter can be achieved without any direct anode cooling. The electrons that are not absorbed by the rotating target are called *secondary electrons*. These electrons, while lower in energy than the incident beam, still carry considerable kinetic energy that must be removed from the system with an electron collector, usually a sizeable piece of copper with a slit in it to let the x rays pass. The high power density of rotating anode sources usually results in larger x-ray focal spot sizes (roughly 0.1 mm to about 3 mm in diameter) than in stationary target sources. Although the x-ray power of rotating anode sources is a couple of orders of magnitude higher than for stationary targets, with accordingly more intense x-ray beams, rotating anode systems typically are larger and more costly to maintain. They also generally require larger source to window distances, limiting the available acceptance angle for small x-ray optics.

Liquid-Metal Anodes

All commercial x-ray sources have solid phase target materials. However, over the years liquid metal targets have been repeatedly proposed. The basic concept consists of a liquid metal, e.g., GaInSn or hot Sn (300°C), flowing turbulently past a thin barrier, e.g., molybdenum foil or glass, that is reasonably electron transparent; the barrier separates the liquid metal from the vacuum in which the cathode has to operate. The electrons from the cathode pass through the vacuum/liquid barrier and interact with the liquid target, producing x rays in much the same way they would from a solid target, with the liquid metal removing the heat generated during electron impact as the metal flows away from the impact region. The bremsstrahlung energy distribution from a liquid target is slightly different than that from a solid target, since the reduced density of atoms in a liquid will skew the energy distribution toward lower energies compared to those from a solid target. Liquid metal anode sources have the potential for much improved continuous loading capability and significantly higher maximum loading (a few hundred kW/mm x-ray beam diameter).

Source Shape

In many reflection sources, the electron beam intersects the target in a line rather than a circular spot. The x rays are then viewed at a small takeoff angle with respect to the target, resulting in what appears to be a nearly square or circular source spot, as shown in Fig. 5. Because the apparent size of the beam is reduced by the sine of the take-off angle, the brightness in the resultant x-ray beam is increased to

$$\text{Brightness} \propto (\text{power})/(\text{apparent area}) = (\text{power})/(A \sin \alpha) \quad (10)$$

where A is the actual electron beam spot area on the target and α is the takeoff angle (see Fig. 5). This can considerably improve the brightness for some applications. However, the x rays are not produced at the surface of the target, but a few microns into the target, depending on the electron stopping power of the target material and the electron voltage. Absorption of the x rays traveling out of the target material will eventually reduce the source brightness at small grazing angles. The radiated power is reduced exponentially with the path length,

$$\text{Brightness} \propto \frac{\text{power out}}{A \sin \alpha} = \frac{\text{power}}{A} \frac{e^{-(d/\sin \alpha)/D}}{\sin \alpha} \quad (11)$$

where d is the depth in which x rays are produced and D the absorption length for the x rays. The factor on the right is plotted in Fig. 6 for $d/D = 10$, which is approximately the case for a typical Cu K_α system. The brightness is usually maximized for a takeoff angle from 6° to 12°. Excessively small takeoff angles are more affected by surface roughness, which may increase with target use.

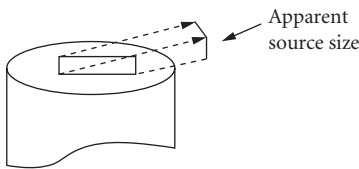


FIGURE 5 Apparent x-ray beam size from a line focus source when viewed at an angle.

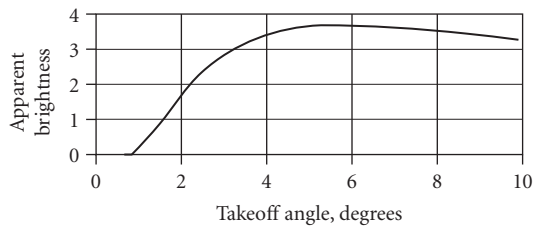


FIGURE 6 Source brightness as a function of takeoff angle for an x-ray production depth to absorption length ratio of 10.

Optics with sufficient depth of field to collect from the entire length of the line source can take advantage of the increased power emitted by line-focus sources compared to a smaller square source. It should be noted, however, that optics designed for very short source-to-optic distances often do not have large depths of field.

Source Depth and Size Measurement

The solid angle from which an optic can collect x rays depends on how close the optic can be placed to the x-ray generation point inside a source. For optics placed external to a source, this is limited by the distance from the vacuum window to the electron beam spot on the target. One method to measure this distance in practice is to move a pinhole in a series of steps transverse to the beam and record the intensity in multiple exposures, as shown in Fig. 7.⁸ The source position X relative to the pinhole is related to the image position Y and the source spot to pinhole distance L by

$$L = \frac{X}{Y} D \quad (12)$$

where D is the pinhole to image plate distance.

Most optics can only collect from relatively small source spots. Thus, source size measurement is very important for optics and system design, and also for optic characterization. Source sizes can be deduced from pinhole images. Because small pinholes may be thick compared to their diameter, it is important to ensure that the measured size is not influenced by pinhole shape, and so multiple images are required for accurate assessment.

Several images should be taken with the detector at different distances D from the pinhole, as shown in the sketch of Fig. 8. The slope of a plot of the image size W versus pinhole-to-detector distance D yields the source diameter S .

$$W = \frac{(S+2r)D}{L} + 2r \quad (13)$$

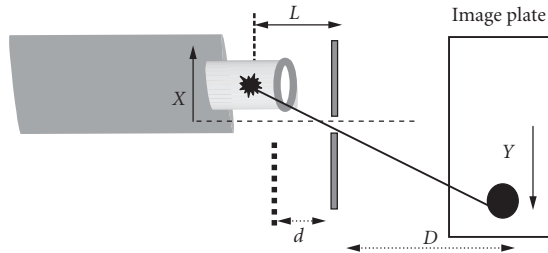


FIGURE 7 Schematic diagram of the setup for a source depth measurement.

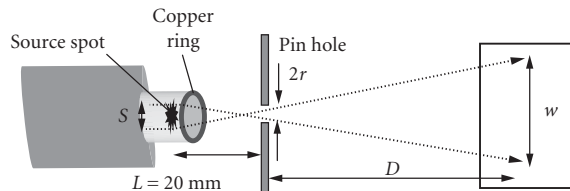


FIGURE 8 Schematic diagram of the setup for source spot measurement.

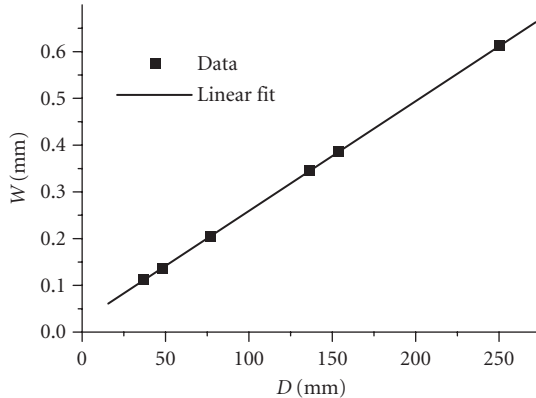


FIGURE 9 Image size W versus pinhole to image plate distance D taken with a 25- μm pinhole. The slope is 0.0024, yielding a source size of 25 μm , in good agreement with the nominal pinhole size.

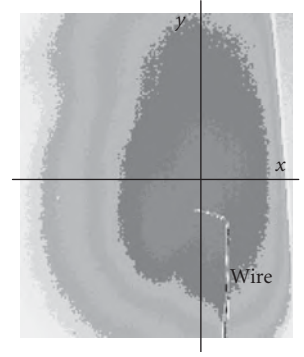


FIGURE 10 Source image at 10 cm from the source spot.

There should be good agreement between the intercept value and the pinhole radius, as shown in Fig. 9.

The beam emitted from a standard tube is often restricted in one direction due to shadowing by the anode, as shown in Fig. 10, which has a cone angle of approximately 15° in the horizontal direction.

Effective Source Size for X-Ray Optics

When applying Liouville's theorem to x-ray optics, the limiting factor is not the source focal spot size, but the area over which the optic can collect photons. Even if the optic can be placed close to the source, the optic collection area may be smaller than the actual focal spot, wasting source photons. However, large source sizes do not detract from beam brightness, if the power density is maintained and the optic only collects from a small area suitable for a specific application. For sources with nonuniform brightness, typically high peak brightness in a central region of the beam, the optic has to be carefully optimized to collect x rays from the smallest possible region that will provide sufficient power for the selected application.

54.5 GENERAL OPTIMIZATION

Trade-Offs between Low- and High-Power Sources

Low-power sources typically have shorter distances between the x-ray generation point on the target and the source exit window, where x-ray optics are usually positioned. Such small source-to-optic distances allow the optics to collect significantly larger solid angles than similar optics applied to larger sources. For applications, e.g., diffraction, where x-ray pencil beams are employed, or microanalysis, which requires small x-ray beam spots, x-ray optics coupled to low-power sources can provide similar intensities to higher-power sources, with greater convenience, portability, ease of maintenance, and lower cost. However, applications such as many medical and nondestructive imaging modalities, that require irradiating a large area in a short time use higher-power sources. These allow shorter exposure times, but are not conducive to efficient x-ray optic coupling.

Parallel Beam

As a result of Liouville's theorem, the x-ray source has to be smaller than the desired parallel beam diameter in order to achieve an intensity gain. The gain is produced by collecting over a solid angle from the source that is larger than the output divergence of the collimated beam. Therefore, it is particularly challenging to provide intense, small-cross-section, nearly parallel beams, as required by some microdiffraction applications.

Focused Beam

An optic can achieve large gains if it collects a diverging beam from a point source and refocuses it into a converging beam. The diameter of the optic is not limited by the application, but by manufacturing and integration constraints. For any optic, the minimum focal spot size will be smaller for a smaller optic-to-focus distance, as the output divergence of the optic is never zero. The best spatial definition of the beam will occur for smaller focal distances.

Effect of Different Optics Types on Brightness

Imaging optics such as pinholes and mirrors are able to conserve brightness. Pinholes are the most basic optic and can be used to define unity gain. They are also useful to further condition the beam and eliminate scatter at the sample location. A straight, untapered, single capillary tube preserves the beam condition across its length, if transport losses are ignored. Therefore, it is identical in performance to placing a pinhole at the input of the capillary. Straight single capillaries can be useful as improved pinholes or to provide more convenient geometries.⁸ (Capillaries are discussed in Chap. 52.) Mirrors are imaging optics and can preserve brightness. An ideal symmetric mirror produces a focal spot equal in size to the source spot. In order to demagnify the source focal spot size, the output focal distance has to be reduced from the symmetric case. (Mirrors are discussed in Chaps. 44 to 47.) The same considerations for mirrors are valid for diffractive optics, with the additional restriction that the Bragg diffraction condition must be met for the characteristic energy of the source target material, to obtain the most intense x-ray beam. (Diffractive optics are discussed in Chaps. 39 to 43.) Tapered single capillary tubes can provide a significantly enhanced flux over simple pinholes. If shaped appropriately, the tapered capillaries can serve as imaging micro-mirrors and fall under the mirror category. Polycapillary optics use many consecutive reflections to achieve a large overall deflection. The acceptance from the multiple capillary tubes can be overlapped so that the optic collects from a large solid angle with correspondingly large total power. However, polycapillary optics are not imaging optics and cannot conserve brightness. (Polycapillary optics are discussed in Chap. 53.)

Choosing a Source/Optic Combination

Choosing an appropriate source/optic combination depends on analyzing the requirements of the application for photon energy, energy bandwidth, beam size, beam divergence, and total power. If total x-ray power is the only consideration, a large focal spot source with high Z target material, high voltage, high current, small source-to-sample distance, and no optic is the most appropriate choice. If an optic is to be employed, it must be carefully designed to take into account the source geometry. Large focal spot sources with large source-to-output window distances are not well matched with optics designed to produce small bright beams. Detector properties, discussed in Chaps. 60 to 62, also affect the system optimization.

54.6 REFERENCES

1. H. Compton and S. K. Allison, *X Rays in Theory and Experiment*, 2nd ed., D. Van Nostrand, New York, 1935.
2. H. A. Kramers, "On the Theory of X-Ray Absorption and of the Continuous X-Ray Spectrum," *Phil. Mag.* **46**:836, 1923.
3. W. Potts, "Electron Impact X-Ray Sources," in *X-Ray Science and Technology*, A. G. Michette and C. J. Buckley (eds.), IOP Publishing, London, 1993.
4. J. L. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, C. Fiori, and E. Lifshin, *Scanning Electron Microscopy and X-Ray Microanalysis*, Plenum Press, New York, 1981, p. 97.
5. N. Nakamori, K. Yamano, M. Yamada, and H. Kanamori, "Effect of Electron Energy Distribution on Bremsstrahlung Spectrum," *Jpn. J. Appl. Phys.* **32**:4019–4025, 1993.
6. Birch and M. Marshal, "Computation of Bremsstrahlung X-Ray Spectra and Comparison with Spectra Measured with a Ge(Li) Detector," *Phys. Med. Biol.* **24**:505–517, 1979.
7. M. Tucker, G. T. Barnes, and D. P. Chakraborty, "Semiempirical Model for Generating Tungsten Target X-Ray Spectra," *Med. Phys.* **18**:211–218, 1991.
8. N. Mail, W. M. Gibson, and C. A. MacDonald, "Molybdenum Microfocus Source Coupling to Polycapillary Optics for Powder Diffraction," in *Advances in Laboratory-Based X-Ray Sources and Optics III*, A. M. Khounsary and C. A. MacDonald, (eds.), *SPIE* **4781**:87–95, 2002.
9. Thompson, D. Atwood, E. Gullikson, et al., "X-Ray Data Booklet" CXRO: Berkeley, 2002. Web site: <http://www-cxro.lbl.gov>, accessed 31 May 2009.
10. F. O'Foghludha and G. A. Johnson, "Voltage Waveform Effects on Output and Penetration of W- and Mo-Anode Mammographic Tubes," *Phys. Med. Biol.* **26**:291–303, 1981.

This page intentionally left blank

SYNCHROTRON SOURCES

Steven L. Hulbert

*National Synchrotron Light Source
Brookhaven National Laboratory
Upton, New York*

Gwyn P. Williams

*Free Electron Laser
Thomas Jefferson National Accelerator Facility
Newport News, Virginia*

55.1 INTRODUCTION

Synchrotron radiation is a bright, broadband, polarized, pulsed source of electromagnetic radiation extending from the far-infrared to the hard x-ray region. *Brightness*, defined as flux per unit area per unit solid angle, is normally a more important quantity than flux or intensity, particularly in throughput-limited applications which utilize only a small fraction of the transverse phase space of the emitted radiation or a small energy bandwidth, or both.

It is well known from classical theory of electricity and magnetism that accelerating charges emit electromagnetic radiation. In the case of synchrotron radiation, relativistic electrons are accelerated in a circular orbit and emit electromagnetic radiation in a broad spectral range. The visible portion of this spectrum was first observed on April 24, 1947 at General Electric's Schenectady facility by Floyd Haber, a machinist working with the synchrotron team, although the first theoretical predictions were by Liénard¹ in the latter part of the 1800s. An excellent early history with references is presented by Blewett² and a history covering the development of the utilization of synchrotron radiation is presented by Hartman.³

Synchrotron radiation covers the entire electromagnetic spectrum from the far-infrared (or THz) and infrared, through the visible, ultraviolet, and x-ray regions and into the very hard x-ray range up to energies of 100 kilovolts and above. If the charged particles are of low mass, such as electrons, and if they are traveling relativistically, the emitted radiation is very intense and highly collimated, with opening angles, which depend inversely on the energy of the particle, on the order of 1 milliradian. In electron storage rings there are two distinct types of sources of synchrotron radiation: dipole (bending) magnets and insertion devices. Insertion devices are further classified as either wigglers, which act like a sequence of bending magnets with alternating polarities, or undulators, which are also multiperiod alternating magnet systems but in which the beam deflections are small resulting in coherent interference of the emitted light.

In typical storage rings used as synchrotron radiation sources, several bunches of up to $\sim 10^{12}$ electrons circulate in vacuum, guided by magnetic fields. The bunches are typically several 10s of centimeters long, so that the light is pulsed, being on for a few 10s to a few 100s of picoseconds, and off for several 10s to a few 100s of nanoseconds depending on the particular machine and the radio-frequency cavity which restores the energy lost to synchrotron radiation. The revolution time for a ring of circumference 30 m is 100 ns, so that each bunch of $\sim 10^{12}$ electrons is seen 10^7 times per second, giving a

current of ~ 1 A (assuming single bunch filling). In linacs that are used as light sources the electrons are also in bunches, but these are usually much shorter resulting in pulses of < 1 ps.

The most important characteristic of accelerators built specifically as synchrotron radiation sources is that they have a magnetic focusing system which is designed to concentrate the electrons into bunches of very small transverse cross section and to keep the electron transverse velocities small. The combination of high intensity with small opening angles and small source dimensions results in the very high brightness.

The first synchrotron radiation sources to be used were operated parasitically on existing high-energy physics or accelerator development programs. These were not optimized for brightness, and were usually accelerators rather than storage rings, meaning that the electron beams were constantly being injected, accelerated, and extracted. Owing to the successful use of these sources for scientific programs, a second generation of dedicated storage rings was built starting in the early 1980s. In the mid 1990s, a third generation of sources was built, this time based largely on insertion devices, especially undulators of various types. A fourth generation of accelerator-based photon sources is now coming on line, based on what is called multiparticle coherent emission, in which coherence along the path of the electrons, or longitudinal coherence, plays the major role. This is achieved by microbunching the electrons on a length scale comparable to or smaller than the scale of the wavelengths emitted. The emission is then proportional to the square of the number of electrons N which, if N is 10^{12} , can be a very large enhancement. These sources can approach the theoretical diffraction limit of source emittance (the product of solid angle and area).

55.2 THEORY OF SYNCHROTRON RADIATION EMISSION

General

The theory describing synchrotron radiation emission is based on classical electrodynamics and can be found in the works of Tomboulion and Hartman,⁴ Schwinger,⁵ Jackson,⁶ Winick,⁷ Hofmann,⁸ Krinsky, Perlman, and Watson,⁹ and Kim¹⁰. A quantum description, presented by Sokolov and Ternov,¹¹ is quantitatively equivalent.

Here we present a phenomenological description in order to highlight the general concepts involved. Electrons in circular motion radiate in a dipole pattern as shown schematically in Fig. 1a.

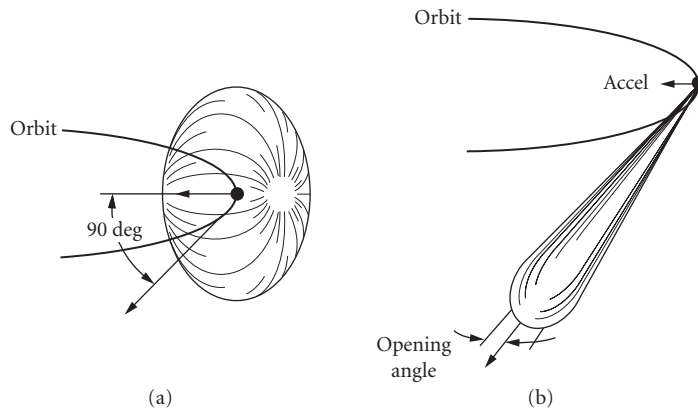


FIGURE 1 Conceptual representation of the radiation pattern from a charged particle undergoing circular acceleration at (a) nonrelativistic and (b) relativistic velocities.

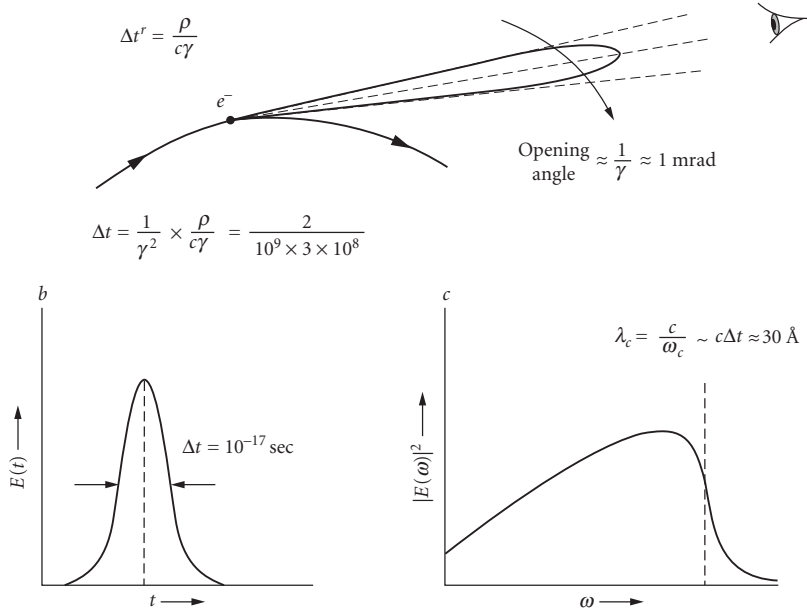


FIGURE 2 Illustration of the derivation of the spectrum emitted by a charged particle in a storage ring.

As the electron energies increase and the particles start traveling at relativistic velocities, this dipole pattern appears different to an observer in the rest frame of the laboratory. Special relativity tells us that angles θ_t in a transmitting object are related to those in the receiving frame θ_r by

$$\tan \theta_r = \frac{\sin \theta_t}{\gamma (\cos \theta_t - \beta)} \quad (1)$$

with γ the ratio of the mass of the electron to its rest mass, being given by $E/m_0 c^2$, E being the electron energy, m_0 the electron rest mass, and c the velocity of light. β is the ratio of electron velocity v to the velocity of light c . Since $\beta \approx 1$ for electrons travelling at relativistic energies, the peak of the dipole emission pattern in the particle frame, $\theta_t = 90^\circ$, transforms to $\theta_r \approx \tan \theta_r \approx \gamma^{-1}$ in the laboratory frame as shown in Fig. 1b. Thus γ^{-1} is a typical opening angle of the radiation in the laboratory frame. For an electron viewed in passing by an observer, as shown in Fig. 2, the duration of the pulse produced by a particle under circular motion of radius ρ will be $\rho/\gamma c$ in the particle frame, or $\rho/\gamma c \times 1/\gamma^2$ in the laboratory frame owing to time dilation. The Fourier transform of this function will contain frequency components up to the reciprocal of this time interval. For a storage ring with a radius of 2 meters and $\gamma = 1000$, corresponding to a stored electron beam energy of ~ 500 MeV, the time interval is 10^{-17} seconds, which corresponds to light of wavelength 30 Å.

Bending Magnet Radiation

For an electron storage ring, the relationship between the electron beam energy E , bending radius ρ , and field B is

$$\rho = \frac{E}{ecB} = \frac{E[\text{GeV}]}{0.300B[\text{T}]} \quad (2)$$

where γ the ratio of the mass of the electron to its rest mass is given by $\gamma = E/m_0 c^2 = E/0.511 \text{ MeV} = 1957E$ [GeV] and λ_c which is defined as the wavelength for which half the power is emitted above and half below, is

$$\lambda_c = \frac{4\pi\rho}{3\gamma^3} = 5.59 \text{ \AA} \frac{\rho[m]}{E^3[\text{GeV}^3]} = \frac{18.6 \text{ \AA}}{B[T]E^2[\text{GeV}^2]} \quad (3)$$

The critical frequency and photon energy are

$$\omega_c = \frac{2\pi c}{\lambda_c} = \frac{3c\gamma^3}{2\rho} \quad \varepsilon_c[eV] = \hbar\omega_c(eV) = 665.5E^2[\text{GeV}^2]B[T] \quad (4)$$

The angular distribution of synchrotron radiation flux emitted by electrons moving through a bending magnet with a circular trajectory in the horizontal plane is given¹⁰ by

$$\begin{aligned} \frac{d^2 F_{bm}(\omega)}{d\theta d\psi} &= \frac{3\alpha}{4\pi^2} \gamma^2 \frac{\Delta\omega}{\omega} \frac{I}{e} \left(\frac{\omega}{\omega_c} \right)^2 (1 + \gamma^2 \psi^2)^2 \left[K_{2/3}^2(\xi) + \frac{\gamma^2 \psi^2}{1 + \gamma^2 \psi^2} K_{1/3}^2(\xi) \right] \\ &= 1.326 \times 10^{13} \text{ photons/sec/mrad}^2 / 0.1\% \text{ bandwidth} \\ &\quad \times E^2[\text{GeV}^2] I[A] (1 + \gamma^2 \psi^2)^2 \left(\frac{\omega}{\omega_c} \right)^2 \left[K_{2/3}^2(\xi) + \frac{\gamma^2 \psi^2}{1 + \gamma^2 \psi^2} K_{1/3}^2(\xi) \right] \end{aligned} \quad (5)$$

where θ is the observation angle in the horizontal plane, ψ the observation angle in the vertical plane, α the fine structure constant ($1/137$), ω the light frequency, I the beam current, and $\xi = (\omega/2\omega_c)(1 + \gamma^2 \psi^2)^{3/2}$. The subscripted K s are modified Bessel functions of the second kind. The $K_{2/3}$ term represents light linearly polarized parallel to the electron orbit plane, while the $K_{1/3}$ term represents light linearly polarized perpendicular to the orbit plane.

If one integrates over all vertical angles, then the total intensity is

$$\begin{aligned} \frac{dF_{bm}(\omega)}{d\theta} &= \frac{\sqrt{3}}{2\pi} \alpha \gamma \frac{\Delta\omega}{\omega} \frac{I}{e} \frac{\omega}{\omega_c} \int_{\omega/\omega_c}^{\infty} K_{5/3}(y) dy \\ &= 2.457 \times 10^{13} \text{ photons/sec/mrad} / 0.1\% \text{ bandwidth} \times E[\text{GeV}] I[A] \frac{\omega}{\omega_c} \int_{\omega/\omega_c}^{\infty} K_{5/3}(y) dy \end{aligned} \quad (6)$$

The Bessel functions can be computed easily using algorithms of Kostroun:¹²

$$K_\nu(x) = h \left\{ \frac{e^{-x}}{2} + \sum_{r=1}^{\infty} e^{-x \cosh(rh)} \cosh(vrh) \right\} \quad (7)$$

and

$$\int_x^{\infty} K_\nu(\eta) d\eta = h \left\{ \frac{e^{-x}}{2} + \sum_{r=1}^{\infty} e^{-x \cosh(rh)} \frac{\cosh(vrh)}{\cosh(rh)} \right\} \quad (8)$$

for all x and for any fractional order ν , where h is some suitable interval such as 0.5. In evaluating the series, the sum is terminated when the r th term is small, $<10^{-5}$, for example.

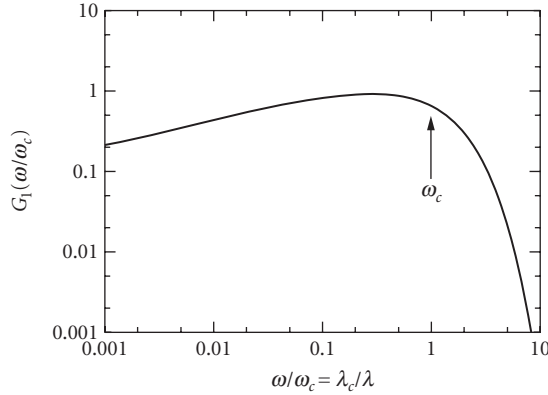


FIGURE 3 Universal synchrotron radiation output curve.

In Fig. 3 we plot the universal function

$$G_1\left(\frac{\omega}{\omega_c}\right) = \frac{\omega}{\omega_c} \int_{\omega/\omega_c}^{\infty} K_{5/3}(y) dy$$

from Eq. (6), so that the photon energy dependence of the flux from a given ring can be calculated readily. It is found that the emission falls off exponentially as $e^{-\lambda_c/\lambda}$ for wavelengths shorter than λ_c , but only as $\lambda^{-1/3}$ at longer wavelengths.

The vertical angular distribution is more complicated. For a given ring and wavelength, there is a characteristic natural opening angle for the emitted light. The opening angle increases with increasing wavelength. If we define ψ as the vertical angle relative to the orbital plane, and if the vertical angular distribution of the emitted flux is assumed to be Gaussian in shape, then the *rms* divergence σ_ψ is defined as $1/\sqrt{2\pi}$ times the ratio of Eqs. (5) and (6) evaluated at $\psi = 0$:

$$\sigma_\psi = \sqrt{\frac{2\pi}{3}} \frac{1}{\gamma} \left(\frac{\omega}{\omega_c}\right)^{-1} \frac{\int_{\omega/\omega_c}^{\infty} K_{5/3}(y) dy}{K_{2/3}^2(\omega/2\omega_c)} \quad (9)$$

In reality, the distribution is not Gaussian, especially in view of the fact that the distribution for the vertically polarized component vanishes in the horizontal plane ($\psi = 0$). However, σ_ψ defined by Eq. (9) is still a simple and useful measure of the angular divergence. The photon energy (ω) dependence of the electron-energy-independent quantity $\gamma\sigma_\psi$ is plotted in Fig. 4. At $\omega = \omega_c$, $\sigma_\psi = 0.647/\gamma$. The asymptotic values of σ_ψ can be obtained from the asymptotic values of the Bessel functions and are

$$\sigma_\psi \approx \frac{1.07}{\gamma} \left(\frac{\omega}{\omega_c}\right)^{-1/3} \quad \omega \ll \omega_c \quad (10)$$

and

$$\sigma_\psi \approx \frac{0.58}{\gamma} \left(\frac{\omega}{\omega_c}\right)^{-1/2} \quad \omega \gg \omega_c \quad (11)$$

In Fig. 5 we show examples of the normalized vertical angular distributions of both parallel and perpendicularly polarized synchrotron radiation for a selection of wavelengths.

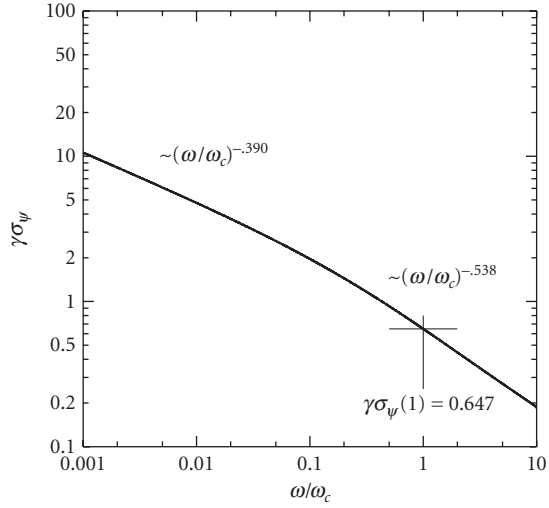


FIGURE 4 Plot of the normalized vertical opening angle $\gamma\sigma_\psi$ for bending magnet radiation.

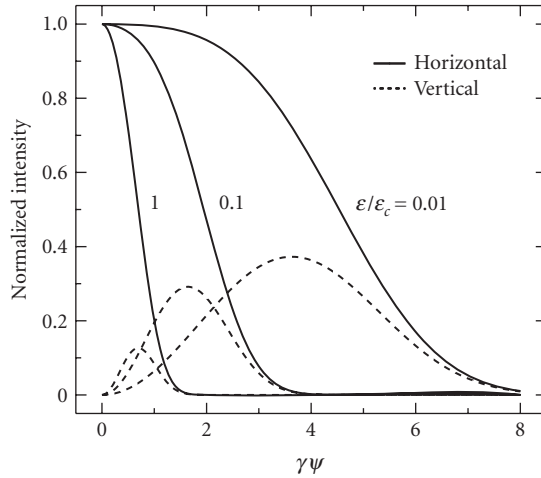


FIGURE 5 Normalized intensities of horizontal and vertical polarization components, as functions of the vertical observation angle for different photon energies.

Circular Polarization and Aperturing for Magnetic Circular Dichroism

Circularly polarized radiation is a valuable tool for the study of electronic, magnetic, and geometric structure of a wide variety of materials. The dichroic response in the soft x-ray spectral region (100 to 1500 eV) is especially important because in this energy range almost every element has a strong dipole transition from a sharp core level to its lowest unoccupied state.¹³

The production of bright sources of circularly polarized *soft x-rays* is therefore a topic of keen interest, and is a problem which has seen a multitude of solutions, from special insertion devices

(crossed undulators, helical undulators, elliptically polarized undulators/wigglers) to optical devices (multiple-bounce reflectors/multilayers and quarter-wave plates). However, standard bending magnet synchrotron radiation sources are good sources of elliptically polarized soft x-rays when viewed from either above or below the orbital plane.

As discussed by Chen,¹³ a practical solution involves acceptance of a finite vertical angular range, $\psi_{\text{off}} - \Delta\psi/2 < \psi < \psi_{\text{off}} + \Delta\psi/2$ centered about any vertical offset angle $\psi = \psi_{\text{off}}$ or, equivalently, about $\psi = -\psi_{\text{off}}$. This slice of bending magnet radiation exhibits a circular polarization¹⁴

$$P_c = -\frac{2A_h A_v}{(A_h^2 + A_v^2)} \quad (12)$$

where $A_h = K_{2/3}(\xi)$ and $A_v = \gamma\psi/(1 + \gamma^2\psi^2)^{1/2} K_{1/3}(\xi)$ are proportional to the square-roots of the horizontally and vertically polarized components of bending magnet flux [Eq. (5)], i.e., A_h and A_v are proportional to the horizontal and vertical components of the electric field, respectively. P_c depends on the vertical angle ψ , electron energy γ and, through ξ , the emitted photon energy ω/ω_c . In Fig. 6 we plot values of P_c vs $\gamma\psi$ and ω/ω_c for $\gamma = 1565$ ($E = 0.8\text{GeV}$) and $\rho = 1.91\text{m}$ ($h\nu_{\text{crit}} = 594\text{ eV}$).

Magnetic circular dichroism (MCD) measures the normalized difference of the absorption of right circular and left circular light. Assuming no systematic error, the signal to noise ratio in such a measurement defines a figure of merit.

$$\text{MCD figure of merit} = (\text{average circular polarization}) \times (\text{flux fraction})^{1/2} \quad (13)$$

where

$$\text{average circular polarization} = \frac{\int_{\psi_{\text{off}} - \frac{\Delta\psi}{2}}^{\psi_{\text{off}} + \frac{\Delta\psi}{2}} P_c(\psi) \frac{dF}{d\psi} d\psi}{\int_{\psi_{\text{off}} - \frac{\Delta\psi}{2}}^{\psi_{\text{off}} + \frac{\Delta\psi}{2}} \frac{dF}{d\psi} d\psi} \quad (14)$$

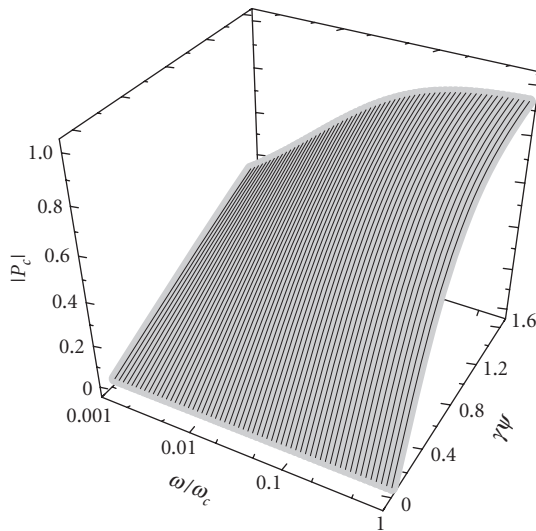


FIGURE 6 P_c versus $\gamma\psi$ versus ω/ω_c

and the fraction of the total (vertically-integrated) flux emitted into the vertical slice $\psi = \psi_{\text{off}} \pm \Delta\psi/2$ is

$$\text{flux fraction} = \frac{1}{dF_{bm}(\omega)/d\theta} \int_{\psi_{\text{off}} - \Delta\psi/2}^{\psi_{\text{off}} + \Delta\psi/2} \frac{d^2F_{bm}(\omega)}{d\theta d\psi} d\psi \quad (15)$$

Here $d^2F_{bm}(\omega)/d\theta d\psi$ is the angular dependence of the bending magnetic flux from Eq. (5) and $dF_{bm}(\omega)/d\theta$ is the vertically integrated flux from Eq. (6). For a 0.8 GeV storage ring (e.g., the VUV ring at the National Synchrotron Light Source (NSLS), Upton, NY USA), the choices of ψ and $\Delta\psi$ that maximize the MCD figure of merit are 0.5 mrad and 0.66 mrad, respectively. This yields a flux fraction ~ 0.3 , a circular polarization ~ 0.65 and a figure of merit ~ 0.35 .

Bending Magnet Power

Integration of $\hbar\omega(d^2F_{bm}(\omega)/d\theta d\psi)$ from Eq. (5) over all frequencies ω yields the angular distribution of power radiated by a bending magnet:

$$\begin{aligned} \frac{d^2P_{bm}}{d\theta d\psi} &= \int_0^\infty \hbar\omega \frac{d^2F_{bm}(\omega)}{d\theta d\psi} d\omega = \frac{I}{e} \frac{\alpha hc \gamma^5}{2\pi\rho} \frac{7}{16} F(\gamma\psi) \\ &= 18.082 \text{ W/mrad}^2 \times \frac{E^5[\text{GeV}^5] I[\text{A}]}{\rho[\text{m}]} F(\gamma\psi) \end{aligned} \quad (16)$$

which is independent of the horizontal angle θ as required by symmetry, and the vertical angular dependence is contained in the factor

$$F(\gamma\psi) = \frac{1}{(1 + \gamma^2\psi^2)^{5/2}} \left[1 + \frac{5}{7} \frac{\gamma^2\psi^2}{(1 + \gamma^2\psi^2)} \right] \quad (17)$$

The first term in $F(\gamma\psi)$ represents the component of the bending magnet radiation parallel to the orbital plane, while the second represents the perpendicular polarization component. $F(\gamma\psi)$ and its polarization components are plotted versus $\gamma\psi$ in Fig. 7. Note that the area under the F_{parallel} curve is approximately seven times greater than that for $F_{\text{perpendicular}}$.

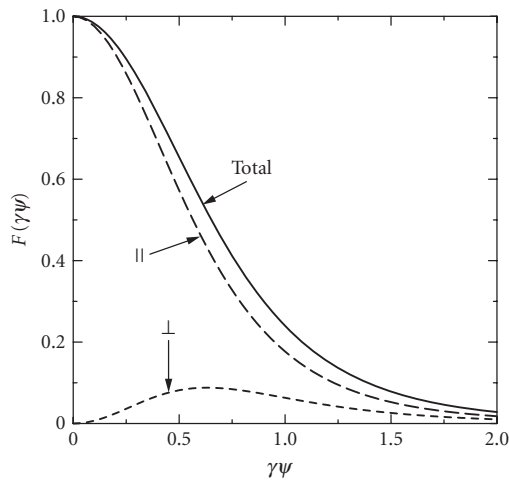


FIGURE 7 Vertical angle dependence of bending magnet power, $F(\gamma\psi)$, versus $\gamma\psi$.

Integrating Eq. (17) over the out-of-orbital-plane (vertical) angle ψ yields the total power radiated per unit in-orbital-plane (horizontal) angle θ :

$$\frac{dP_{bm}}{d\theta} = \frac{I}{e} \frac{\hbar c \alpha \gamma^4}{3\pi\rho} = 14.08 \text{ W/mrad} \times \frac{E^4 [\text{GeV}^4] I [A]}{\rho [m]} \quad (18)$$

For example, a 1.0-GeV storage ring with 2 m radius bends generates 7.04 W/mrad/Amp of stored current. By contrast, a 2.5 GeV machine with 7-m radius bends generates 78.6 W/mrad/A and a 7 GeV machine with 39 m radius bends generates 867 W/mrad/A.

Bending Magnet Brightness

Thus far we have calculated the emitted flux in photons per second per milliradian² of solid angle. In order to calculate the brightness we need to include the source size. In these calculations we calculate the central (or maximum) brightness, for which we use the natural opening angle to define both the horizontal and vertical angles. Using vertical angles larger than this will not increase the flux as there is no emission. Using larger horizontal angles will increase the flux proportionately as all horizontal angles are filled with light, but owing to the curvature of the electron trajectory, the *average* brightness will actually be less. The brightness expression^{15,16} is

$$B_{bm} = \frac{\frac{d^2 F_{bm}}{(d\theta d\psi)} \big|_{\psi=0}}{2\pi \sum_x \sum_y} \quad (19)$$

where

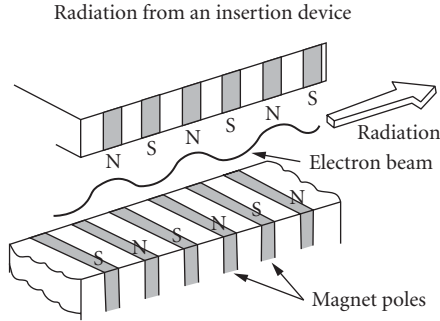
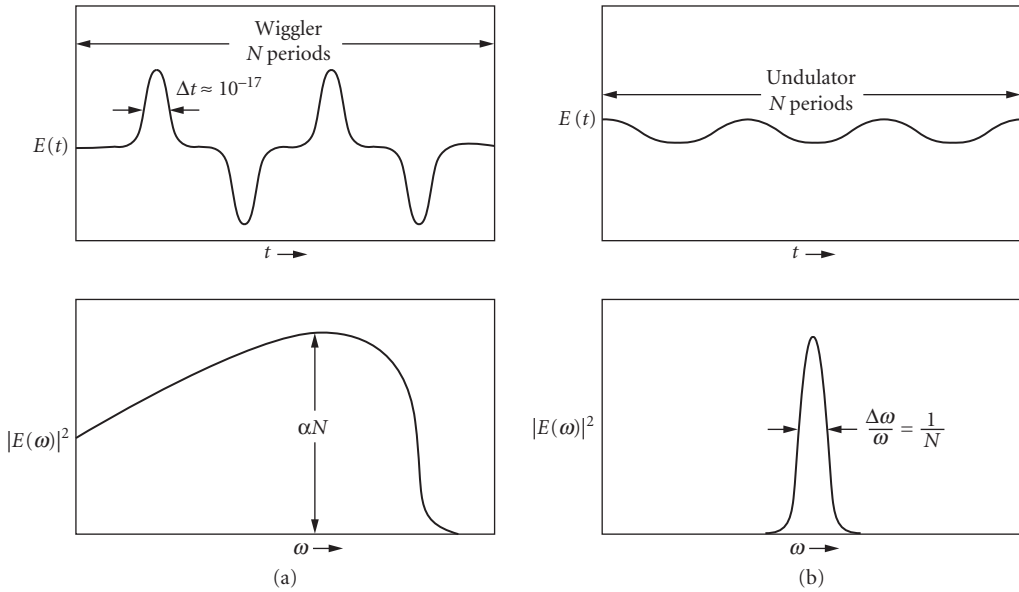
$$\sum_x = [\epsilon_x \beta_x + \eta_x^2 \sigma_E^2 + \sigma_r^2]^{1/2} \quad \text{and} \quad \sum_y = \left[\epsilon_y \beta_y + \sigma_r^2 + \frac{\epsilon_y^2 + \epsilon_y \gamma_y \sigma_r^2}{\sigma_\psi^2} \right]^{1/2} \quad (20)$$

ϵ_x and ϵ_y are the electron beam emittances in the horizontal and vertical directions respectively, β_x and β_y are the electron beam beta functions in the horizontal and vertical planes, η_x is the dispersion function in the horizontal plane, and σ_E is the rms value of the relative energy spread. All the electron beam parameters are properties of a particular storage ring. The diffraction-limited source size is $\sigma_r = \lambda/4\pi\sigma_\psi$. The effective source sizes ($\sum_x$ and $\sum_y$) are photon energy dependent via the natural opening angle σ_ψ and the diffraction-limited source size σ_r .

55.3 INSERTION DEVICES (UNDULATORS AND WIGGLERS)

General

Insertion devices are periodic magnetic structures installed in straight sections of storage rings, as illustrated in Fig. 8, in which the vertical magnetic field varies approximately sinusoidally along the axis of the undulator. The resulting motion of the electrons is also approximately sinusoidal, but in the horizontal plane. One can understand the nature of the spectra emitted from these devices by again studying the electric field as a function of time, and this is shown in Fig. 9. This shows that the electric field and hence its Fourier transform, the spectrum, depend critically on the magnitude of the beam deflection in the device. At one extreme, when the magnetic fields are high, as in Fig. 9a, the deflection is large and the electric field is a series of pulses similar to those obtained from a

**FIGURE 8** Schematic of an insertion device.**FIGURE 9** Conceptual representation of the electric fields emitted as a function of time by an electron in (a) a wiggler and (b) an undulator, with the corresponding spectra.

dipole. Such a device is termed a “wiggler.” The Fourier transform for the wiggler is N times that of a single dipole. At the other extreme, as in Fig. 9b, the deflection of the electron beam is such that the electric field as a function of time is sinusoidal, and the Fourier transform is then a single peak with a width proportional to the inverse of the length of the wavetrain L^* according to $\lambda^2/\Delta\lambda = L^*$. L^* is obtained by dividing the real length of the device L by γ^2 because of relativistic effects. Thus for a meter long device emitting at a wavelength $\lambda = 10$ nm in a machine of energy 0.5 GeV ($\gamma \sim 1000$), corresponding to, for example, 1 cm period length and magnetic field strength $B = 1.5$ T, we get $\lambda^2/\Delta\lambda = 10^{-6}$ meters, and $\lambda/\Delta\lambda = 100$. Note that $\Delta\lambda/\lambda \sim 1/N$ as expected. Interference occurs in an undulator since the electric field from one part of the electron path is added coherently to that from adjacent parts.

Formal Treatment

We assume that the motion of an electron in an insertion device is sinusoidal, and that we have a magnetic field in the vertical (y) direction varying periodically along the z direction, with

$$B_y = -B_0 \sin(2\pi z/\lambda_u) \quad 0 \leq z \leq N\lambda_u \quad (21)$$

where B_0 is the peak magnetic field, λ_u is the period length, and N the number of periods. By integrating the equation of motion, the electron transverse velocity $c\beta_x$ is found to be

$$\beta_x = \frac{K}{\gamma} \cos(2\pi z/\lambda_u) \quad (22)$$

where

$$K = \frac{eB_0\lambda_u}{2\pi mc} = 0.934\lambda_u[cm] B_0[T] \quad (23)$$

is a dimensionless parameter which is proportional to the deflection of the electron beam. The maximum slope of the electron trajectory is $\delta = (K/\gamma)$. In terms of δ , we define an undulator as a device in which $\delta \leq \gamma^{-1}$, which corresponds to $K \leq 1$. When K is large, the device is called a wiggler. In most insertion devices the field can be changed either electromagnetically or mechanically, and in some cases K can vary between the two extremes of undulator and wiggler operation.

Wigglers

For the wiggler, the flux distribution is given by $2N$ (where N is the number of magnetic periods) times the appropriate bending magnet formulae in Eqs. (5) and (6). However, ρ or B must be taken at the point in the path of the electron which is tangent to the direction of observation. For a horizontal angle θ ,

$$\varepsilon_c(\theta) = \varepsilon_{c\max} \sqrt{1 - (\theta/\delta)^2} \quad (24)$$

where

$$\varepsilon_{c\max}[keV] = 0.665 E^2[GeV^2] B_0[T] \quad (25)$$

from Eq. (4). Integration over θ , which is usually performed numerically, gives the wiggler flux.

The calculation of the brightness of wigglers needs to take into account the depth-of-field effects, i.e., the contribution to the apparent source size from different poles. The expression for the brightness of wigglers is

$$B_W = \frac{d^2 F_W}{d\theta d\psi} \sum_{\pm} \sum_{n=-\frac{1}{2}N}^{\frac{1}{2}N} \frac{1}{2\pi} \times \frac{\exp\left[-\frac{1}{2}\left(\frac{x_o^2}{\sigma_x^2 + z_{n\pm}^2 \sigma_x'^2}\right)\right]}{\left[(\sigma_x^2 + z_{n\pm}^2 \sigma_x'^2) \left(\frac{\varepsilon_y^2}{\sigma_\psi^2} + \sigma_y^2 + z_{n\pm}^2 \sigma_y'^2\right)\right]^{1/2}} \quad (26)$$

where $z_{n\pm} = \lambda_w[n \pm (1/4)]$, λ_w is the wiggler period, and σ_ψ is identical to Eq. (9), but evaluated, in the wiggler case, as the instantaneous radius at the tangent to the straight-ahead ($\theta = \psi = 0$) direction (i.e., minimum ρ , maximum ε_c), $\sigma_x = \sqrt{\varepsilon_x \beta_x}$ and $\sigma_y = \sqrt{\varepsilon_y \beta_y}$ are the rms transverse beam sizes,

while $\sigma'_x = \sqrt{\varepsilon_x/\beta_x}$ and $\sigma'_y = \sqrt{\varepsilon_y/\beta_y}$ are the angular divergences of the electron beam in the horizontal and vertical directions respectively. The exponential factor in Eq. (26) arises because wigglers have two source points separated by $2x_o$, where

$$x_o = \frac{K}{\gamma} \frac{\lambda_w}{2\pi} \quad (27)$$

The summations in Eq. (26) must be performed for each photon energy because σ_ψ is photon-energy dependent.

Undulators

The interference which occurs in an undulator, i.e., when K is moderate ($K \leq 1$), produces sharp peaks in the forward direction at a fundamental ($n = 1$) and all odd harmonics ($n = 3, 5, 7, \dots$) as shown for a zero emittance ($\varepsilon = 0$) electron beam in Fig. 10a (dotted line). In the $\varepsilon = 0$ case, the even harmonics ($n = 2, 4, 6, \dots$) peak off-axis and do not appear in the forward direction. For real ($\varepsilon \neq 0$) electron beams, the spectral shape, angular distribution, and peak brightness are strongly dependent on the emittance and energy spread of the electron beam as well as the period and magnitude of the insertion device field.

In general, the effect of electron beam emittance is to cause all harmonics to appear in the forward direction (solid line in Fig. 10a). The effect of angle integration on the spectrum in Fig. 10a is shown in Fig. 10b, a spectrum which is independent of electron beam emittance except for the presence of “noise” in the zero emittance case. The effect of electron beam emittance on the angular distribution of the fundamental, second, and third harmonics of this device is shown in Fig. 10c, which also nicely demonstrates the dependence on harmonic number.

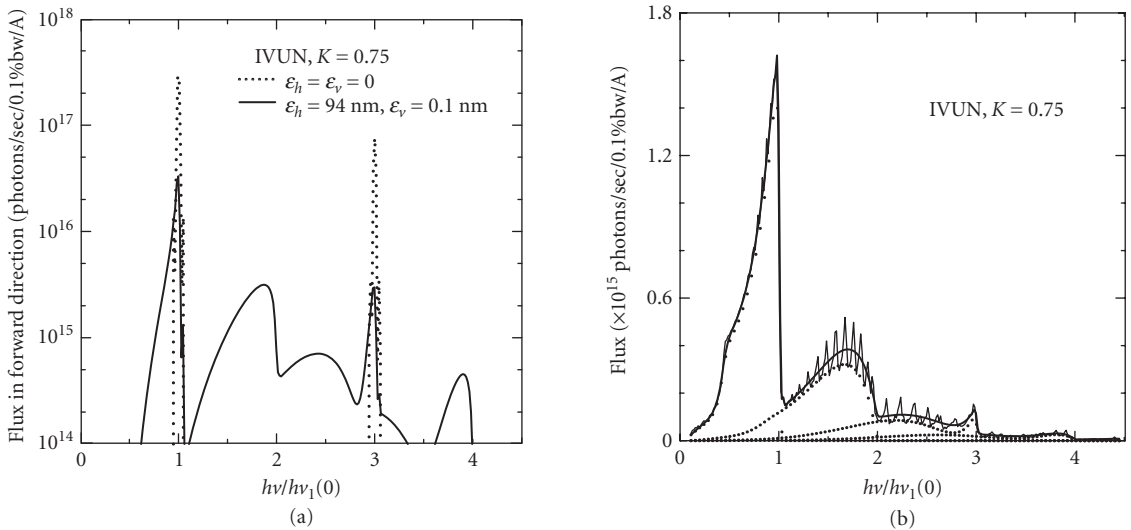


FIGURE 10 Spectral output and angular distribution of the emission from the early-2000s vintage NSLS In-Vacuum Undulator (IVUN) for $K = 0.75$. (a) Spectral output in the forward direction, with (solid line) and without (dotted line) the effect of electron beam emittance; (b) angle-integrated spectral output with (solid line) and without (faint solid line) the effect of electron beam emittance, and the decomposition into harmonics ($n = 1, 2, 3, 4$) (dotted lines); and (c) angular distribution of the first three harmonics ($n = 1, 2, 3$), with and without the effect of electron beam emittance. Emittance values used: 94-nm horizontal, 0.1-nm vertical.

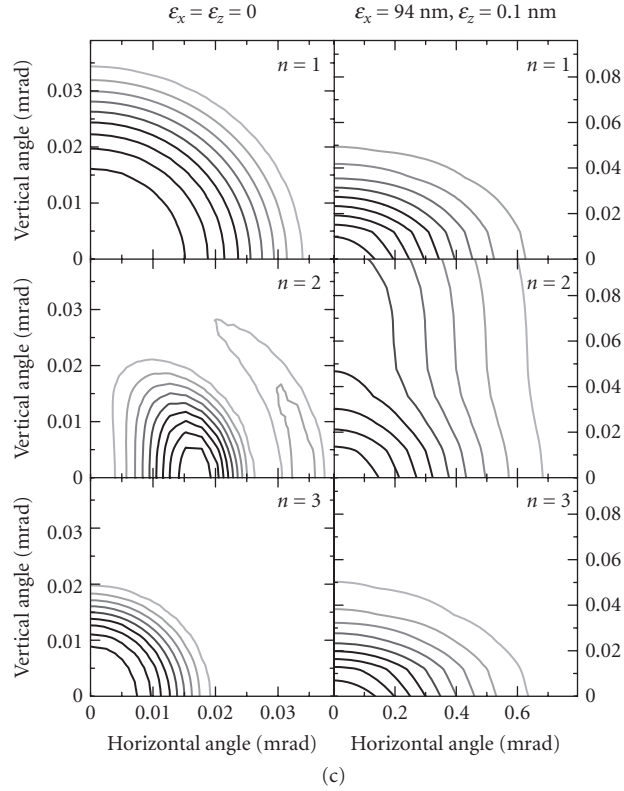


FIGURE 10 (Continued)

The peak wavelengths of the emitted radiation λ_n are given by

$$\lambda_n = \frac{\lambda_u}{2n\gamma^2} \left(1 + \frac{K^2}{2} + \gamma^2 \theta^2 \right) \quad n=1, 3, 5, \dots \quad (28)$$

where λ_u is the undulator period length. They soften as the square of the deviation angle θ away from the forward direction.

Of main interest is the intense central cone of radiation. An approximate formula for flux integrated over the central cone (for the odd harmonics) is

$$\begin{aligned} F_u(K, \omega) &= \pi \alpha N \frac{\Delta \omega}{\omega} \frac{I}{e} Q_n(K) \quad n=1, 3, 5, \dots \\ &= 1.431 \times 10^{14} \text{ photons/sec/0.1\% bandwidth} \times I[A] N Q_n(K) \end{aligned} \quad (29)$$

where

$$Q_n(K) = \left(1 + \frac{K^2}{2} \right) \frac{F_n(K)}{n} \quad n=1, 3, 5, \dots \quad (30)$$

and

$$F_n(K) = \frac{K^2 n^2}{(1 + K^2/2)^2} \left\{ J_{(n-1)/2} \left[\frac{nK^2}{4 \left(1 + \frac{1}{2}K^2\right)} \right] - J_{(n+1)/2} \left[\frac{nK^2}{4 \left(1 + \frac{1}{2}K^2\right)} \right] \right\}^2 \quad (31)$$

Here $J_{(n \pm 1)/2}$, $n = 1, 3, 5, \dots$ are the integer Bessel functions of the first kind: $J_0, J_1, J_2, \dots$

To calculate the undulator flux angular distribution and spectral output into arbitrary solid angle, one can use freely available codes such as Urgent¹⁷ (R. P. Walker and B. Diviacco). To include magnetic field errors (e.g., measured values), use Ur¹⁸ (R. J. Dejus and A. Luccio), SRW¹⁹ (O. Chubar and P. Elleaume), or Spectra²⁰ (T. Tanaka and H. Kitamura).

The brightness of an undulator B_u is approximated by dividing the central cone flux by the effective angular divergence, $\Sigma'_x(\Sigma'_y)$, and by the effective source size, $\Sigma_x(\Sigma_y)$, in the horizontal (vertical) directions. These are given by convolution of the Gaussian distributions of the electron beam and the diffraction-limited photon beam, in both angle and space:

$$\Sigma_{x'} = \sqrt{\sigma_{x'}^2 + \sigma_r^2} \quad \Sigma_{y'} = \sqrt{\sigma_{y'}^2 + \sigma_r^2} \quad (32)$$

$$\Sigma_x = \sqrt{\sigma_x^2 + \sigma_r^2} \quad \Sigma_y = \sqrt{\sigma_y^2 + \sigma_r^2} \quad (33)$$

Thus, B_u is given by

$$B_u = \frac{F_u}{(2\pi)^2 \Sigma_x \Sigma_y \Sigma_{x'} \Sigma_{y'}} \quad (34)$$

The diffraction-limited emittance of a photon beam is the minimum value in the inequality

$$\varepsilon = \sigma_r \sigma_{r'} \geq \frac{\hbar}{2} = \frac{\lambda}{4\pi} \quad (35)$$

where ε is the photon emittance and λ is the wavelength, in direct analogy to the Heisenberg uncertainty principle in nonrelativistic quantum mechanics. The space versus angle separation of this minimum emittance is energy and harmonic dependent.²¹ For the exact harmonic frequency in the forward direction, given by Eq. (28) with $\theta = 0$, there appears to be consensus that σ_r and $\sigma_{r'}$ are given by

$$\sigma_{r'} = \sqrt{\frac{\lambda}{2L}} \quad \text{and} \quad \sigma_r = \frac{\sqrt{2\lambda L}}{4\pi} \quad (36)$$

On the other hand, at the peak of the angle-integrated undulator spectrum, which lies a factor of $[1 - (1/nN)]$ below the exact harmonic energy, σ_r and $\sigma_{r'}$ are given by

$$\sigma_{r'} = \sqrt{\frac{\lambda}{L}} \quad \text{and} \quad \sigma_r = \frac{\sqrt{\lambda L}}{4\pi} \quad (37)$$

It is clear from Eqs. (32) and (33) that the choice of expression for σ_r and $\sigma_{r'}$ can have a non-negligible effect on the undulator brightness value especially for small beam size and opening angle. Lacking a specific functional form for σ_r and $\sigma_{r'}$ as a function of photon energy, we generally use Eq. (36) in evaluating the expression for undulator spectral brightness from Eq. (34).

Insertion Device Power

The Schwinger⁵ formula for the distribution of radiated power from an electron in a sinusoidal trajectory, which applies with reasonable approximation to undulators and, to a lesser extent, wigglers, reduces²² to

$$\frac{d^2P}{d\theta d\psi} = P_{\text{total}} \frac{21\gamma^2}{16\pi K} G(K) f_K(\gamma\theta, \gamma\psi) \quad (38)$$

where the total (angle-integrated) radiated power is

$$P_{\text{total}} = \frac{N}{6} \frac{Z_0 I 2\pi e c}{\lambda_u} \gamma^2 K^2 = 633.0 \text{ W} \times E^2 [\text{GeV}^2] B_0^2 [T^2] L [m] I [A] \quad (39)$$

where N is the number of undulator or wiggler periods, Z_0 is the vacuum impedance (377Ω), I is the storage ring current, e is the electronic charge, c is the speed of light, $L = N\lambda_u$ is the length of the insertion device,

$$G(K) = \frac{K}{(1 + K^2)^{7/2}} \left(K^6 + \frac{24}{7} K^4 + 4K^2 + \frac{16}{7} \right) \quad (40)$$

and

$$f_K(\gamma\theta, \gamma\psi) = \frac{16}{7\pi} \frac{K}{G(K)} \int_{-\pi}^{\pi} d\alpha \left(\frac{1}{D^3} - \frac{4(\gamma\theta - K \cos \alpha)^2}{D^5} \right) \sin^2 \alpha \quad (41)$$

where

$$D = 1 + \gamma^2 \psi^2 + (\gamma\theta - K \cos \alpha)^2 \quad (42)$$

The integral in the expression for f_K is best evaluated numerically.

For $K > 1$, which includes all wigglers and much of the useful range of undulators, an approximate formula for the angle dependence of the radiated power is

$$f_K(\gamma\theta, \gamma\psi) = \sqrt{1 - \left(\frac{\gamma\theta}{K} \right)^2} F(\gamma\psi) \quad (43)$$

where $F(\gamma\psi)$ is the bending magnet formula from Eq. (17). This form clearly indicates the strong weakening of insertion device power as θ increases, vanishing at $\theta = \pm K/\gamma$.

Since $f_K(0,0)$ is normalized to unity, the radiated power density in the forward direction (i.e., along the undulator axis) is

$$\frac{d^2P}{d\theta d\psi}(\theta = 0, \psi = 0) = P_{\text{total}} \frac{21\gamma^2}{16\pi K} G(K) = 10.84 \text{ W/mrad}^2 \times B_0 [T] E^4 [\text{GeV}^4] I [A] N G(K) \quad (44)$$

Polarization of Undulators and Wigglers

The polarization properties of the light emitted by wigglers are similar to those of dipoles. For both sources the radiation is elliptically polarized when observed at some angle away from the orbital plane as given by Eq. (5). For radiation from planar undulators, however, the polarization is always linear. The polarization direction, which is in the horizontal plane when observed from that plane,

rotates in a complicated way at other directions of observation. A comprehensive analysis of the polarization from undulators has been carried out by Kitamura.²³ The linear polarization of the undulator radiation is due to the symmetry of the electron trajectory within each period. The polarization can in fact be controlled by a deliberate breaking of this symmetry. Circularly polarized radiation can be produced by a helical undulator, in which the series of dipole magnets is arranged such that each period is rotated by a fixed angle with respect to the previous one. To generate variable polarization, one can use a pair of planar undulators oriented at right angles to each other. The amplitude of the radiation from these so-called crossed undulators is a linear superposition of two parts, one linearly polarized along the x direction and another linearly polarized along the y direction, x and y being orthogonal to the electron beam direction. By varying the relative phase of the two amplitudes by means of a variable-field magnet between the undulators, it is possible to modulate the polarization in an arbitrary way. The polarization can be linear and switched between two mutually perpendicular directions, or it can be switched between left and right circular polarization. For this device to work, it is necessary to use a monochromator with a sufficiently small band-pass, so that the wave trains from the two undulators are stretched and overlap. Also the angular divergence of the electron beam should be sufficiently small or the fluctuation in relative phase will limit the achievable degree of polarization. A planar undulator whose pole boundaries are tilted away from a right angle with respect to the axial direction can be used as a helical undulator if the electron trajectory lies a certain distance above or below the midplane of the device.

Transverse Spatial Coherence

As shown by Kim²⁴ and utilized in the brightness formulae given earlier, in wave optics the phase-space area of a radiation beam is given by the ratio of flux (F_0) to brightness (B_0). A diffraction-limited photon beam (no electron size or angular divergence contribution) occupies the minimum possible phase-space area. From Eqs. (32) to (37) this area is

$$(2\pi\sigma_r\sigma_{r'})^2 = (2\pi\epsilon)^2 = \left(\frac{\lambda}{2}\right)^2 \quad (45)$$

Thus, the phase space occupied by a single Gaussian mode radiation beam is $(\lambda/2)^2$, and such a beam is referred to as completely transversely coherent. It then follows that the transversely coherent flux of a radiation beam is

$$F_{\text{coherent}} = \left(\frac{\lambda}{2}\right)^2 B_0 \quad (46)$$

and the degree of transverse spatial coherence is

$$\frac{F_{\text{coherent}}}{F_0} = \left(\frac{\lambda}{2}\right)^2 \frac{B_0}{F_0} \quad (47)$$

Conversely, the number of Gaussian modes occupied by a beam is

$$\frac{F_0}{F_{\text{coherent}}} = \frac{F_0}{B_0(\lambda/2)^2} \quad (48)$$

Transverse spatial coherence is the quantity which determines the throughput of phase-sensitive devices such as Fresnel zone plates used for x-ray microscopy.

55.4 COHERENCE OF SYNCHROTRON RADIATION EMISSION IN THE LONG WAVELENGTH LIMIT

We now discuss coherent effects which depend upon the phase coherence between electric fields emitted by different electrons within a bunch. These effects naturally become stronger as the emission wavelength increases to values greater than the bunch length, the so-called long wavelength limit, which is generally in the far-infrared range or beyond for standard storage ring parameters. In the evolution from 3rd to 4th generation sources, the bunch length decreases, thereby pushing the coherent emission to shorter wavelengths, into the UV, VUV, and beyond. One way to generate shorter bunches is with single-pass devices such as linacs, wherein the electrons can be closer packed since relaxation processes which occur in storage rings may not have had time to develop.

In this section, we confine our discussion to coherent synchrotron radiation emission in the long wavelength limit. We describe two types of coherence: (i) longitudinal, or temporal (sometimes called multiparticle or super-radiant) coherence; and (ii) transverse or lateral coherence.

Temporal (Longitudinal/Multiparticle/Super-Radiant) Coherence

Under normal circumstances in a storage ring, the bunch length is considerably longer than the emitted wavelength, and since the electrons are randomly distributed, there is no phase correlation between the emitted electric fields from different electrons. For N_e particles emitting, then, the amplitude of the electric field is generated from the statistical noise, which is proportional to $N_e^{1/2}$. This implies that the power, which goes like to E^2 , is proportional to N_e .

In situations in which the wavelength is much longer than the bunch length, the phase differences between the electric fields of the electrons are small compared to the wavelength, and the field is N_e times that of a single electron. In this regime, the intensity is N_e^2 times that of a single electron, or N_e times greater than the incoherent process. This effect is very large because N_e is generally quite large: the number of electrons per bunch can easily be in the nanoCoulomb range, i.e., containing 10^{10} to 10^{11} electrons.

Long-wavelength coherent emission from bunches of relativistic charged particles was first described theoretically by Nodvick and Saxon²⁵ in 1954. For the situation of multiple electrons, the more general expression for the flux emitted by an electron *bunch* as a function of frequency (ω) and solid angle (Ω) is derived by extending the expressions derived earlier for a single electron [Eq. (5)], to a system of N_e electrons, thus:

$$\begin{aligned} \frac{d^2 F_{bm}(\omega)}{d\theta d\psi} &= 1.326 \times 10^{13} \text{ photons/sec/mrad}^2 / 0.1 \% \text{ bandwidth} \times [N_e [1 - f(\omega)] + N_e^2 f(\omega)] \\ &\times E^2 [\text{GeV}^2] I [A] (1 + \gamma^2 \psi^2)^2 \left(\frac{\omega}{\omega_c} \right)^2 \left[K_{2/3}^2(\xi) + \frac{\gamma^2 \psi^2}{1 + \gamma^2 \psi^2} K_{1/3}^2(\xi) \right] \end{aligned} \quad (49)$$

where the term $N_e^2 f(\omega)$ represents the coherent enhancement and includes the form factor $f(\omega)$, which is the Fourier transform of the normalized longitudinal particle distribution within the bunch, i.e.,

$$f(\omega) = \left| \int_{-\infty}^{\infty} e^{i\omega \hat{n} \cdot \vec{z}/c} S(z) dz \right|^2 \quad (50)$$

where $S(z)$ is the distribution function for particles in the bunch, measured relative to the bunch center.

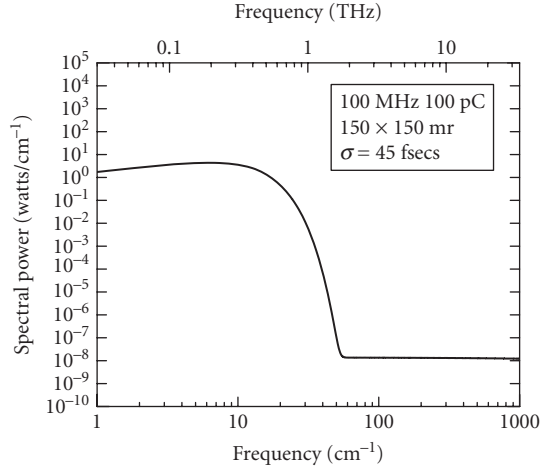


FIGURE 11 Long-wavelength coherent enhancement of synchrotron radiation spectra power for short electron bunches. Electron beam parameters used: 100-pC charge, 100-MHz repetition rate, 45-fs ($1 - \sigma$) bunch length. Emission angle range considered: 150 mrad (horiz.), 150 mrad (vert.).

If we assume that the electron bunch has a longitudinal Gaussian particle distribution of width σ_z , the form factor will be

$$f(\omega) = e^{-\left(\frac{\omega\sigma_z}{c}\right)^2} = e^{-4\pi^2\sigma_z^2\left(\frac{1}{\lambda}\right)^2} \quad (51)$$

where λ is the wavelength of the light at frequency ω .

The long-wavelength coherent enhancement is shown graphically for a chosen set of beam parameters in Fig. 11.

Transverse, or Lateral, Coherence

Transverse coherence was defined and described earlier in this chapter in the context of emission from undulators. In the long wavelength limit, even the emission from bending magnets becomes transversely coherent. In this section, we derive the interesting result that, in the long wavelength limit, the flux emitted into the natural opening angle (transverse coherent limit) is constant for all wavelengths, independent of storage ring parameters such as electron energy and bend radius.

1. Natural opening angle for synchrotron radiation for $\omega < \omega_c$:

$$\theta_{\text{rms}} = 0.8282 \left(\frac{\lambda(m)}{2\pi\rho(m)} \right)^{1/3} \text{ radians} \quad (52)$$

2. For $\omega < \omega_c$, the universal function

$$G_1\left(\frac{\omega}{\omega_c}\right) = \frac{\omega}{\omega_c} \int_{\omega/\omega_c}^{\infty} K_{5/3}(y) dy$$

from Eq. (6) and Fig. 3 can be approximated by^{8,11}

$$G_1\left(\frac{\omega}{\omega_c}\right) = \frac{1.33 \times 8 \times \pi}{9 \times \sqrt{3}} \left(\frac{\omega}{\omega_c}\right)^{1/3} = 2.144 \left(\frac{\omega}{\omega_c}\right)^{1/3}$$

3. Therefore, the formula for angle integrated flux (F) for $\omega < \omega_c$ into horizontal collection angle θ is

$$F = 5.27 \times 10^{16} \text{ photons/sec/0.1\% bandwidth} \times \theta [\text{rad}] E [\text{GeV}] I [\text{A}] \left(\frac{\omega}{\omega_c}\right)^{1/3} \quad (53)$$

4. Critical wavelength λ_c is given by

$$\lambda_c [m] = 5.59 \times 10^{-10} \frac{\rho [m]}{E^3 [\text{GeV}^3]} = \frac{2\pi c [\text{m/sec}]}{\omega_c [\text{sec}^{-1}]} \quad (54)$$

so that

$$\omega_c [\text{sec}^{-1}] = \frac{2\pi c [\text{m/sec}] E^3 [\text{GeV}^3]}{5.59 \times 10^{-10} \rho [m]} \quad (55)$$

5. Substituting Eqs. (52) and (55) into Eq. (53) and using $4 \times \theta_{\text{rms}}$ for the horizontal collection angle θ , we obtain:

$$F = 5.374 \times 10^{16} \text{ photons/sec/0.1\% bandwidth} \times E [\text{GeV}] I [\text{A}] \times \left(\frac{5.59 \times 10^{-10} \omega [\text{sec}^{-1}] \rho [m]}{2\pi c [\text{m/sec}] E^3 [\text{GeV}^3]} \right)^{1/3} \times 4 \times 0.8282 \left(\frac{2\pi c [\text{m/sec}]}{2\pi \rho [m] \omega [\text{sec}^{-1}]} \right)^{1/3} \quad (56)$$

or

$$F = 7.94 \times 10^{13} \text{ photons/sec/0.1\% bandwidth} \times I [\text{A}] \quad (57)$$

Figure 12 shows a flux plot based on an exact calculation for a synchrotron storage ring source with 3-GeV electron energy and 5-meter bend radius. This plot agrees with Eq. (57) from $\sim 2 \times 10^{-6} \omega_c$

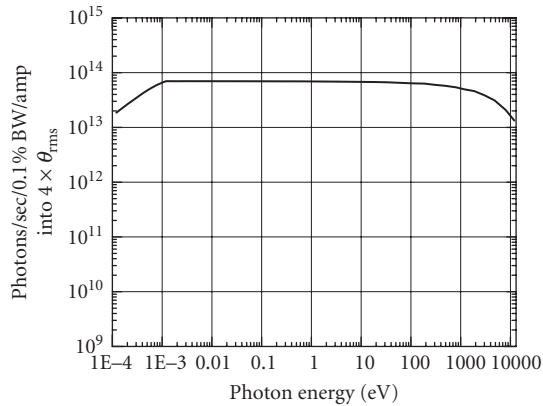


FIGURE 12 Synchrotron radiation spectral flux emitted into $4\theta_{\text{rms}}$ (horiz. $\times$ vert.) for a storage ring with 3-GeV electron energy and 5-meter bend radius.

to $\sim\omega_c$. The curve turns downward at the low-energy end owing to the finite size of the ring vacuum chamber. At the upper energy end it also turns downward owing to a failure of the approximation in Eq. (53) at and above ω_c .

For consistency as well as interest, another route to the result in Eq. (57) is via the undulator flux formula given by Eq. (29), with $K = 1$, $n = 1$, and $N = 1$. In this case, intended to approximate a bending magnet, $F_n(K) = 0.37$ one obtains, from Eq. (29), a flux value of $F = 7.99 \times 10^{13}$ photons/sec/0.1% bandwidth $\times I$ [A], in remarkable agreement with Eq. (57).

55.5 CONCLUSION

We have attempted to compile the formulae needed to calculate the flux, brightness, polarization (linear and circular), and power produced by the three standard storage ring synchrotron radiation sources: bending magnets, wigglers and undulators. Where necessary, these formulae have contained reference to the emittance (ϵ) of the electron beam, as well as to the electron beam size (σ) and its divergence (σ'). For all three types of sources, the source phase space area, i.e., the spatial and angular extent of the effective (real) source, is a convolution of its electron and photon components. For a more detailed description of these properties, see Ref. 26 and references therein.

55.6 REFERENCES

1. A. Liénard, "Champ électrique et magnétique, produit par une charge électrique concentrée en un point et animée d'un mouvement quelconque," *L'Eclairage Electrique* **16**:5 (1898).
2. J. P. Blewett, "Synchrotron Radiation—1873–1947," *Nucl. Instrum. Methods* **A266**:1 (1988).
3. P. C. Hartman, "Introductory Remarks," *Nucl. Instrum. Methods* **195**:1 (1982).
4. D. H. Tombouliau and P. L. Hartman, "Spectral and Angular Distribution of Ultraviolet Synchrotron Radiation from the 300-MeV Cornell Synchrotron," *Phys. Rev.* **102**:1423 (1956).
5. J. Schwinger, "On the Classical Radiation of Accelerated Electrons," *Phys. Rev.* **75**:1912 (1949).
6. J. D. Jackson, *Classical Electrodynamics* (Wiley, New York) 1975.
7. H. Winick, *Synchrotron Radiation Research* (Plenum Press, New York) Chapter 2, 1980.
8. A. Hofmann, *The Physics of Synchrotron Radiation*, (Cambridge University Press, Cambridge, England) 2004, ISBN 0 521 30826 7; A. Hofmann, *Phys. Rep.* **64**:253 (1980).
9. S. Krinsky, M. L. Perlman, and R. E. Watson, *Handbook of Synchrotron Radiation*, E. E. Koch (ed.) (North-Holland, Amsterdam) Chapter 2, 1983.
10. K. J. Kim, "Physics of Particle Accelerators," *AIP Proceedings* **184**:565 (1989).
11. A. A. Sokolov and I. M. Ternov, "Synchrotron Radiation" (Cambridge University Press, Cambridge, England) 1968, ISBN 0 521 30826 7.
12. V. O. Kostroun, "Simple Numerical Evaluation of Modified Bessel Functions $K_\nu(x)$ of Fractional Order and the Integral $\int_{-\infty}^{\infty} K_\nu(\eta)\delta\eta$," *Nucl. Instrum. Methods* **172**:371 (1980).
13. C. T. Chen, "Raytracing, Chopper and Guideline for Double-Headed Dragon Monochromators," *Rev. of Scientific Instr.* **63**:1229 (1992).
14. M. Born and E. Wolf, *Principles of Optics*, (Pergamon Press, London) 1964.
15. Lawrence Berkeley Laboratory Publication 643 Rev. 2, "X-Ray Data Booklet," Berkeley, Calif. (1989).
16. S. L. Hulbert and J. M. Weber, "Flux and Brightness Calculations for Various Synchrotron Radiation Sources," *Nucl. Instrum. Methods* **A319**:25 (1992).
17. R. P. Walker and B. Diviacco, "URGENT—A Computer Program for Calculating Undulator Radiation Spectral, Angular, Polarization and Power Density Properties," *Rev. Scientific Instr.* **63**:392 (1992).
18. R. P. Dejus and A. Luccio, "Program UR: General Purpose Code for Synchrotron Radiation Calculations," *Nucl. Instrum. Methods* **A347**:61 (1994).

19. O. Chubar and P. Elleaume, "Accurate and Efficient Computation of Synchrotron Radiation in the Near Field Region," *Proc. of the EPAC98 Conf.* **22–26**:1177–1179 (1998).
20. T. Tanaka and H. Kitamura, "SPECTRA: A Synchrotron Radiation Calculation Code," *J. Synch. Rad.* **8**:1221 (2001).
21. M. R. Howells and B. M. Kincaid, *LBL Report #34751* (1993).
22. K. J. Kim, "Angular Distribution of Undulator Power for an Arbitrary Deflection Parameter K," *Nucl. Instrum. Methods* **A246**:67 (1986).
23. H. Kitamura, "Polarization of Undulator Radiation," *Jpn. J. Appl. Phys.* **19**:L185 (1980).
24. K. J. Kim, "Brightness, Coherence and Propagation Characteristics of Synchrotron Radiation," *Nucl. Instrum. Methods* **A246**:71 (1986). See also earlier work of A. M. Kondratenko and A. N. Skrinsky, "Use of Radiation of Electron Storage Rings in X-Ray Holography of Objects," *Opt. Spectroscopy* **42**:189 (1977) and F. Zernike, "The Concept of Degree of Coherence and Its Application to Optical Problems," *Physica V* **785** (1938).
25. J. S. Nodvick and D. S. Saxon, "Suppression of Coherent Radiation by Electrons in a Synchrotron," *Phys. Rev.* **96**:180 (1954).
26. S. L. Hulbert and G. P. Williams, "Synchrotron Radiation Sources," in *Vacuum Ultraviolet Spectroscopy I*, J. A. R. Samson and D. L. Ederer (eds.), *Experimental Methods in the Physical Sciences* **31**:1 (Academic Press, San Diego, Calif.) 1998.

This page intentionally left blank

LASER-GENERATED PLASMAS

Alan Michette

*King's College London
United Kingdom*

56.1 INTRODUCTION

Laser plasmas are produced by focusing a pulsed laser beam, e.g., Nd:YAG at 1.064 μm , possibly frequency multiplied, or KrF excimer at 249 nm, onto a target, typically a tape,¹ liquid droplet or jet² or a gas-jet.³ The required irradiance, that is the focused intensity per unit area per unit time, is in the range $\sim 10^{17}$ to 10^{19} W m^{-2} , which heats the target material to $\sim 10^6 \text{ K}$ and thus ionizes it to produce the plasma. The need for high irradiance means that high beam energies, small focused spot sizes and short pulse lengths must be used. For repetitive systems¹ pulse energies in the range $\sim 10 \text{ mJ}$ to 1 J are used with focal spot sizes of $\sim 10 \mu\text{m}$. For single pulse systems, using much higher pulse energies, focal spot sizes can be much larger. Normally pulse lengths are in the range of a few picoseconds to several nanoseconds, but lasers with much shorter pulse lengths, in the tens of femtoseconds range, have also been used.⁴

For picosecond–nanosecond systems the emission is generated from the ionized material. The spectral characteristics depend mainly on the target material, with the proviso that the irradiance must be sufficient to produce the ionic state required to give a particular spectral feature. The use of tape targets, compared to liquids or gases, allows a wider range of materials to be used, but the effects of increased particulate debris emission must be alleviated by a low-pressure buffer gas. For light materials it is common to almost strip the atoms and emission is then from hydrogen- and helium-like ions. The spectrum largely consists of characteristic line emission, with small contributions from bremsstrahlung and recombination radiation. The lines typically have bandwidths $\Delta\lambda/\lambda \sim 10^{-4}$.

For a repetitive source operated at $\sim 100 \text{ Hz}$, the brilliance in a particular line, e.g., the H-like carbon Lyman- α line at a wavelength of 3.37 nm, can be comparable to that of a second-generation synchrotron, albeit only at this wavelength. Use of a heavier Z target, such as copper or gold, means that the ions are less fully stripped and the emission from many closely spaced ionic energy levels merges into a quasi-continuum. Such targets also give rise to more bremsstrahlung. Although the overall emission is higher than that from light targets it is considerably lower than that of a synchrotron, and the peak emission can be much less than that from light materials.

56.2 CHARACTERISTIC RADIATION

Characteristic radiation arises as a direct consequence of the quantum nature of the plasma ions. Electrons bound in ionic systems can only occupy discrete energy levels, with only one electron in each quantum state—the Pauli Exclusion Principle. In the ground state all the lower energy levels will be filled. If, by some process, an electron is removed from a filled level, either by promoting it to a higher energy level (excitation) or by completely removing it from the ion (ionization), then the empty level may be filled by an electron from a higher occupied state, which loses energy by emitting radiation. Because of the discreteness of the energy levels the transition between two defined states will always give a photon of the same energy, which is characteristic of the ion involved since the energies of the bound states depend on the nuclear charge and on the screening effects of other electrons.

Energies of Spectral Lines

For an ion with just one electron—a hydrogenic ion—the energies of the electron states are

$$E_n = -\frac{1}{2n^2} \frac{\mu}{\hbar^2} \left(\frac{Ze^2}{4\pi\epsilon_0} \right)^2 \quad (1)$$

where the negative sign indicates a bound state, n is the principal quantum number, Z is the nuclear charge, and μ is the reduced mass of the electron-nucleus system

$$\mu = \frac{m_e M}{m_e + M} \quad (2)$$

M being the mass of the nucleus. The energy of the photon emitted when an electron makes a transition from an initial level n_i to a final level n_f is thus

$$E = \hbar\omega = \frac{1}{2} \frac{\mu}{\hbar^2} \left(\frac{Ze^2}{4\pi\epsilon_0} \right)^2 \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right) \quad (3)$$

As is well known, for atomic hydrogen the transitions result in long wavelength radiation, for example, the Lyman series in the ultraviolet, corresponding to $n_f = 1$. For hydrogenic ions, Eq. (3) shows that the spectral line energies are a factor of Z^2 higher than those of atomic hydrogen (and the wavelengths are correspondingly shorter); for hydrogenic carbon, for example, the Lyman- α line is at a wavelength of 3.37nm, compared to 121.6nm in atomic hydrogen.

Spectral Line Intensities

Characteristic radiation as described here corresponds to spontaneous emission, for which the transition rate, in the electric-dipole approximation, is⁵

$$R_{if} = \frac{\omega_{if}^3}{3\pi\epsilon_0\hbar c^3} |\langle \psi_f | \mathbf{er} | \psi_i \rangle|^2 \quad (4)$$

where $\langle \psi_f | \mathbf{er} | \psi_i \rangle$ is the matrix element describing the transition from the initial state with wave function ψ_i via the electric dipole operator $\mathbf{er}$ to the final state ψ_f . The intensity of the emission is given by multiplying Eq. (4) by the energy $\hbar\omega_{if}$ of each emitted photon, the statistical weight g_i of the initial state, which is the number of degenerate configurations that can give rise to the transition,

and the probability F_f that an electron has been removed from the level to which the system eventually relaxes. Thus the intensity of a characteristic line is given by

$$I_{if} = \frac{g_i F_f \omega_{if}^4}{3\pi \epsilon_0 c^3} |\langle \psi_f | e \mathbf{r} | \psi_i \rangle|^2 \quad (5)$$

If the matrix element in the electric dipole transition rate is zero, then the transition cannot take place. This is embodied in the familiar electric dipole selection rules,

$$\Delta l = \pm 1 \quad \Delta m = 0, \pm 1 \quad \Delta j = 0, \pm 1 \quad \text{with } j = 0 \rightarrow 0 \text{ forbidden} \quad (6)$$

where l is the orbital angular momentum, m is the magnetic quantum number and j is the total angular momentum.

For higher-order transitions the selection rules are more complicated. For electric quadrupole radiation, for example,

$$\Delta l = 0, \pm 2 \quad \Delta j = 0, \pm 1, \pm 2 \quad \text{with } 0 \rightarrow 0 \quad \frac{1}{2} \rightarrow \frac{1}{2} \text{ \& } 0 \leftrightarrow 1 \text{ forbidden} \quad (7)$$

and the parities of the initial and final states must be the same. For magnetic dipole transitions the parity must also remain unchanged, with

$$\Delta l = 0 \quad \Delta j = 0, \pm 1 \quad \text{with } 0 \rightarrow 0 \text{ forbidden} \quad (8)$$

The relative intensities in different spectral lines can, in principle, be calculated from Eq. (5), so long as all states with the same energy difference $E_i - E_f$ are taken into account. The intensities are characterized by the *oscillator strength*

$$\phi_{if} = \frac{2m_e \omega_{if}}{3\hbar} |\langle \psi_f | \mathbf{r} | \psi_i \rangle|^2 \quad (9)$$

where, for hydrogenic systems,

$$\sum_f \phi_{if} = 1 \quad (10)$$

the sum being over all final states that contribute to emission or absorption of a photon of energy $\hbar \omega_{if}$, including the continuum if appropriate. In terms of the oscillator strength, the transition rate, Eq. (4), becomes

$$R_{if} = \frac{2\hbar \alpha}{m_e c^2} \omega_{if}^2 |\phi_{if}| \quad (11)$$

where α is the fine structure constant and the modulus of ϕ_{if} is used since, by convention, the oscillator strength is defined as negative for emission ($E_f < E_i$) and positive for absorption. As hydrogenic wave functions can be calculated exactly, the oscillator strengths and transition rates are straightforward to determine.

For ions with Z electrons the sum rule, Eq. (10), becomes

$$\sum_f \phi_{if} = Z \quad (12)$$

and the oscillator strengths can only be estimated since approximate methods must be used to determine the wave functions. The oscillator strengths depend on the quantum numbers n , l , and m of the initial and final states, particularly on the magnetic quantum number m and hence on the

polarization of the radiation. It is often useful to use an oscillator strength averaged over the initial and final state magnetic quantum numbers, since this is independent of the polarization,

$$\bar{\phi}_{if}(n_i, l_i, n_f, l_f) = \frac{1}{2l_i + 1} \sum_{m_i = -l_i}^{l_i} \sum_{m_f = -l_f}^{l_f} \phi_{if}(n_i, l_i, m_i, n_f, l_f, m_f) \quad (13)$$

The sum rule, Eq. (10) or (12), is still obeyed for the average oscillator strength. The corresponding relative line intensities depend on the statistical weights $2(2l + 1)$ of the initial state and the spontaneous emission coefficient as well as on the oscillator strengths.⁶

For line emission from hydrogenic ions, as the atomic number increases relativistic effects become increasingly important.⁷ This causes the average oscillator strengths to decrease; for a low- Z element such as carbon the decreases are only $\sim 0.1\%$, for a medium- Z element such as copper they are about 3%, while for high- Z elements such as gold or uranium they are about 20% and 30%, respectively.

Spectral Line Shapes and Widths

The discussion of line emission so far has implicitly assumed that each spectral line has an exactly defined energy, with corresponding frequency and wavelength. However, in practice there is always a spread of energies in each line as the result of several factors.⁸ Since the spread depends upon the plasma properties, such as temperature and density, measurement of line shapes and widths provides a useful diagnostic of the plasma state.

Natural Line Width Natural line widths arise since atomic and ionic levels do not have precisely determined energies, because interactions with real or virtual photons result in transitions between levels. Hence the levels have finite lifetimes τ and so, as a result of the uncertainty principle, there is an energy spread

$$\Delta E \sim \frac{h}{\tau} \quad (14)$$

with a corresponding frequency spread

$$\Delta \nu \sim \frac{1}{\tau} \quad (15)$$

For transitions from a state i to a state j , the lifetime τ_{ij} is given by

$$\tau_{ij} = \frac{2}{\sum_{k < i} A_{ik} + \sum_{k < j} A_{jk}} \quad (16)$$

where A_{ij} is the spontaneous emission coefficient for the transition $i \rightarrow j$ and the sums are over all spontaneous transitions from the initial and final states, respectively. If the final state is the ground state then the second summation is equal to zero.

The amplitude $A(t)$ of a state with lifetime τ decays exponentially with time

$$A(t) \propto e^{-t/\tau} e^{2\pi i \nu_0 t} \quad (17)$$

where ν_0 is the central transition frequency. The shape of the spectral line is given by the square of the Fourier transform of the amplitude, resulting in a Lorentzian distribution

$$I(\nu) = I(\nu_0) \frac{1}{1 + [2\pi(\nu - \nu_0)\tau]^2} \quad (18)$$

where $I(\nu)$ is the intensity at frequency ν . The full width at half maximum (FWHM) of this distribution is

$$\Delta\nu_L = \frac{1}{\pi\tau} \quad (19)$$

In terms of wavelength, which is often how spectra are presented, the line shape is still (obviously) Lorentzian and the FWHM is

$$\Delta\lambda = \frac{\lambda_0^2}{\pi\tau c} \quad (20)$$

Doppler Broadening If the emitting ions are in thermal motion, there will be a Doppler shift $\Delta\nu_D$ on the frequency of the emitted radiation. In a plasma, the particles are normally nonrelativistic,

$$\Delta\nu_D = |\nu - \nu_0| = \nu_0 \frac{v}{c} \quad (21)$$

where v is the particle speed along the line of sight. Radiation from particles with a velocity component away from an observer will thus be red shifted, while a velocity component toward the observer causes a blue shift. For particles in random motion the spectral lines are therefore broadened and, if the velocity distribution $f(v)$ is Maxwellian with a mean speed v_m , then

$$f(v) \propto e^{-v^2/2v_m^2} \quad (22)$$

The resulting line profile, found simply by making the substitution $v \rightarrow c(\nu/\nu_0 - 1)$ in Eq. (22), is

$$I(\nu) = I(\nu_0) \exp \left[-\frac{(\nu - \nu_0)^2 c^2}{2v_m^2 \nu_0^2} \right] \quad (23)$$

which is a Gaussian distribution with FWHM

$$\Delta\nu_G = \nu_0 \frac{v_m}{c} 2(2\ln 2)^{1/2} \quad (24)$$

Clearly, the broadening is larger if the mean speed is higher, so that Doppler broadening can dominate in hot plasmas and is a measure of the plasma temperature. In terms of the wavelength, Eqs. (23) and (24) become

$$I(\lambda) = I(\lambda_0) \exp \left[-\frac{(\lambda - \lambda_0)^2 c^2}{2v_m^2 \lambda_0^2} \right] \quad (25)$$

and

$$\Delta\lambda_G = \lambda_0 \frac{v_m}{c} 2(2\ln 2)^{1/2} \quad (26)$$

Pressure Broadening In plasmas neighbouring particles affect the emission profiles of spectral lines. There are several contributions to this *pressure broadening*. Collisions, of either ions or electrons, with the emitting ion can change the ionic state, effectively shortening the lifetimes of the energy levels involved in the emission. By a similar argument to that in the discussion of natural broadening this leads to a Lorentzian profile but with a larger width. The width is governed by the

mean time between collisions, and so this process can be dominant in dense plasmas, and is a measure of the plasma density.

Nearby ions and electrons also generate electric and magnetic fields, which change the ionic energy levels of the emitting ion and hence contribute to the line width. The major contributor in plasmas is normally the electric field which, in general, produces a shift in the energy levels proportional to the square of the electric field—the quadratic Stark effect. This leads to asymmetric broadening and hence to shifts in the central frequencies of spectral lines. In hydrogenic ions the effect is linear, except in the ground state where the linear effect disappears, and hence the broadening is symmetric with no shift of the central frequency. The effect on line shapes, apart from in hydrogenic ions, is very complicated and requires each case to be treated individually,⁸ beyond the scope of the present discussion. However, it is possible to use simple arguments to obtain an idea of how line profiles and widths scale with plasma properties.⁹

In the nearest neighbour approximation, an ion in a range of distances dr about r from another ion experiences a high field; the probability of it being affected is proportional to the volume $4\pi r^2 dr$. The frequency shift $|v - v_0|$ caused by this is proportional to the field strength, i.e., to r^{-2} , so that

$$\begin{aligned} |v - v_0| &\propto r^{-2} & r &\propto |v - v_0|^{-1/2} & dr &\propto |v - v_0|^{-3/2} dv \\ r^2 dr &\propto |v - v_0|^{-5/2} dv \end{aligned} \quad (27)$$

and the resulting line profile is

$$I(v) \propto |v - v_0|^{-5/2} \quad (28)$$

This clearly breaks down for frequencies close to the central frequency, since otherwise the intensity $I(v_0)$ would become infinite. The line profile, Eq. (28), was derived using the nearest neighbour approximation which is only valid for the distance below which the electric field of the perturbing ion dominates over that due to other ions. This distance is determined by the volume containing about one ion, i.e., $r \approx [3/(4\pi n_i)]^{1/3}$, where n_i is the ion number density. Thus the line width, which is proportional to r^{-2} , is

$$\Delta v_s \propto n_i^{2/3} \quad (29)$$

Combinations of Line Broadening Mechanisms The preceding discussions show that, if the line broadening mechanism can be identified, properties of the plasma, such as density and temperature, can be determined from measured line profiles and widths. In general, however, the emission lines will be affected by more than one process, and thus combinations of the various line broadening mechanisms must be considered. In addition, the measured profiles will be modified by the instrument used to obtain the information; this may be dominant and will not, in general, provide a well defined profile.

For two independent effects, with profiles $I_1(v)$ and $I_2(v)$ leading to FWHM line widths Δ_1 and Δ_2 , the combined profile is the convolution

$$I(v) = \int I_1(v - v') I_2(v') dv' \quad (30)$$

If the two profiles are the same the combined profile has the same shape; for Lorentzian distributions the combined width is

$$\Delta_L = \Delta_1 + \Delta_2 \quad (31)$$

while for Gaussian distributions it is

$$\Delta_G = (\Delta_1^2 + \Delta_2^2)^{1/2} \quad (32)$$

If the two profiles are not the same then there is no general simple form for the convolved line shape. The combination of a Lorentzian and a Gaussian profile leads to the Voigt function

$$I_V(q) \propto \int_{-\infty}^{\infty} \frac{e^{-y^2}}{(x-y)^2 + a^2} dy \quad (33)$$

where

$$q = \frac{2(\ln 2)^{1/2}}{\Delta v_G} (v - v_0) \quad \text{and} \quad a = (\ln 2)^{1/2} \frac{\Delta v_L}{\Delta v_G} \quad (34)$$

Equation (33) does not have an analytic solution; the integration has to be done numerically for each frequency involved and so is very expensive computationally. An alternative to the Voigt function, which has been used in many other fields¹⁰⁻¹⁷ but not widely in the analysis of x-ray emission from plasmas, is the Pearson VII function.¹⁸⁻²⁰ This has the form

$$I_p(\lambda) = I(\lambda_0) [1 + P(\lambda)^2]^{-M} \quad (35)$$

where

$$P(\lambda) = \frac{2(\lambda - \lambda_0) \sqrt{2^{1/M} - 1}}{\Delta_p} \quad (36)$$

In Eqs. (35) and (36) M is the Pearson parameter and Δ_p is the FWHM. If $M = 1$ the distribution is clearly Lorentzian, while as M becomes large the distribution is indistinguishable from a Gaussian. For intermediate values of M the Pearson VII distribution is more general than the Lorentzian and Gaussian functions due to the extra free parameter (M), and it makes no assumptions about the line profile. It can also provide a good approximation to a Voigt function, compared to which it is much less computationally intensive since it does not require numerical integration. For $M < 1$ the distribution has broader wings than a Lorentzian function; this corresponds to no known broadening mechanism but even then the Pearson VII distribution can still provide a convenient functional form for line profile fitting.

A disadvantage of the Pearson VII distribution is that there are no direct relationships between the plasma properties and the fit parameters Δ_p and M . The widths of the Lorentzian and Gaussian distributions, which can be extracted from fits to the Voigt function can, however, be related to plasma properties such as density and temperature. But if the spectra are time and spatially integrated the plasma parameters cannot readily be extracted, so that there is no advantage in using the Lorentzian, Gaussian, or Voigt profiles. Unlike the Voigt function, the Pearson VII distribution can be integrated analytically to provide the total line intensity

$$I_p^{\text{tot}} = \frac{\Delta_p I_0 \sqrt{\pi}}{2\sqrt{2^{1/M} - 1}} \frac{\Gamma\left(M - \frac{1}{2}\right)}{\Gamma(M)} \quad (37)$$

where Γ is the gamma function,

$$\Gamma(g) = \int_0^{\infty} y^{g-1} e^{-y} dy \quad (38)$$

A common characteristic of the Lorentzian, Gaussian, Voigt, and Pearson VII profiles is that they are symmetrical about the mean. However, some line-broadening mechanisms, e.g., Stark broadening, can result in asymmetric profiles. In this case a suitable fitting function is the Pearson IV distribution^{18,20}

$$I_{PIV}(\lambda) = I_{PVII}(\lambda) \exp[-a \tan^{-1} P(\lambda)] \quad (39)$$

which clearly reduces to the Pearson VII distribution when a , the asymmetry parameter, is zero. The disadvantages of the Pearson IV distribution are the extra parameter required in the fit, and the total

intensity in the line has to be obtained by numerical integration since the distribution is not analytically integrable. Hence the fitting is more computationally expensive than for the Pearson VII distribution, but the Pearson IV function may be useful in analysis of high spectral resolution data where Stark broadening may be observed.

Reabsorption of Spectral Lines The discussion of spectral lines has so far assumed that the emission profiles are not modified by reabsorption within the plasma. This may not always be the case, especially in optically thick plasmas.

The linear absorption coefficient α_λ , which is a function of the wavelength of the radiation, describes the amount of attenuation over a thickness x of a medium

$$I_\lambda(x) = I_\lambda(0)e^{-\alpha_\lambda x} \quad (40)$$

The transmission of radiation can be the same through a tenuous extended plasma or through a dense compact plasma. It may not be necessary, or possible, for a remote observer to distinguish these, in which case the optical thickness (or depth) τ_λ is a useful concept; this is defined as the absorption coefficient (which may depend upon position) integrated between two points x_1 and x_2 ,

$$\tau_\lambda = \int_{x_1}^{x_2} \alpha_\lambda dx \quad (41)$$

An optically thin medium has $\tau_\lambda < 1$ while for an optically thick medium $\tau_\lambda \gg 1$; a plasma may be optically thin for some wavelengths and optically thick for others.

In an optically thick plasma, some of the radiation in a spectral line may be reabsorbed, which can alter the line profile and apparent FWHM. If there is no reabsorption, a beam of radiation passing through the plasma will gain in intensity due to spontaneous and stimulated emission. However, this cannot continue indefinitely since the intensity cannot become larger than that of a blackbody at the effective plasma temperature. Hence, as the blackbody level is approached, absorption must balance the emission and the intensity saturates. Since a spectral line is most intense at its central wavelength the emitted intensity close to the line center is the first to be limited. Thus the line becomes flattened near to its center, which modifies the profile and increases the apparent line width.

In a nonuniform plasma, for example, if the temperature decreases near to the edge—which is often the case—the observed central line intensity can be decreased since the observed emission near to the line center is from the cooler plasma, which has a lower blackbody intensity. This is known as line inversion and can lead to confusion in the interpretation of plasma spectra, as it may appear as two closely spaced lines.

56.3 BREMSSTRAHLUNG

The preceding discussion of line emission refers to the situation in which an electron bound in an ion makes a transition to another bound state of the same ion (bound-bound radiation). In this section the transition of a free electron (or collection of electrons) to another free state is considered—free-free radiation, caused by the acceleration of the electron in the Coulomb field of an ion. According to Maxwell's equations, accelerated charges emit electromagnetic radiation and so they lose energy; the emitted radiation is thus called “braking radiation” or *bremsstrahlung*. An important difference between bound-bound and free-free radiation is that free electrons do not have quantised energies, and hence the emission spectrum is a continuum.

Although in practice many electrons will encounter many ions, it is instructive to consider, in the first instance, the interaction of a single electron with a single ion. In this case the radiated energy in the frequency range $d\omega$ is given by

$$\left. \frac{dW}{d\omega} \right|_{\omega=\omega_0} \approx \frac{Z^2 e^6}{(4\pi\epsilon_0)^3} \frac{8}{3\pi m_e^2 c^3} \frac{1}{v^2 b^2} \quad (42)$$

where v is the electron speed, b is the impact parameter (the perpendicular distance between the ion and the initial electron direction) and ω_0 is the central frequency of the emitted radiation. If the single electron collides with a random assembly of ions, with a number density $n_i \text{ m}^{-3}$, the power spectrum is given by multiplying Eq. (42) by $n_i v$ and integrating over the impact parameter

$$\frac{dP}{d\omega} = n_i v \int_{b_{\min}}^{b_{\max}} \frac{dW}{d\omega} 2\pi b db \approx \frac{Z^2 e^6}{(4\pi\epsilon_0)^3} \frac{16n_i}{3\pi m_e^2 c^3 v} \int_{b_{\min}}^{b_{\max}} \frac{1}{b} db = \frac{Z^2 e^6}{(4\pi\epsilon_0)^3} \frac{16n_i}{3\pi m_e^2 c^3 v} \ln \frac{b_{\max}}{b_{\min}} \quad (43)$$

Naïvely, Eq. (43) might suggest that a single electron encountering a group of ions would radiate infinite power, with a minimum impact parameter $b_{\min} = 0$ (head-on collision) and a maximum of $b_{\max} = \infty$. This is clearly incorrect; the explanation is that $b_{\max}$ is determined by shielding—since the electron is only accelerated if the nuclear charge is not shielded by surrounding particles—and $b_{\min}$ is governed by the Uncertainty Principle, since the electron is only close to an ion for a finite time and so the impact parameter cannot be determined precisely.

In the nonquantum, nonrelativistic limit the ratio $b_{\max}/b_{\min}$ may be obtained in terms of the impact parameter b_{90} for which the electron is scattered through an angle of 90° . For the case $b \gg b_{90}$ this leads to

$$\frac{b_{\max}}{b_{\min}} = \frac{2\gamma v}{\omega b_{90}} \quad (44)$$

where γ is Euler's constant

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \ln n \right) = 0.5772157... \quad (45)$$

Equation (43) may generally be written as

$$\frac{dP}{d\omega} = \frac{Z^2 e^6}{(4\pi\epsilon_0)^3} \frac{16n_i}{3m_e^2 c^3 v} \frac{\pi}{\sqrt{3}} G \quad (46)$$

where, for $b \gg b_{90}$,

$$G = \frac{\sqrt{3}}{\pi} \ln \frac{2\gamma v}{\omega b_{90}} \quad (47)$$

Equation (46) can be used for other cases, including when quantum and relativistic effects are important (neither of which are normally appropriate for laser-generated plasmas used as x-ray sources), so long as an appropriate choice is made for the Gaunt factor, G , which never differs substantially from unity; in many cases using $G = 1$ is a sufficient approximation. Several authors have calculated Gaunt factors for bremsstrahlung under a range of conditions, including relativistic, and extensive tabulations or graphical representations are available.^{21,22}

Bremsstrahlung from a Collection of Electrons In a plasma the electron density n_e is usually larger than the ion density n_i , and so bremsstrahlung from a collection of electrons must be considered. The radiated power must then be integrated over the electron velocity distribution $f(\mathbf{v})$ and the spectral power emitted into the whole solid angle per unit angular frequency is then

$$4\pi w(\omega) = \frac{Z^2 e^6}{(4\pi\epsilon_0)^3} \frac{16\pi n_i}{3\sqrt{3}m_e^2 c^3} \int \frac{G(\omega, \mathbf{v})}{v} f(\mathbf{v}) d^3 \mathbf{v} \quad (48)$$

For an isotropic Maxwellian velocity distribution, as usually assumed,

$$4\pi w(\omega) = n_e n_i \frac{Z^2 e^6}{(4\pi \epsilon_0)^3} \frac{16\pi}{3\sqrt{3} m_e^2 c^3} \left(\frac{2m_e}{\pi T_{eV}} \right)^{1/2} e^{-\hbar\omega/T_{eV}} \bar{G}(\omega, T_{eV}) \quad (49)$$

where T_{eV} is the plasma temperature in electronvolts and $\hbar\omega$ is the emitted photon energy. Values of the Maxwell-averaged Gaunt factor, $\bar{G}(\omega, T_{eV})$ are typically ~ 1 , and this is normally sufficient to give an accurate enough approximation of the bremsstrahlung intensity.

56.4 RECOMBINATION RADIATION

In some circumstances in plasmas recombination radiation, emitted when an electron is captured by an ion, can be a dominant part of the spectrum, competing with and modifying the bremsstrahlung component. In recombination radiation the emitted energy is clearly greater than the free-electron kinetic energy, and the final-state electron is bound with a discrete energy spectrum E_n . For hydrogenic ions of charge Ze , the emitted spectral power is given by

$$4\pi w(\omega) = n_e n_i \frac{Z^2 e^6}{(4\pi \epsilon_0)^3} \frac{16\pi}{3\sqrt{3} m_e^2 c^3} \left(\frac{2m_e}{\pi T_{eV}} \right)^{1/2} e^{-\hbar\omega/T_{eV}} F_n \quad (50)$$

where

$$F_n = \frac{Z^2 e^4 m_e}{(4\pi \epsilon_0 \hbar)^2 n^3} \frac{1}{T_{eV}} G_n e^{-E_n/T_{eV}} \quad (51)$$

and G_n is the Gaunt factor for capture of the free electron to the energy level with principal quantum number n and energy E_n . Equation (50) is the same as that for bremsstrahlung from a collection of electrons, Eq. (49), except that F_n has replaced the Maxwell-averaged Gaunt factor $\bar{G}(\omega, T_{eV})$. The ratio of powers emitted in recombination radiation and bremsstrahlung is thus

$$\frac{F_n}{\bar{g}(\omega, T_{eV})} = \frac{Z^2 e^4 m_e}{(4\pi \epsilon_0 \hbar)^2 n^3} \frac{1}{T_{eV}} \frac{G_n}{\bar{g}(\omega, T_{eV})} e^{-E_n/T_{eV}} \quad (52)$$

It must be remembered that this is only for hydrogenic ions, but Eq. (52) gives an idea of the order of magnitude of recombination radiation compared to bremsstrahlung. Putting in the values of the constants, and recalling that the ratio of the Gaunt factors will be of order unity,

$$\frac{F_n}{\bar{G}(\omega, T_{eV})} \approx \frac{27.2 Z^2}{n^3 T_{eV}} e^{13.6 Z^2 / n^2 T_{eV}} \quad (53)$$

which shows that recombination radiation can be many orders of magnitude more intense than bremsstrahlung for low temperature plasmas and for recombination to inner ionic levels, i.e., small n .

56.5 REFERENCES

1. I. C. E. Turcu, I. N. Ross, P. Tenda, C. W. Wharton, R. A. Meldrum, H. Daido, M. S. Schulz, et al., "Picosecond Excimer Laser-Plasma X-Ray Source for Microscopy, Biochemistry, and Lithography," *Proc. SPIE* **2015**:243–260 (1993).
2. H. M. Hertz, L. Rymell, M. Berglund, and L. Malmqvist, "Debris-Free Liquid-Target Laser-Plasma Soft X-Ray Source for Microscopy and Lithography," in *X-Ray Microscopy and Spectromicroscopy*, Berlin: Springer, 1998, pp. V-3–V-13.

3. H. Fiedorowicz, A. Bartnik, H. Szczurek, H. Daido, N. Sakaya, V. Kmetik, Y. Kato, et al., "Investigation of Soft X-Ray Emission from a Gas Puff Target Irradiated with a Nd:YAG Laser," *Opt. Comm.* **163**:103–114 (1999).
4. C. Reich, P. Gibbon, I. Uschmann, and E. Förster, "Yield Optimization and Time Structure of Femtosecond Laser Plasma K_α Sources," *Phys. Rev. Lett.* **84**:4846–4849 (2000).
5. B. H. Bransden and C. J. Joachain, *Physics of Atoms and Molecules* 2nd ed., London: Longmans, 2002, pp. 195–197.
6. H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms*, Berlin: Springer, 1957, p. 265.
7. Y. G. Pal'chikov, "Relativistic Transition Probabilities and Oscillator Strengths in Hydrogen-Like Atoms," *Physica Scripta* **57**:581–593 (1998).
8. R. G. Breene, *The Shift and Shape of Spectral Lines*, Oxford: Pergamon Press, 1962.
9. I. H. Hutchinson, *Principles of Plasma Diagnostics*, Cambridge: Cambridge University Press, 1987, pp. 218–221.
10. H. Toraya, M. Yoshimura, and S. Somiya, "A Computer Program for the Deconvolution of X-Ray Diffraction Profiles with the Composite of Pearson Type VII Functions," *J. Appl. Cryst.* **16**:653–657 (1983).
11. C. P. Lafrance, J. Debigare, and R. E. Prudhomme, "Study of Crystalline Orientation in Drawn Ultra-High Molecular Weight Polyethylene Films," *J. Polym. Sci. Pol. Phys.* **31**:255–264 (1993).
12. M. Oetzel and G. Heger, "Laboratory X-Ray Powder Diffraction: A Comparison of Different Geometries with Special Attention to the Usage of the $\text{Cu } K_\alpha$ Doublet," *J. Appl. Cryst.* **32**:799–807 (1999).
13. A. Santoro, R. J. Cava, D. W. Murphy, and R. S. Roth, "Use of the Pearson Type VII Distribution in the Neutron Profile Refinement of the Structures of LiReO_3 and Li_2ReO_3 ," *Proc. AIP* **89**:162–165 (1982).
14. D. R. Noakes, "Magnetic Field Distributions at Interstitial Sites in Nondilute Alloys," *Phys. Rev. B* **44**:5064–5072 (1991).
15. V. Villani, R. Pucciariello, and G. Ajroldi, "Calorimetric Study of the Room-Temperature Transitions of Polytetrafluoroethylene—The Influence of Thermal History," *J. Polym. Sci. Pol. Phys.* **29**:1255–1259 (1991).
16. R. Oven, D. G. Ashworth, and M. D. J. Bowyer, "Formulas for the Distribution of Ions Under an Ideal Mask," *J. Phys. D* **25**:1235–1237 (1992).
17. W. Wulfhekel and J. M. Cadogan, "Mössbauer Line-Sharpening—Application to Magnetically Split Spectra," *Hyperfine Interact.* **92**:1195–1202 (1994).
18. W. P. Elderton and N. L. Johnson, *Systems of Frequency Curves*, London: Cambridge University Press, 1969.
19. A. G. Michette and S. J. Pfauntsch, "Laser Plasma X-Ray Line Spectra Fitted Using the Pearson VII Function," *J. Phys. D Appl. Phys.* **33**:1186–1190 (2000).
20. S. J. Pfauntsch, *Developments in Soft X-Ray Laboratory Systems for Microscopy and Cellular Probing*, Ph.D. Thesis (London University), 2001.
21. P. J. Brussard and H. C. van der Hulst, "Approximation Formulas for Nonrelativistic Bremsstrahlung and Average Gaunt Factors for a Maxwellian Electron Gas," *Rev. Mod. Phys.* **34**:507–520 (1962).
22. T. R. Carson, "Coulomb Free-Free Gaunt Factors," *Astron. Astrophys.* **189**:319–324 (1988).

This page intentionally left blank

PINCH PLASMA SOURCES

Victor Kantsyrev

*Physics Department
University of Nevada
Reno, Nevada*

57.1 INTRODUCTION

In pinch systems, the source of x-ray radiation is the plasma that is produced by an electrical current discharge and compressed by the associated magnetic field. Numerous forms of pinch plasma are used as radiation sources (Z-pinch,¹ θ -pinch,² etc.). The most common type of pinch plasma source is the Z-pinch. In the Z-pinch system, plasma is created by applying a high-voltage pulse from a fast generator between an anode and cathode; its name refers to the direction of the earliest experimental facility, where the current flowed down the vertical discharged tube, the z axis on a normal mathematical diagram. Both electrodes have cylindrical geometry and are placed in a vacuum chamber, where vacuum is typically held at 10^{-4} to 10^{-5} torr. The anode-cathode gap is initially filled with gas from a fast valve gas-puff system or connected by an array of tiny (micrometers in diameter) wires, placed around the z axis. The typical Z-pinch plasma implosion process is illustrated in Fig. 1, where the current starts to flow through outer layers of plasma due to a skin effect, and the Z-pinch plasma implodes as a shell. After the discharge current I starts to flow axially through anode-cathode gap, it ionizes the wires or gas, produces the azimuthal magnetic field $\mathbf{B}$, and implodes the plasma by magnetic field forces. During compression of plasma into a hot dense column and its stagnation, the gas dynamic pressure equalizes the magnetic forces (Phase III in Fig. 1),³ and the kinetic energy is converted to thermal energy and to a radiation pulse. The implosion typically lasts from 50 ns to several hundred microseconds. For a plasma that consists of higher atomic number elements, the radiative collapse can happen when strong line or continuum radiation leads to plasma cooling, which decreases the gas-dynamic pressure, and allows more significant plasma compression, to a diameter of 10 to 100 μm . Under such conditions, Z-pinch loads reveal small, localized sources of x-ray radiation (“hot” or “bright” spots, the size of which may be as small as 1 to 5 μm) formed at random points along the Z-pinch axis.

In Z-pinch plasmas, EUV and x-ray radiation occurs by ionization, excitation, and recombination processes which involve multicharged ions,⁴ as shown in Fig. 2.⁵ Bremsstrahlung emission is negligible. For high-current Z-pinch systems (more than 3 to 10 MA) high-atomic-number plasmas (Pt, W, etc.), are typically optically thick, and plasma radiation can be approximated by a black-body spectrum with some presence of spectral lines of multicharged ions.¹ At university-scale Z-pinch generators, with currents of about 0.5 to 1 MA, the total EUV/x-ray radiation yield can be as high as 10 to 30% of the electromagnetic energy delivered to plasma.⁶

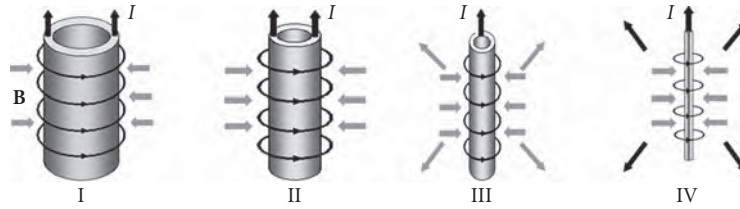


FIGURE 1 Z-pinch plasma implosion, I indicating current, and B , the magnetic field. The grey arrows pointing to the central axis show magnetic forces that compress plasma; the grey arrows outside the plasma show optical and extreme ultraviolet (EUV) emission, and black arrows showing x-ray emission. Phase I: acceleration of plasma shell from diameter 2 to 12 cm; Phase II: run-in, start of optical and EUV emission; Phase III: stagnation at diameter 1 to 4 mm; generation of softer x-rays and keV bursts; plasma reaches the thermalization stage with high concentration of multi-charged ions; Phase IV: for plasma that consists of higher atomic number elements, radiative collapse can happen when strong line radiation occurrence leads to plasma radiation cooling, a decrease in gas-dynamic pressure and additional plasma implosion to diameter 10 to 100 μm .

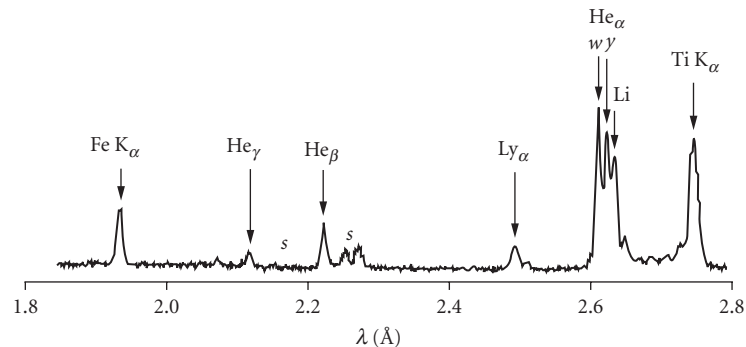


FIGURE 2 Typical spectrum of Ti X-pinch obtained at the Z-pinch Zebra generator. Peak current is 1 MA. The most intense and important spectral lines for diagnostics include the He-like lines, resonance line $\text{He}_\alpha(w)$, intercombination line y , resonance lines He_β and He_γ , and H-like resonance line Ly_α . They also include the satellite lines Li and S of Li-like ions to above mentioned resonance lines and “cold” $K\alpha$ lines of Ti and Fe. The “cold” lines were generated by electron beam existing in X-pinch plasma; Ti line was from material of wires and Fe line was from stainless steel anode.

57.2 TYPES OF Z-PINCH RADIATION SOURCES

Z-pinch plasma radiation sources have higher conversion coefficients of input electrical power to EUV/x-ray radiation compared to laser plasma or conventional x-ray sources. Z-pinch systems can use a wide range of materials for creation of plasma, including metals, dielectrics, and gases. Z-pinch radiation sources can be divided into multishot and single-shot systems. The first type includes plasma focus, gas-puff, vacuum spark, and capillary discharge devices. They can operate even in a repetition pulse regime (up to 0.1 to 0.5 Hz), but not for a lengthy period of time (from tens to several thousand shots). Their parameters are shown in Table 1.

TABLE 1 Characteristics of Pinch Sources

Source ^a	Mode ^b / Current ^c (MA)	Average Source Size ^d (mm)	X-Ray Burst Duration ^e (ns)	Total Yield ^f (kJ)
Plasma focus	multi 0.1–2	1–50	20–100	0.1–50
Gas-puff	multi 0.1–15	0.5–50	5–100	0.1–100
Vacuum spark	multi 0.05–0.4	0.01–5	5–20	0.05–0.25
Wire-array	single 0.3–20	0.5–20	5–50	0.1–1,800
X-pinch	single 0.1–1	0.002–5	0.1–10	0.01–10

^aThe minimum values given for sources parameters are for table-top devices.

^bThe mode refers to multishot or single-shot operational mode.

^cCurrent is peak current.

^dAverage source size is the size of the emitting region observed in single shot.

^eX-ray burst duration is the FWHM.

^fThe total yield is the total EUV/x-ray radiation yield.

The plasma focus is similar to a plasma gun device, where plasma sheath ejects between two cylindrical coaxial electrodes (positioned in chamber fill with H₂, D₂, He, or another gas under pressure of several torr) and implodes on central axis near electrodes' edge.⁷ Plasma focus facilities with current from 100 kA to 2 MA are mainly applied in fusion research, x-ray lithography and micromachining, surface material modification, and x-ray sources for medical purposes.

For a gas-puff source,⁸ the anode-cathode gap in the vacuum chamber fills first with a supersonic hollow gas stream from a fast valve (50 to 100-μsec-duration pulse), and then high voltage is applied to electrodes and the plasma implodes, generating an x-ray burst 10 to 100 ns in duration. The gas-puff devices cover the range from compact table-top machines with current near 100 kA to generators with 5 to 15 MA of current. The large facilities are mainly pulse sources of K-shell radiation (Ne, Ar, and Kr) with a yield more than 30 kJ, applied in fusion research and x-ray laser studies. Table-top gas-puff devices are used in fusion studies, x-ray lithography and micromachining, x-ray laser studies, and x-ray spectroscopy.

A vacuum spark⁹ is a simple compact device that consists of two electrodes positioned in a vacuum chamber and separated by a 5 to 10 mm gap. A needle-shaped anode electrode is connected to a fast capacitor with an initial stored energy of 1 to 3 kJ. The cathode electrode includes a small plasma gun. After the plasma gun ejects a cold plasma cloud into the inter-electrode gap, the current begins to flow, evaporating and ionizing anode material. For designs without a plasma gun, the cold plasma is created by a pulsed laser beam that focuses on flat cathode. The vacuum spark current peak is typically smaller than 200 to 400 kA. The plasma column implodes, with the formation of hot spots that generate x-ray bursts. Vacuum spark generators are used in x-ray spectroscopy, x-ray lithography and micromachining, and as a testing facility for new diagnostics. The vacuum spark's main shortcoming is instabilities of the x-ray yield from shot to shot.

In capillary discharge devices,¹⁰ the plasma is formed on the inner surface of a ceramic capillary (with inner diameter several millimeters, and length 100 to 150 mm) under a high-voltage pulse, and implodes in a narrow column on the capillary central axis. Such devices are table-top and generate mainly EUV bursts. They became widely applied in EUV/X-ray lasers¹⁰ in the last decade.

Single-shot x-ray sources are wire array and X-pinch. In these sources, the wires, anode, and cathode electrodes should be changed after each shot. These sources generate maximum radiation yields and powers in EUV and x-ray spectral regions. X-ray yields of 1.8 MJ and powers of 250 TW were obtained for facilities such as Sandia National Laboratories Z-generator with wire array loads.¹¹ In a Z-pinch source with a conventional wire array load, as shown in Fig. 3a,¹² the anode and cathode are initially connected by a cylindrical array (or double cylindrical array, also called nested array) of wires several micrometers in diameter that are placed around a central z axis. The wire number varies from 4 to 60 (on university-scale generators) to 200 to 300 (on multi-MA machines); array diameter is from 8 to 10 mm to 30 to 40 mm. The length of wires connecting anode and cathode is 10 to 20 mm. The mechanisms of the implosion and x-ray burst generation are generally

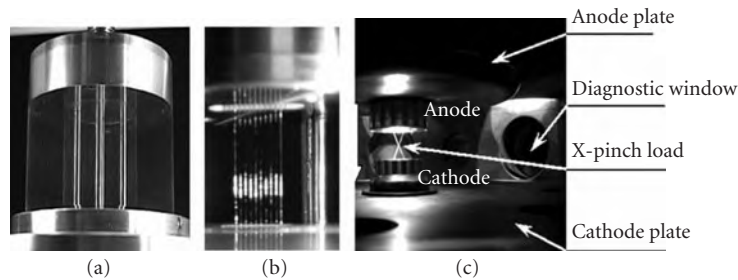


FIGURE 3 (a) A nested cylindrical wire array.¹² The central rods that are supporting the frame should be removed before the shot. (b) Single planar wire array.¹³ The gap between the wires is 1 mm. The supporting rod at the right should be removed before the shot. (c) X-pinch load in the vacuum chamber.¹⁴

the same, as described in Fig. 1. Source parameters are shown in Table 1. For the development of more effective sources than conventional cylindrical arrays, new load geometries would help to achieve plasma of higher temperature and density, and reduce the source size and possibly shape of radiation pulse. A single planar array, as shown in Fig. 3b, consists of tiny wires placed in one-plane row parallel to each other between anode and cathode plates.¹³ Multiplanar arrays consist of 2 to 3 such rows placed parallel to each other between the anode and cathode plates.⁶ Compact single planar wire array and multiplanar wire arrays (with the width of just several mm) as well as compact cylindrical wire arrays (diameter smaller than 4 to 6 mm) show better data than any other tested loads at 1-MA generators. The maximum electron temperature T_e and density N_e are 1 to 1.4 keV and 10^{21} to 10^{22} cm⁻³, respectively.^{6,13} The maximum radiation power in short nanosecond-scale rise-time x-ray bursts is around 1 TW and the yield is 23 to 25 kJ from 50 to 60 kJ of electrical energy delivered by the generator to the X-pinch load.^{6,13}

An X-pinch load is formed by crossing two or more wires at one point, as shown in Fig. 3c. X-pinch plasmas have been actively investigated for currents from 0.1 to 1 MA.^{14,15} The unique property of X-pinch as “a point source” is that the emitting region is localized near the wire crossing point and the source size varies from 1 to 2 μ m up to hundreds of micrometers. X-pinch plasmas can radiate just one or two bursts with emission times ranging from a few nanoseconds to less than 100 ps,¹⁵ as described in Table 1. At 1-MA university-scale generators, the total radiation yield can be 10 kJ, which is half of the output of planar wire array and compact cylindrical array sources, but, the burst power is comparable due to the shorter X-pinch x-ray burst duration.¹⁴

57.3 CHOICE OF OPTICS FOR Z-PINCH SOURCES

All Z-pinch plasma sources are characterized by the presence of not only powerful x-ray/EUV bursts, but also intense neutral and ion beams that can damage the filters (micrometer-thick plastic or metal films) typically used to protect x-ray/EUV optics as zone plates, transmission gratings, and multilayer and grazing incidence systems (see Chaps. 38, 40, 41, 44, 48, and 49). One of the solutions is the application of strong permanent magnets positioned between plasma and optics. The magnetic field deflects ion beams from the filter and significantly increases the filter lifetime. Table-top plasma focus, gas-puff, vacuum spark, and capillary discharge devices can be used with zone plates in applications which require monochromatic radiation, such as x-ray microscopy. Those sources also can be applied in x-ray lithography, which needs high x-ray flux, with multilayer or grazing incidence optics (high-bandpass x-ray optical systems).

For more powerful pinch sources, like 1-MA University-scale generators, the situation is more difficult, because the sensitive x-ray optical systems, even with magnetic protection, sustain damage even at a distance of 1 to 1.5 m from plasma,¹⁴ and the efficiency of the optics decreases dramatically.

Relatively inexpensive high-bandpass EUV/x-ray glass capillary optics (see Chaps. 52 and 53) can be used with such powerful x-ray Z-pinch sources.¹⁴ These can be placed at a distance of several centimeters from a z-pinch source and changed after each shot. Glass capillary optics can be adapted for any source configuration, such as a plasma column in wire array source or a point-type source in an x-pinch system. Estimates show¹⁴ that for a 1-MA wire array (linear source) or x-pinch (point source) with glass capillary optics with a 2 to 3 mm diameter focusing spot, it is possible to reach energy densities of 2 to 10 J/cm² and flux densities of $(1 \text{ to } 5) \times 10^9 \text{ W/cm}^2$. Because a typical surface melting threshold is about 1 to 2 J/cm², such Z-pinch sources with glass capillary optics can be used for surface modification research.

57.4 REFERENCES

1. C. Deeney, C. A. Coverdale, M. R. Douglas, et al., "Titanium K-shell X-Ray Radiation from High Velocity Wire Array Implosions on the 20-MA Z Accelerator," *Phys. Plasmas* **6**:2081 (1999).
2. M. H. Elghazaly, A. M. Abd Elbaky, A. H. Bassyouni, et al., "Spectroscopic Studies of Plasma Temperature in a Low-Pressure Hydrogen in Plasma," *J. Quant. Spectrosc. Rad. Transf.* **61**:503 (1999).
3. A. Velikovich, J. Davis, Y. K. Chong, et al., "Implosion Instabilities and Mitigation," *Minicourse on the Physics of Z-Pinches*, ICOPS, Monterey, CA, 2005.
4. N. R. Pereira and J. Davis, "X-Rays from Z-Pinches on Relativistic Electron-Beam Generators," *J. Appl. Physics* **64**:R1 (1988).
5. A. S. Shlyaptseva, S. B. Hansen, V. L. Kantsyrev, et al., "X-Ray Spectropolarimetry of High-Temperature Plasmas," *Rev. Scientific Instr.* **72**:1241 (2001).
6. V. L. Kantsyrev, L. I. Rudakov, A. S. Safronova, et al., "Double Planar Wire Array as a Compact Plasma Radiation Source," *Phys. Plasmas* **15**:030704 (2008).
7. N. V. Filipov, T. I. Fillipova, and V. P. Vinogradov, "Dense High-Temperature Plasma in a Noncylindrical Z-Pinch Compression," *Nucl. Fusion Suppl.* **2**:577 (1962).
8. F. C. Young, R. J. Commisso, D. P. Murphy, et al., "Measurement and Analysis of Continuum Radiation from a Large-Diameter Long Implosion Time Argon Gas Puff Z-Pinch at 6 MA," *IEEE Trans. Plasma Sci.* **34**(5):2312 (2006).
9. H. Flemberg, "X-Ray Spectra of High-Temperature Plasma," *Ark. Mat. Astron. Fys. A* **28**(18):1 (1942).
10. J. J. Rocca, D. P. Clark, J. L. A. Chilla, et al., "Energy Extraction and Achievement of the Saturation Limit in a Discharge-Pumped Table-Top Soft X-Ray Amplifier," *Phys. Rev. Lett.* **77**:1476 (1996).
11. M. K. Matzen, M. A. Sweeney, R. G. Adams, et al., "Pulsed-Power-Driven High Energy Density Physics and Inertial Confinement Fusion Research," *Phys. Plasmas* **12**:055503 (2005).
12. C. Deeney, "History of Z-Pinches," *Minicourse on the Physics of Z-Pinches*, ICOPS, Monterey, CA, 2005.
13. V. L. Kantsyrev, L. I. Rudakov, A. S. Safronova, et al., "Properties of a Planar Wire Arrays Z-Pinch Source and Comparisons with Cylindrical Arrays," *High Energ. Dens. Physics* **3**:136 (2007).
14. V. L. Kantsyrev, D. A. Fedin, A. S. Shlyaptseva, et al., "High-Z 0.9–1.0 MA X-Pinch as a Possible Backlighter in 50–100 keV and Sub-keV–10 Kev Spectral Regions and a Powerful Soft X-Ray Source for Surface Modification Research," *Rev. Scientific Instr.* **74**:1935 (2003).
15. S. A. Pikuz, D. B. Sinars, T. A. Shelkovenko, et al., "High Energy Density Z-Pinch Plasma Conditions with Picosecond Time Resolution," *Phys. Rev. Lett.* **89**:035003 (2002).

This page intentionally left blank

X-RAY LASERS

Greg Tallents

*University of York
York, United Kingdom*

58.1 FREE-ELECTRON LASERS

The original concept for a free-electron laser was proposed by Madey in 1971¹ with free-electron laser output observed by Elias in 1976². In a free-electron laser, stimulated emission occurs from relativistic electrons passing through a spatially periodic transverse magnetic field (known as an undulator). In free-electron lasers operating at wavelength λ , the electrons interact with the laser electric field and arrange themselves into “microbunches” separated by one wavelength λ so that the electric field arising from the stimulated emission by each electron is coherently additive at each spatial point. As the intensity of laser output is proportional to the square of the laser electric field, if N electrons are present, the free-electron laser intensity is proportional to N^2 . The simplest form of free-electron lasing is where spontaneously emitted photons are amplified, so-called self-amplified spontaneous emission (SASE) output. The frequency bandwidth of output in SASE mode is usually large ($\nu/\Delta\nu \approx 10$), potentially enabling ultrashort laser pulses ($<10^{-15}$ s) to be produced. However, large frequency bandwidth implies a short coherence length that can be limiting for some applications. To reduce the free-electron laser bandwidth, increase the coherence length and increase the output brilliance (photons $\text{s}^{-1}\text{mm}^{-2}\text{mrad}^{-2}$ 0.1%BW); it will be possible to “seed” the amplification process with photons produced by infrared laser harmonic radiation. The production of harmonic radiation is discussed in this chapter.

In 2008, free-electron laser output of ≈ 70 μJ per pulse had been produced down to a fundamental wavelength $\lambda \approx 6$ nm together with significantly shorter wavelengths, but less intense harmonic components at the FLASH free-electron laser facility at Deutsches Elektronen-Synchrotron (DESY) in Hamburg.³ Free-electron laser operation will be extended to an unprecedented short wavelength of ≈ 0.1 nm with the European X-Ray Free-Electron Laser (XFEL) to be also constructed at DESY with the Linac Coherent Light Source (LCLS) at the Stanford Linear Accelerator Center (SLAC) due to come on-line in 2009 and with the Spring-8 Compact SASE Source (SCSS) under development in Japan.

Free-electron lasers (FELs) are intense (peak brilliance up to 10^{34} photons $\text{s}^{-1}\text{mm}^{-2}$ 0.1% bandwidth is expected for XFEL),³ have high transverse coherence, and may be able to be ultimately focused to focal diameters $\sim \lambda$. The high-focused irradiance (initially up to 10^{18} Wcm^{-2} , ultimately possibly up to 10^{30} Wcm^{-2} assuming focusing to λ dimensions) and high photon energy (ultimately exceeding 10 keV) will enable many novel experiments in high-energy density physics, biological imaging, and other fields.

Future possibilities for basic physics research include the production of Unruh radiation⁴ and electron-positron pair production. However, FELs are big, multiuser facilities of large capital and running cost. A review in 2008 led to the decision in the United Kingdom not to proceed with the 4GLS free-electron laser source producing extreme ultraviolet (EUV) radiation at $\lambda > 10$ nm.⁵ The review suggested that FELs are uniquely useful for harder x-ray experiments ($\lambda < 5$ nm), but that they have strong competitors from other laser sources in the EUV.

58.2 HIGH HARMONIC PRODUCTION

Experiments with 10 to 100-fs-duration visible and near-infrared laser pulses incident into inert gas targets such as Ne, Ar, or Xe targets have shown that harmonics are produced at wavelengths >4 nm with an efficiency $\sim 10^{-6}$ into each harmonic.^{6,7} Electrons are tunnel-ionised from gas atoms by the laser and then accelerated in the laser electric field. Provided linear polarisation is employed, the electron is accelerated back to the atom when the laser electric field reverses direction. Returning back to the atom, the electron recombines, emitting radiation up to the energy of the accelerated electron ($\approx$ atom ionisation energy + $3.17 \times$ electron ponderomotive energy).⁸ Due to this harmonic production being closely associated with each oscillation of the driving laser electric field, the harmonics have to a good approximation the temporal duration and beam qualities of the optical laser creating them. The harmonics are each produced with approximately equal efficiency out to a maximum harmonic number (in the so-called “plateau” regime). The maximum harmonic photon energy $h\nu_{\max}$ in the plateau is

$$h\nu_{\max} \cong I_p + \left(3.17 \times 10^{-13} \frac{eV}{(W/cm^2)\mu m^2} \right) I_{\text{laser}} \lambda_{\mu m}^2 \quad (1)$$

where I_p is the atom ionization energy (in eV), I_{laser} is the laser irradiance (in Wcm^{-2}) and $\lambda_{\mu m}$ is the driving laser wavelength (in microns). For example, a focused laser irradiance of $3 \times 10^{14} Wcm^{-2}$ can be expected to produce strong harmonics up to $h\nu_{\max} \approx 100$ eV, equivalent to a wavelength of 12 nm.

Harmonics from the laser irradiation of solid targets have been recently demonstrated down to wavelengths ≈ 0.33 nm with 10^{-6} efficiency from 500-fs laser pulses.⁹ Such a development is exciting as the harmonic wavelength is in the x-ray regime. To achieve x-ray harmonics, a large laser system capable of focusing to $\sim 10^{21} Wcm^{-2}$ with high contrast ($>10^{10}$) is required. Much less stringent requirements are needed for the production of EUV harmonics. An intensity of $10^{19} Wcm^{-2}$ and contrast $> 10^{-6}$ is sufficient.¹⁰ Harmonics from solid surfaces are emitted in the specular reflection direction for the incident laser in a narrow cone much smaller than the incident laser cone size. As they arise from the relativistic oscillation by the incoming laser electric field of the vacuum-solid interface, the beam quality of the harmonic radiation depends strongly on the uniformity of the solid target irradiation. However, comprehensive measurements of solid target high harmonic beam quality have not yet been made.

58.3 PLASMA-BASED EUV LASERS

Until the late 1990s, plasma-based EUV lasers had only been pumped with high-energy laser pulses. For example, the record shortest wavelength EUV laser with saturated output was produced at 5.9 nm, but required ~ 60 to 120 J of energy per pulse.¹¹ It was not feasible to propose that such lasers could be used to produce EUV lasing for applications where high average power was required as it is not possible to have a high-repetition-rate laser of such energy per pulse. Developments in the last 10 years have completely changed this picture and have enabled demonstrations showing repetitively pulsed EUV laser action (10 to 50 nm) at high average power. Discharge pumped EUV laser action

at 46.9 nm was demonstrated with a capillary discharge,¹² but significantly shorter wavelength (<20 nm), high repetition rate lasing has been achieved with infrared laser pumping in plasmas.¹³

Repetitively pulsed (>10 Hz), short duration (ps) infrared lasers capable of pumping EUV lasing from initially solid targets were developed in the late 1990s. Resulting EUV gain durations are short with <ps pumping pulses, so “traveling wave” pumping is needed whereby the pumping laser is incident along the ~3 to 10 mm target length coincident with the speed of propagation ($\approx c$) of the amplifying EUV beam. The efficiency of generating EUV lasing from solid targets works best with electron collisional excitation via laser pumping using two-pulse irradiation. A prepulse laser generates an expanding plasma. A main pulse heats, ionizes, and excites this plasma to generate gain via electron collisional excitation at some later time. The prepulse irradiation ensures that the volume of the gain region is large, that the main laser pulse is strongly absorbed, and that density gradients at the time of arrival of the main pulse are small so that refraction of the EUV lasing beam is reduced. A recent significant development, grazing incidence pumping,¹⁴ improves the pumping efficiency even more by offering a controlled way of achieving pumping at the optimum electron density in the plasma, as shown in Fig. 1. There is an optimum density for gain above which collisional de-excitation destroys the population inversion and below which the gain coefficient drops due to the decrease in density. In grazing incidence pumping, laser light penetrates to a turning point determined by the angle of incidence tuned to correspond to the optimum plasma density. In grazing-incidence pumping, there is also an inherent (close to velocity c) traveling wave excitation of gain. EUV lasing has been produced with laser pumping energies <1 J using the grazing incidence pumping technique.¹⁵

The development of plasma-based extreme ultraviolet (EUV) lasers has been a very successful investigation that initially started in the 1970s, but has matured to a state where saturated output has been achieved down to the record short wavelength of 5.9 nm and the atomic, plasma, and propagation physics are known to good accuracy (see Ref. 16). Plasma-based EUV lasers (10 to 20 nm) offer a route to the achievement of both high peak brightness (typically 10^{24} photons $\text{s}^{-1} \text{mm}^{-2} \text{mrad}^{-2}$ has been achieved) and high average power (e.g., 10^{14} photons $\text{s}^{-1} \text{mm}^{-2} \text{mrad}^{-2}$ under operation at 10 Hz). This compares to, for example, third generation synchrotron light sources that produce smaller to comparable peak brightness and comparable average brightness for a spectral bandwidth $\Delta\nu/\nu$ of 0.1% over similar spectral ranges. Plasma-based EUV laser output is extremely narrowband ($\Delta\nu/\nu < 10^{-4}$), so coherence lengths are typically 100 μm to 1 mm. The pulse energy in a plasma-based EUV laser is

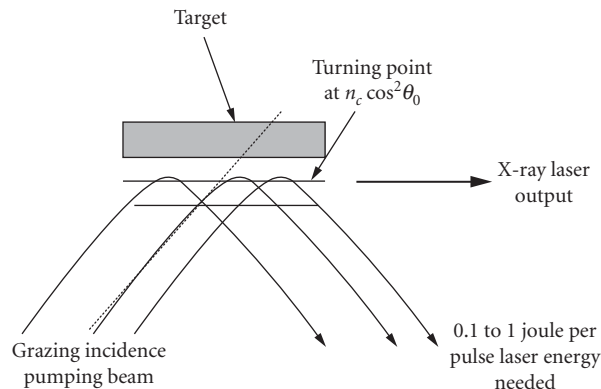


FIGURE 1 Schematic diagram illustrating grazing-incidence pumping into a preformed plasma. Heating occurs preferentially at the turning point of the beam (at electron density $n_c \cos^2 \theta_0$, where n_c is the electron critical density for the pump laser wavelength). The angle of incidence θ_0 is tuned so that the pump laser beam turning point corresponds to the optimum density for EUV laser gain (0.1 to 1 joule per pulse laser energy needed).

typically 1 μJ to 1 mJ, giving 0.01- to 10-mW average power (at rep rates of 10 Hz) and 1- to 1000-MW peak power.

An outstanding issue with plasma-based EUV lasers is the beam spatial coherence. The phase profile of these lasers needs to be improved considerably while maintaining their power output to enable their use for the high quality optics testing needed, for example, by the EUV lithography industry. At present plasma-based EUV lasers have high Fresnel number ($\sim 10^4$) with EUV lasers effectively composed of a large number ($\approx$ Fresnel number) of “beamlets” that interfere to create output that is a complex speckle pattern.¹⁷ With short pulse ($\sim \text{ps}$) pumping, each beamlet results from a single spontaneous emission and the phase of each beamlet is approximately uniform transversely, but random with respect to other beamlets. The temporal duration of each beamlet is typically 3 ps and is close to the Fourier transform limit.¹⁸

58.4 REFERENCES

1. J. M. Madey, “Stimulated Emission of Bremsstrahlung in a Periodic Magnetic Field,” *J. Appl. Phys.* **42**:1906 (1971).
2. L. R. Elias, “Observation of Stimulated Emission Radiation by Relativistic Electrons in Spatially Periodic Transverse Magnetic Field,” *Phys. Rev. Lett.* **36**:717 (1976).
3. V. Ayvazyan, R. Baboi, J. Bähr, V. Balandin, B. Beutner, A. Brandt, I. Bohnet, et al., “First Operation of a Free-Electron Laser Generating GW Power Radiation at 32 nm Wavelength,” *Eur. J. Phys.* **D37**:207 (2006); see also <http://xfelinfo.desy.de/en/start/2/index.html>, accessed 6 May 2009.
4. P. Chen and T. Tajima, “Testing Unruh Radiation with Ultraintense Lasers,” *Phys. Rev. Lett.* **83**:256 (1999).
5. B. W. J. McNeil, N. R. Thompson, D. J. Dunning, J. G. Karssenberg, P. J. M. van der Slot, and K.-J. Boller, “A Design for the Generation of Temporally Coherent Radiation Pulses in the VUV and beyond by a Self-Seeding High Gain Free Electron Laser Amplifier,” *New J. Phys.* **9**:82 (2007).
6. C. Altucci, R. Bruzzese, C. de Lisio, M. Nisoli, G. Cerullo, S. Stagira, S. De Silvestri, et al., “Features of High-Order Harmonic Generation in the 30 fs and the Sub-10 fs Regimes,” *J. Opt.* **A2**:289 (2000).
7. T. Ditmire, J. K. Crane, H. Nguyen, L. B. DaSilva, and M. D. Perry, “Energy-yield and Conversion Efficiency Measurements of High Order Harmonic Radiation,” *Phys. Rev.* **A51**:R902 (1995).
8. P. Corkum, “Plasma Perspectives on Strong Field Multiphoton Ionization,” *Phys. Rev. Lett.* **71**:1994 (1993).
9. B. Dromey, S. Kar, C. Bellei, D. C. Carroll, R. J. Clarke, J. S. Green, S. Kneip, et al., “Bright Multi-keV Harmonic Generation from Relativistically Oscillating Plasma Surfaces,” *Phys. Rev. Lett.* **99**:085001 (2007).
10. P. A. Norreys, M. Zepf, S. Moustazis, A. P. Fewes, J. Zhang, P. Lee, M. Bakarezos, et al., “Efficient Extreme UV Harmonics Generated from Picosecond Laser Pulse Interactions with Solid Targets,” *Phys. Rev. Lett.* **76**:1832 (1996).
11. R. Smith, G. J. Tallents, J. Zhang, G. Eder, S. McCabe, G. J. Pert, and E. Wolfrum, “Saturation Behavior of Two X-ray Lasing Transitions in Ni-Like Dy,” *Phys. Rev.* **A59**:R47 (1999).
12. B. R. Benware, C. D. Macchietto, C. H. Moreno, and J. J. Rocca, “Demonstration of a High Average Power Tabletop Soft X-Ray Laser,” *Phys. Rev. Lett.* **81**:5805 (1998).
13. Y. Wang, M. A. Larotonda, B. M. Luther, D. Alessi, M. Berrill, V. N. Shlyaptsev, and J. J. Rocca, “Demonstration of High Repetition Rate Tabletop Soft X-Ray Lasers with Saturated Output at Wavelengths Down to 13.9 nm and Gain Down to 10.9 nm,” *Phys. Rev.* **A72**:053807 (2005).
14. R. Keenan, J. Dunn, P. K. Patel, D. F. Price, R. F. Smith, and V. N. Shlyaptsev, “High Repetition Rate Grazing Incidence Pumped X-Ray Laser Operating at 18.9 nm,” *Phys. Rev. Lett.* **94**:103901 (2005).
15. J. Tummeler, K. A. Janulewicz, G. Priebe, and P. V. Nickles, “10-Hz Grazing-Incidence Pumped Ni-Like Mo X-Ray Laser,” *Phys. Rev.* **E72**:037401 (2005).
16. G. J. Tallents, “The Physics of Soft X-Ray Lasers Pumped by Electron Collisions in Laser Plasmas,” *J. Phys.* **D36**:R259 (2003).
17. O. Guilbaud, A. Klisnick, K. Cassou, S. Kazamias, D. Ros, G. Jamelot, D. Joyeux, and D. Phalippou, “Origin of Microstructures in Picosecond X-Ray Laser Beams,” *Europhys. Lett.* **74**:823 (2006).
18. P. Mistry, M. H. Edwards, and G. J. Tallents, “X-Ray Laser Pulses at the Fourier Transform Limit,” *Phys. Rev.* **A75**:013818 (2006).

INVERSE COMPTON X-RAY SOURCES

Frank Carroll

MXISystems

Nashville, Tennessee

59.1 INTRODUCTION

Beams of pulsed, tunable, monochromatic hard x rays have been a long sought after goal. Such “dream beams,” that are tunable from energies as low as 8 keV to well into the *hundreds of keV*, have now become available in compact units that can rival synchrotrons in many ways, delivering fluxes of photons of varying bandwidths that approach or exceed the output of those much larger facilities at some energies. Additionally, these monochromatic beams of x rays can be produced in cone beam geometries that are very useful for covering larger areas, even as large as humans.

All of this has come about because we have learned to harness a process called inverse Compton scattering (Thomson backscattering). In normal Compton scattering, a photon incoherently scattering from a nearly stationary electron exchanges some of its energy and momentum with the electron, resulting in a lower energy photon. In inverse Compton scattering, the electron is traveling at high velocity, and the photon gains energy. In one embodiment of this process, an electron beam is accelerated to near-relativistic energies typically between 12 and 50 MeV and then focused down to a focal spot that is anywhere from 3 to 100 micron in size in an area designated the interaction zone (IZ). In like fashion, an intense (terawatt) laser beam is also focused down to a similar sized focal spot and counterpropagated against the packet of electrons in a head-to-head collision (180° geometry), as shown in Fig. 1.

The Rayleigh ranges of the beams are aligned so that they completely overlap in the IZ and such that the packets of light and electrons both reach the IZ at the same instant. In that collision, the laser photons are Doppler shifted by the inverse Compton process to x-ray energies. Hence a light photon goes in and an x-ray photon comes out. X rays are generated in a somewhat slowly diverging cone beam along the axis and in the direction traveled by the electron beam, as shown in Fig. 1. These x rays typically exit the machine through a beryllium vacuum window for use in various applications. Since the accelerator can be tuned, the x rays emanating from the machine are tunable. Since all of the electrons are not at exactly the same energy, some reduction in monochromaticity of the x rays is seen, but since the laser light is nearly monochromatic, the x rays, on the whole, are nearly monochromatic.^{1–5}

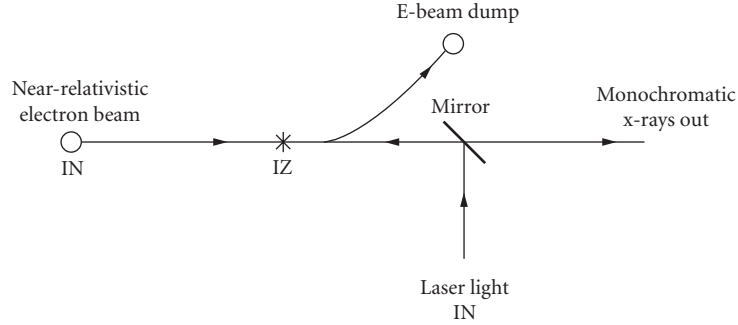


FIGURE 1 Inverse Compton process in practice.

59.2 INVERSE COMPTON CALCULATIONS

The theoretical x-ray yields from such sources can be computed directly from the Thomson scattering cross section and some basic beam-geometry and flux considerations. The Thomson cross-section σ_T is

$$\sigma_T = \frac{8\pi}{3} r_e^2 = 6.652 \times 10^{-29} \text{ m}^2 \quad (1)$$

where r_e is the classical electron radius. If one uses a diffraction-limited optical beam and an electron beam that is assumed to be smaller than the optical spot size, then this gives an x-ray yield of

$$N_x(\Delta V_x) = \frac{Qe U_L \sigma_T}{hc q Z_r \Delta V_x} \quad (2)$$

where Qe is the charge in an electron bunch, U_L is the energy (in joules) in a laser pulse, h is Planck's constant, c is the speed of light, q is the electron charge, Z_r is the Rayleigh range of the laser focal spot (which is matched to the electron beam focal parameters for a given desired ΔV_x), and ΔV_x is the fractional energy spread of the x-ray output which is useful to the application. For a system with 10 J of laser light, 1 nC of electron charge, and an f/10 final focus of the laser, the output is about 5×10^9 x rays into a 10% bandwidth. The brilliance, is $7 \times 10^{31} [\text{m}^2 \cdot \text{sr} \cdot \text{s} \cdot (10\% \text{ bandwidth})]^{-1}$ assuming a 10-ps output pulse.

Modifications to such machines can include using light of a shorter wavelength, thereby keeping gradients within the accelerator smaller. In addition, lasers used for these machines can be run at 20 Hz instead of 10 Hz, yielding almost double the light and hence greater x-ray flux.

An additional and extremely important enhancement comes about by using smaller electron beam focal spots at the interaction zone, which, when using a higher current of electrons, increases the effective cross section for interaction with the light photons, creating significantly more x rays as well.

When all of these factors are used together, one gains anywhere from one to three orders of magnitude in the output flux of the x rays.⁶⁻⁸

59.3 PRACTICAL DEVICES

Synchrotrons are, of course, broadband white radiation sources (see Chap. 55), which use monochromators to deliver beams of narrow bandwidth (see Chap. 39) to multiple user beamlines. Inverse Compton sources typically deliver bandwidths of 0.1 to 10%, and can be further monochromatized to narrower bandwidth using various x-ray optics as well;⁹⁻¹² but this diminishes their brightness relative to the typical synchrotron source. However, Compton sources are much more compact,

TABLE 1 Comparison of Synchrotrons to Available Compton Sources

	(Generation 2) Machine*	(Generation 3) Compton Source†	Typical Synchrotrons‡
E-beam:	Linac running in “single pulse” mode Up to 75 MeV 0.5 nanocoulombs/pulse Emittance $\cong 3.00\pi$ mm-mrad Copper photocathode	Same Same 2–3 nanocoulombs/pulse Emittance $\cong 0.7\pi$ mm-mrad Copper-Mg photocathode	Synchrotron ring 0.1–7.0 GeV 50–5200 mA current Variable N/A
Laser:	Tabletop terawatt 1054-nm IR light 20 J/5 ps pulse Once every 5 min	Tabletop terawatt 532-nm green light 1.5 J/5 ps pulse 10–20 Hz rep rate	N/A N/A N/A N/A
X-ray beam:	$10^{10}/5$ ps shot 3×10^7 photons/s Tunable from 12–50 keV Change energy 30 min to 4 h 0.1–10% bandwidth Conebeam geometry (large areas covered in a single shot) Effective focal spot 100–300 μ	10^{9-10} photons/5 ps shot $2.23 \times 10^{11-12}$ photons/s Tunable from 10–100 keV Change energy in 1/10th of a second Same Same but focusable to 100 μ (round beam) or 300 μ in a square beam, if desired Effective focal spot 3–6 μ	Quasi-CW Typically $1.4 \times 10^{13-16}$ photons/s Variable from 1–57 keV Variable 1×10^{-4} 100 $\times$ 200 μ Effectively collimated
Shielding vault:	None	None	Complex

*Inverse Compton source operational at the Vanderbilt University Medical Free-Electron Laser Facility, Nashville, TN since April 2001.^{4,5}

†MXISystems, Inc., Inverse Compton source, 2008.¹⁹

‡Advanced Design Consulting USA, Inc., Synchrotron Primer—World’s Light Sources. Section 4-2.¹³

affordable, and available. A comparison of typical synchrotron beamlines to two published compact Compton sources is shown in Table 1.

While inverse Compton x-ray sources have been proposed by many over the years, there have only been a few devices successfully built and operated. (Ting 500 eV,¹⁴ Carroll 14 to 18 keV,^{1,2} Ruth 10 to 35 keV,^{10,15} Carroll 12 to 53 keV,^{3,4} Gibson/Pelaides 70 keV,¹⁶ and more recently Barty/T-Rex¹⁷ at 776 keV.) Some of the more compact devices are now being commercialized for both medical and nonmedical purposes in the ranges from 8 to 120 keV (MXISystems, Inc.)^{18,19} and 10 to 35 keV (Lyncean Technologies, Inc.)^{20,21}

59.4 APPLICATIONS

With the advent of this enabling technology come broad-ranging applications. In medicine these uses encompass both the diagnostic and therapeutic arenas.

Due to inherent contrast differences between tissues in the human body, monochromatic x-ray beams can make cancerous tissues stand out from normal tissues without the use of contrast materials, while at the same time lowering radiation dose significantly to the patient.^{22,23} This is particularly true in studies such as mammography, where 3-D compressionless studies become feasible, allowing the examination to be performed more accurately and with far less discomfort to the patient while requiring anywhere from 5 to 60 times less radiation dose depending on how the study is performed (see Chap. 31).

The extremely small effective focal spot size at the interaction zone results in sufficient lateral coherence to allow one to take advantage of the disparate refractive indices of various tissues and use the diffractive effects produced at edges within tissues to perform phase contrast imaging (see Chap. 27). This type of imaging may be capable of delivering 100 to 1000 times the information than that achieved using absorption imaging (what we have done for the past 100 years).^{24–30}

By detecting and back projecting low angle scatter emanating from an irradiated tissue or organ, one can conceivably perform noninvasive biopsies of tissues deep within the body of the patient, discerning cancerous from noncancerous tissues in a painless fashion.³¹

Therapeutically, monochromatic x rays can be tuned to the binding energies of the innermost electron shells of any atom in the periodic chart, thereby knocking K- or L-shell electrons from their orbits, causing cascades of energy (the Auger cascade) concentrating all of this energy within nanometers of the atom. By attaching a drug containing a heavy metal target (such as Pt, Gd, or I) to the DNA of cancerous cells, one can stimulate Auger cascades within the DNA causing double-stranded breaks that for the most part will not heal. This translates into treatment of cancers with 3 to 5 times less radiation than is now used by the best techniques.^{32,33}

59.5 INDUSTRIAL/MILITARY/CRYSTALLOGRAPHIC USES

These narrow bandwidth, pulsed x rays (with pulse lengths of 5 to 10 ps) are produced through widely tunable energy ranges. These are extremely useful in evaluation of defects in newer carbon composite airplane parts,^{34,35} spacecraft parts and subsystems, examination of corrosion and cracks in our aging aluminum aircraft fleet, internal analysis of turbines at full power, performance of protein crystallography,³⁶ and a host of other nondestructive testing applications.

59.6 REFERENCES

1. F. E. Carroll, J. W. Waters, R. R. Price, C. A. Brau, C. F. Roos, N. H. Tolk, D. R. Pickens, and W. H. Stephens, "Near-Monochromatic X-Ray Beams Produced by the Free Electron Laser and Compton Backscatter," *Invest. Radiol.* **25**:465–471, 1990.
2. C. Pellegrini, "On Some Methods of X-Ray Production from Relativistic Electron Beams," *J. X-Ray Sci. Technol.* **4**:275–289, 1994.
3. F. E. Carroll, J. W. Waters, R. H. Traeger, M. H. Mendenhall, W. W. Clark, and D. A. Brau, "Production of Tunable, Monochromatic X-Rays by the Vanderbilt Free-Electron Laser," *Proceedings SPIE. Free Electron Laser Challenges II* **3614**:139–146, 1999.
4. F. E. Carroll, "Perspectives: Tunable, Monochromatic X-Rays: A New Paradigm in Medicine," *AJR* **179**: 583–590, 2002.
5. F. E. Carroll, M. H. Mendenhall, R. H. Traeger, C. Brau, and J. W. Waters, "Pulsed, Tunable, Monochromatic X-Rays from a Compact Source: New Opportunities," *AJR* **181**:1197–1202, 2003.
6. W. J. Brown, S. G. Anderson, C. P. J. Barty, S. M. Betts, R. Booth, J. K. Crane, R. R. Cross, et al., "Experimental Characterization of an Ultrafast Thomson Scattering X-Ray Source with Three-Dimensional Time and Frequency-Domain Analysis," *Phys. Rev. ST Accel. Beams* **7**:060702–12, 2004.
7. F. V. Hartemann, W. J. Brown, D. J. Gibson, S. G. Anderson, A. M. Tremaine, P. T. Springer, A. J. Wooton, E. P. Hartouni, and C. P. J. Barty, "High Energy Scaling of Compton Light Sources," *Phys. Rev. ST Accel. Beams* **8**:100702–17, 2005.
8. C. A. Brau, "Oscillations in the Spectrum of Nonlinear Thomson-Backscattered Radiation," *Phys. Rev. ST Accel. Beams* **7**:020701–9, 2004.
9. M. A. Kumakhov and F. F. Komarov, "Multiple Reflection from Surface X-Ray Optics," *Phys. Rep.* **191**:289, 1990.
10. P. A. Tompkins, C. C. Abreau, F. E. Carroll, Q. F. Xiao, and C. A. MacDonald, "Use of Capillary Optics as a Beam Intensifier for a Compton X-Ray Source," *Med. Phys.* **21**:1777–1784, 1994.

11. www.rigako.com/optics/index.htm, accessed July 16, 2009.
12. F. R. Sugiro, D. Li, and C. A. MacDonald, "Beam Collimation with Polycapillary X-Ray Optics for High Contrast High Resolution Monochromatic Imaging," *Med. Phys.* **31**(12):3288–3297, 2004.
13. www.adc9001.com/index.php?src=primers_synchrotron, accessed Sept. 16, 2008.
14. A. Ting, R. Fischer, A. Fisher, C. I. Moore, B. Hafizi, R. Elton, K. Krushelnick, et al., "Demonstration Experiment of a Laser Synchrotron Source for Tunable, Monochromatic X-Rays at 500 eV," *Nucl. Instrum. Meth. Phys. Res. A* **375**:ABS68–70, 1996.
15. Z. Huang and R. D. Ruth, "Laser-Electron Storage Ring," *Phys. Rev. Lett.* **80**(5):976–9, 1998.
16. D. J. Gibson, S. G. Anderson, C. P. J. Barty, S. M. Betts, R. Booth, W. J. Brown, J. K. Crane, et al., "PLEIADES: A Picosecond Compton Scattering X-Ray Source for Advanced Backlighting and Time-Resolved Material Studies," *Phys. Plasmas* **11**:2857 (2004).
17. Internal LLNL communication. https://publicaffairs.llnl.gov/news/news_releases/2008/NR-08-04-03.html, accessed July 16, 2009.
18. US Patents 6,332,017 B1, 2001 and 6,687,333 B2, 2004. System and Method for Producing Pulsed Monochromatic X-Rays.
19. MXISystems, Inc. www.mxisystems.com, accessed July 16, 2009.
20. International Patent WO 101926 A2, 2005 and US Patents 5,825,847, 1998 and 6,035,015, 2000.
21. Lyncean. Inc. www.lynceantech.com, accessed July 16, 2009.
22. P. C. Johns and M. J. Yaffe, "X-Ray Characterization of Normal and Neoplastic Breast Tissues," *Phys. Med. Biol.* **32**:675–695, 1987.
23. F. E. Carroll, J. W. Waters, W. W. Andrews, R. R. Price, D. R. Pickens, R. Willcott, P. Tompkins, et al., "Attenuation of Monochromatic X-Rays by Normal and Abnormal Breast Tissues," *Invest. Radiol.* **29**:266–272, 1994.
24. T. Takeda, A. Momose, Y. Itai, J. Wu, and K. Hirano, "Phase-Contrast Imaging with Synchrotron X-Rays for Detecting Cancer Lesions: Preliminary Investigation," *Acad. Radiol.* **2**:799–803, 1995.
25. D. Chapman, W. Thomlinson, R. E. Johnston, et al., "Diffraction Enhanced X-Ray Imaging," *Phys. Med. Biol.* **42**:2015–2025, 1997.
26. U. Kleuker, P. Suortti, W. Weyrich, and P. Spanne, "Feasibility Study of X-Ray Diffraction Computed Tomography for Medical Imaging," *Phys. Med. Biol.* **43**:2911–2923, 1998.
27. E. D. Pisano, E. R. Johnston, D. Chapman, G. Iacocca, C. A. Livasy, D. B. Washburn, D. E. Sayers, et al., "Human Breast Cancer Specimens: Diffraction-Enhanced Imaging with Histologic Correlation-Improved Conspicuity of Lesion Detail Compared with Digital Radiography," *Radiology* **214**:895–901, 2000.
28. E. F. Donnelly and R. R. Price, "Quantification of the Effect of KVP on Edge-Enhancement Index in Phase-Contrast Radiography," *Med. Phys.* **29**(6):999–1002, 2002.
29. E. F. Donnelly, R. R. Price, R. David, and D. R. Pickens, "Dual Focal-Spot Imaging for Phase Extraction in Phase-Contrast Radiography," *Med. Phys.* **30**(9), September 2003.
30. A. Momose, "Phase-Sensitive Imaging and Phase Tomography Using X-Ray Interferometers," *Opt. Exp.* **11**(19):2303–2314, 2003.
31. P. C. Johns, R. J. Leclair, and M. P. Wismayer, "Medical X-Ray Imaging with Scattered Photons," *SPIE Reg. Meet. Optoelectron. Photon. Imaging SPIE* **TD01**:355–357, 2002.
32. J. P. Pignol, E. Rakovitch, D. Beachey, and C. L. Sech, "Clinical Significance of Atomic Inner Shell Ionization (ISI) and Auger Cascade for Radiosensitization Using IUDR, BUDR, Platinum Salts, or Gadolinium Porphyrin Compounds," *Int. J. Radiat. Oncol. Biol. Phys.* **55**(4):1082–1091, 2003.
33. F. E. Carroll, "Tunable Monochromatic X-Rays, an Enabling Technology for Molecular/Cellular Imaging and Therapy," *J. Cell. Biochem.* **90**:502–508, 2003.
34. J. Martin-Herrero and Ch. Germain, "Microstructure Reconstruction of Fibrous C/C Composites from X-Ray Microtomography," *Carbon* **45**:1242–1253, 2007.
35. P. M. Pawar and R. Ganguli, "On the Effect of Progressive Damage on Composite Helicopter Rotor System Behavior," *Comp. Struc.* **78**:410–423, 2007.
36. F. V. Hartemann, H. A. Baldis, A. K. Kerman, A. Le Fol, N. C. Luhmann, and B. Rupp, "Three-Dimensional Theory of Emittance in Compton Scattering and X-Ray Protein Crystallography," *Phys. Rev. E* **64**:016501, 2001.

This page intentionally left blank

SUBPART

5.5

X-RAY DETECTORS

This page intentionally left blank

INTRODUCTION TO X-RAY DETECTORS

Walter Gibson*

*X-Ray Optical Systems, Inc.
Rensselaer, New York*

Peter Siddons

*Brookhaven National Laboratory
Upton, New York*

60.1 INTRODUCTION

X rays can interact with gases, liquids, or solids to produce electronic, atomic, molecular, or collective excitations. Measurement of these excitations can be used to determine the x-ray intensity, energy, and position. A large variety of phenomena have been used for such detection. These include ionization of atoms and molecules, production of electrons and holes in semiconductors, excitation of both short-lived and metastable electronic states in liquids and solids, and excitation of phonons and Cooper pairs in superconductors. Selection of the detector or detectors to be used for a given application depends on the particular requirements and constraints inherent in that application, and will typically involve such factors as intensity, energy resolution, position resolution, size, convenience, and cost. To help with consideration of the kinds of trade-offs involved in selection of a particular detector, and, in the context of this *Handbook*, the factors that influence the optimization of a particular optic-detector system, this chapter summarizes the detector choices in terms of their type and function. There are no comprehensive reviews specifically devoted to x-ray detectors. Most books cover detectors for several types of ionizing radiation.¹⁻⁴ A useful summary specifically devoted to x-ray detectors is given by Buckley,⁵ and a summary of x-ray detectors for astronomy is given by Fraser.⁶

60.2 DETECTOR TYPE

Ionization

Ionization Chamber Ionization chambers contain a gas in a region of high electric field, usually produced with flat plates. X rays incident on a gas can excite bound electrons, resulting in their separation from the atom with energy equal to the difference between the absorbed x-ray energy and the binding energy of the electron. The emitted electron can, if it has sufficient energy, result in further ionization.

*This volume is dedicated in memory of Walter Gibson.

The average x-ray (actually secondary electron, because these produce most of the ionization) energy loss to produce an electron-ion pair depends on the gas, but it is typically about 30 eV (electronvolts). This is considerably in excess of the energy needed to ionize the atom, with the difference going into dissociation and excitation of molecules. The electrons and positive ions can then be separated by an applied electric field. The resulting charge current in the external circuit is a measure of the number of electron-ion pairs produced. Because the electrons move much faster than the heavy positive ions, the short-term charge pulse is primarily a measure of the electron motion. Ionization chambers have frequently been used as spectrometers for energetic ions, such as alpha particles or fission fragments, because they have shorter range than x rays, permitting an electrode configuration that allows only electron motion to be measured.

The longer range of x rays precludes easy separation of electron and ion motion in a simple two-plate chamber. The addition of screening grids can allow such a separation, but the typical applications of these devices do not usually justify the additional complexity. Also, the statistical fluctuations are larger than in some other spectrometers, so ion chambers are rarely used to measure x-ray energy distributions. However, ion chambers are the basis of frequently used radiation-monitoring devices, such as the pencil-type electrometers that are used as personal monitors. Another widespread application is to measure the x-ray beam intensity in synchrotron-based experiments. In this application, the ion chamber can be relatively thin (absorbing only a very small but proportional fraction of the incident x rays), and the resulting current flow in the external circuit is used to measure the x-ray intensity. For example, an ion chamber before a sample and one after can be used to measure the x-ray absorption in the sample as in x-ray absorption fine structure (EXAFS)^{7,8} and x-ray absorption near-edge structure (XANES)⁹ experiments (see Chap. 30). The gas in an ion chamber can just be air at atmospheric pressure as in simple monitoring devices. Frequently, when a more consistent measurement is desirable, a rare gas such as helium or argon is used to suppress electron-ion recombination, often with added methane or other gas to increase the electron drift velocity. In the EXAFS applications mentioned here, it is common practice to mix the measurement rare gas with helium in varying fractions in order to tailor the absorbed radiation fraction for optimal measurement precision. Clearly, one does not want to have all of the radiation absorbed by the first of the two chambers. A general discussion of ionization chambers can be found in reviews of radiation detectors.¹⁻⁴

Proportional Counter If the electric field gradient in an ionization chamber is increased high enough, even low-energy electrons released during the x-ray energy loss process can be accelerated enough to produce further ionization. Although, for charged particles or electrons, a parallel plate geometry is frequently used, for x rays, thin wires are often used as the positive electrode so that the acceleration takes place in a relatively confined space, close to the anode. Over a relatively wide applied voltage range, the charge multiplication increases with applied field and is constant at a given applied field. This multiplication results in charge amplification that is fast, low noise, and because it takes place close to the anode, provides effective separation from the ion current. The multiplication factor can be large, typically in the range between 10^4 and 10^6 . The multiplication gain can be stabilized and made less sensitive to the bias voltage as well as the particle energy or counting rate by addition of a small amount of a polyatomic gas to the rare gas. A typical gas mixture is 10 percent methane and 90 percent argon, known as “P-10 gas.” Carbon dioxide is also a common quench gas, and has the added benefit that it is much less prone to polymerization than methane. Polymerized methane can collect on the fine wires and cause severe localized gain reduction. This is particularly undesirable in a position-sensitive detector since it can distort the position linearity. Proportional counters have been extensively used as x-ray spectrometers. They are particularly useful for low-energy x rays because the intrinsic detector amplification reduces the effects of external amplifier noise, and because they can be used in a windowless mode. Measurements have been made for energies less than 1 keV (kiloelectronvolt).¹⁰ At high x-ray energies, the sensitivity is limited by the reduced absorption of x rays in the gas, although the energy discrimination for even high-energy x rays that are absorbed remains good so long as the secondary electrons are stopped in the detector. The effective energy range is dependent on the gas pressure and type, the detector size and shape, and on the presence of applied magnetic or electric fields, which can cause the secondary electrons

to adopt spiral paths. X rays with energy as high as 100 keV have been measured with special detector designs.¹¹

One of the most useful applications of proportional counters for x rays is position determination.¹² If the charge current in an anode wire is collected from both ends of the wire, resistive separation in the wire can be used to determine the position along the wire at which the charge is collected. Position resolution of fractions of millimeters can be achieved this way. Furthermore, if a number of parallel anode wires are used, the relative charge collection on adjacent wires can be used to determine the position of the x-ray absorption between the wires. In this way, two-dimensional position determinations can be achieved. At the same time, energy discrimination can be obtained by combining the charge signal for a given event from both ends of the anode wire and from adjacent wires. This has been used for a number of applications, notably measurement of x-ray diffraction distributions and imaging of astrophysical x-ray sources.¹³ Another popular method of encoding the position is the delay-line technique. In this method, the anode avalanche induces charge in a segmented cathode which is connected to the nodes of a lumped-element L-C delay line. Measurement of the arrival time of the induced charge pulse at either end of the delay line provides a measurement of the position of the event along the wire. This method avoids the need to fabricate the rather delicate resistive wire needed in the charge-division method, and such detectors are therefore more robust. This same method can also be used in the second dimension, encoding the multiple anode wires described above. A third technique uses the same interpolation of partial charges induced on the segmented cathode as described earlier for the multi-anode 2-D detector. In either plane, this technique involves significantly more complexity than the delay-line method.

Geiger Counter If the electric field in an x-ray proportional counter is increased beyond the proportional region (often by making the diameter of the anode wire smaller), the electron multiplication increases catastrophically, resulting in large-current pulses that are no longer proportional to the x-ray energy. This is called the *Geiger region*. In fact, the multiplication is large enough that it is usually necessary to include components in the gas mixture that will quench the resulting electrical discharge. One benefit of this mode of detection is the large amplification that makes it possible to detect x rays with simplified electronics. This is why the most common use of such detectors is in portable or remote radiation monitors.

Semiconductor Detector In gases, x-ray absorption can produce electrons and positive ions that can be separated and measured by an applied electric field. In solids, the situation is somewhat different. In a conductor, the electrons that are excited by absorption of x rays are quickly dissipated in energy and lost among the abundant conduction electrons in the material, and the positive charge is almost immediately neutralized by conduction electrons. Furthermore, an electric field cannot be maintained in the material. In an insulator, it is easy to establish an electric field to sweep out excited electrons. However, the positive charge remains behind, trapped and immobile. The electric field due to this trapped charge quickly cancels the applied field. As a result, initial charge pulses from incident x rays decrease in size and intensity as a result of such polarization effects.¹⁴ However, for semiconductors, the situation is much more favorable. By doping with an appropriate impurity or formation of a surface barrier, a rectifying *pn* junction can be formed that allows an electric field to be established in the solid. The thickness of the field region, called the *depletion layer* because it is free of mobile charge carriers, depends on the applied voltage and on the resistivity of the bulk semiconductor.

X rays absorbed within the depletion layer ionize bound electrons with energy equal to the difference between the absorbed x ray and the binding energy of the electron. These energetic electrons are free to move in the solid and produce further ionization by interaction with bound electrons. In this sense, the situation is similar to x-ray absorption in ionization chambers. Indeed, semiconductor detectors are sometimes referred to as *solid-state ion chambers*. However, there is an important difference. Instead of the ionization producing a heavy, slow moving positive ion as in a gas, or an immobile positive charge as in an insulator, it produces a positively charged "hole." The hole is a vacancy in the electronic valence band of the solid, with low effective mass and high mobility, closer to that of free or conduction electrons than to positive ions. The negatively charged electrons and

positively charged holes are swept apart by the electric field in the depletion layer, producing a current pulse in the external circuit. Contrary to the ion chamber, both the electron and hole motion contribute to the short-term charge pulse. The amplitude of the charge pulse is proportional to the energy of the absorbed x ray. The energy resolution depends on the statistics of the ionization process and on the electronic noise in the associated electronics, which depends in part on the capacitance of the semiconductor *pn* junction. Detailed reviews of semiconductor detectors are given in Refs. 1, 15, and 16.

To increase the depletion layer thickness and therefore the sensitive volume of the detector, very high-resistivity semiconductor material is required. For silicon, this is achieved by drifting lithium ions through the diode at an elevated temperature in the presence of an electric field.^{17–19} The lithium compensates bound impurities in *p*-type silicon. Such detectors are called *lithium-drifted silicon*, or Si(Li), detectors. An alternative is to use very high-purity germanium (HpGe). Both Si(Li) and HpGe detectors are cooled (typically by liquid nitrogen) to reduce electronic noise due to thermally generated electrons and holes [and for Si(Li) to prevent out diffusion of the compensating lithium ions]. Typically, the energy resolution for such detectors is between 120 and 140 eV.

The availability of high-resistivity silicon, grown using the floating zone technique, has made it possible to fabricate detectors using planar fabrication techniques taken from the microchip industry, which perform almost as well as the Si(Li) devices, but without requiring cryogenics.²⁰ Good performance is achieved using Peltier coolers to bring the detector to around -30°C .

This same technology has allowed a new generation of devices based on manipulating the photogenerated charges within the silicon bulk to bring the charge produced over a large area to a low-capacitance collection point several millimeters away from the point of generation. This provides a large-area detector with a small readout capacitance, and hence good collection solid angle while maintaining low electronic noise. These devices are called silicon drift detectors (SDDs).²¹ Single units are commercially available and compete strongly with Si(Li) for low to moderate energies. Since they are low-capacitance devices they can handle large count rates and maintain good resolution. Arrays of such devices are becoming available and should soon become ubiquitous in certain applications.

High electric fields in solid state devices can lead to avalanche charge multiplication in a similar manner to the gas proportional counter. In silicon, these devices are called avalanche photodiodes (APDs).²² Gains of around 100 are possible, and the high fields lead to fast pulses, of order 1 nanosecond. Consequently, they have a high count rate capability, several 10s of megahertz. They have some energy resolution, around 20 percent.

Wide bandgap semiconductors, such as CdZnTe or CdTe, have been used because of their stronger absorption and the possibility of room temperature operation. Such materials frequently exhibit trapping of minority carriers (typically holes) and therefore can show reduced resolution and sometimes polarization effects. They are of particular interest for use in arrays (typically by segmentation of one of the electrodes) for medical or astrophysical applications.^{23,24}

Channel Electron Multipliers X rays and visible photons can be detected by photocathodes and electron multipliers. Photoelectrons that are released from thin films or surfaces, accelerated, and electrostatically focused on another surface produce increasing numbers of secondary electrons as they move from surface to surface, until they are collected on an anode and measured as a charge pulse in an electronic circuit. The first such multipliers, similar to photomultiplier tubes, were introduced in the early 1960s. These were largely replaced with simpler and more stable semiconducting glass or ceramic tubes with a voltage applied between the input and output of the tube. These can be thought of as continuous-dynode electron multipliers. Modern channel electron multipliers (CEM) are made from lead glasses, which contain PbO, SiO₂, and several weight percent of alkali metal oxides. These are light, compact, and quite rugged and are used in x-ray pulse-counting detectors in applications ranging from electron microscopy to x-ray astronomy. The electron gain depends on the secondary electron yield of the surfaces employed, the applied voltage, and the channel diameter and length. If gas is present in the channel, ionization can take place, resulting in further electron multiplication and higher apparent gain. However, ionization

can lead to feedback resulting from positive ions accelerating back up the channel, leading to long-term “ringing” of the signal due to electrons released by accelerated ions. Ion feedback can be suppressed if the channel is operated at very high vacuum. A more convenient solution is to curve the channel so that positive ions do not accelerate to high-enough velocity to produce secondary electrons before they strike the channel wall. Gain saturation results when the supply of charge to the channel wall lags the depletion rate by secondary electron emission. In practice, operational gains of from 10^4 to 10^6 are possible.⁴

Very high counting rates, and especially imaging of x rays, are possible by the use of microchannel plates (MCP), which are arrays of very fine bore CEMs, typically with channel diameters and channel separations of a few 10s of micrometers.²⁵ Curving the channels in MCPs to suppress ion feedback is difficult and expensive, so two (or more) tilted channel MCPs are put together to form “chevron” multipliers in modern detectors. Energy resolution in CEM detectors is poor, and their efficiency decreases with increasing x-ray energy. Their simplicity, low cost, and, for MCPs, high spatial resolution make them very useful for many applications. Microchannel plates are also used as electron multipliers for low-noise amplification when used with scintillation detectors.

Scintillation

Incident x rays can excite electrons in atoms or molecules to excited electronic states, which then deexcite by returning to their ground state with accompanying photon emission. The excitation can be caused directly by the x ray or, more usually, by secondary energetic electrons excited by the x ray. The emitted photons (usually in the optical region of the spectrum) can be measured by film and also by photomultiplier tubes or by other electron multipliers such as MCPs. Various plastics, liquids, organic crystals, inorganic crystals, and gases can act as scintillators.²⁶ Polycrystalline solid light-emitting materials are sometimes called *phosphors*. The deexcitation (and therefore, light emission) can be almost instantaneous or delayed. Both types are used for x-ray detectors.

Activated Phosphor The most used scintillator detectors that also give the photon energy spectrum involve scintillating crystals with rapid response. These are often single crystals of sodium or cesium iodide with thallium added to activate the photo response and shift emission into the visible range. Because alkali halides are frequently hygroscopic, they must be sealed. Scintillators are almost never used for measurement of x rays with energy less than about 1 keV. The scintillating crystal is commonly mounted directly onto the face of the vacuum photomultiplier tube, which provides photoelectron conversion, and low-noise amplification of the electrons. The conversion efficiency, η , which is the fraction of incident energy that appears as scintillation, is typically about 5 percent for CsI(Tl) and 12 percent for NaI(Tl).²⁷ The quantum detection efficiency (QDE) is defined as the number of photoelectrons produced per incident photon and depends on absorption in the cladding, the thickness and absorption coefficient of the scintillator, and the nature of the photocathode of the photomultiplier tube.⁵ The energy resolution of a well-matched NaI(Tl) system is $\Delta E/E \sim 1.7E^{-1/2}$. Activated alkali halide scintillation detectors have working ranges of a few kiloelectronvolts to a few megaelectronvolts. They have good dynamic range with upper detection rates limited by the scintillator decay time, which is approximately 1 μ s for NaI(Tl). Organic scintillators have faster (nanosecond) decay times but have lower efficiency and energy resolution. Scintillation detectors are convenient to use, can be large (several square centimeters), are readily available commercially, and are used for a large number of x-ray applications.

Scintillators are often used for x-ray imaging applications. In such cases, energy discrimination is often not important, and other phosphors, such as rare-earth oxides activated with terbium [e.g., $\text{Gd}_2\text{O}_3\text{S}(\text{Tb})$, $\text{La}_2\text{O}_3(\text{Tb})$, $\text{Y}_2\text{O}_3(\text{Tb})$] or organic phosphors [e.g., tetraphenyl butadiene (TPB)] are used. The phosphor is deposited as a thin layer (1 to 10 μ m) directly on a glass plate, which is viewed by a sensitive television camera, or directly on the face of a fused-glass fiber optic, which is tapered to allow the intensity distribution to match the size and shape of a charge-coupled (CCD) or charge-injected (CID) imaging camera. This type of scintillator-fiber-optic-CCD imaging detector has become the standard for synchrotron-based protein crystallography beamlines, since it

produces a digital image which is amenable to computer analysis. At the time of writing, however, it is becoming clear that large area direct-detection semiconductor detectors will soon overtake these CCD-based detectors for this application, since they provide better point-spread function and faster readout.²⁸ Fast readout becomes important when the exposure time is much less than the detector readout time.

Restimulable Phosphor In the last decade, a different type of scintillation detector, commonly referred to as an imaging or computed radiography plate, has become widely used for x-ray imaging.^{29,30} This makes use of phosphor materials in which the electron excitation is to a metastable state that can have long (minutes to hours) decay time. After exposure, the plate is scanned by a laser beam that stimulates the metastable excited states to deexcite, producing light that is detected by a photomultiplier tube. The time of the light emission gives the position of the x-ray exposure and the intensity of the light, the x-ray intensity. Such detectors have a number of useful features:

1. They produce a digital record that leads conveniently to computer processing, transmission, and storage.
2. They have a linear response (as opposed to film) with a large dynamic range.
3. The thickness and nature of the phosphor used can be varied to accommodate the need (e.g., low-energy x rays, high-energy x rays, neutrons, etc.), and the size of the plate can be adapted to existing systems. For example, such plates are frequently used to replace photographic film.
4. The photostimulable plates can be reused by “erasing” the stored image by exposure to intense white light for a few seconds. At the present time, a restimulable scanning plate system (including exposure plates, scanner, and eraser) can be less expensive and more flexible than other commercially available position-sensitive detectors, although some with automatic readout features are comparable or more expensive. Such systems, although faster than film-based systems, do not have the rapid readout needed for some applications. In particular, they are much slower than the CCD-based detectors discussed above. In addition, the spatial resolution of these systems is limited to around 50 μm , which makes them inferior to film for high-resolution applications.

Film

The oldest, and still the most widely used x-ray detector, is silver halide-based photographic film. Indeed, the mysterious exposure of light-protected film led to discovery of x rays by Roentgen in 1895, and to demonstration of their use for medical imaging, still the most widely used and important application. Metastable excitation of grains of silver halide crystals is induced by x rays or secondary electrons. The resulting latent image is developed by chemical reduction of the excited grains to produce the familiar photographic image. Sometimes a phosphor coating or an adjacent phosphor plate is used to increase the sensitivity to the penetrating x radiation, with the halide excitation being produced by the secondary light emission from the phosphor. An extensive discussion of the use of film for optical imaging is given in Vol. I of this *Handbook*. Much of that discussion also applies to x-ray excitation. Detailed discussion of the use of film for x-ray detection and imaging is given in Refs. 1 and 5.

Cryogenic

Recent development of superconducting cryogenic x-ray detectors provides opportunity for important new applications, particularly in materials analysis and astrophysics. These make use of techniques to reach and maintain very low temperature (30 to 70 mK) for long periods of time.³¹ They use the measurement of the temperature rise (microcalorimeter),^{32–35} or tunnel junction current³⁶ in superconducting materials.

Microcalorimeter Measurement of the temperature rise induced by absorption of an x-ray photon has been demonstrated by the use of semiconductor (typically Si or Ge) thermistors whose electrical conductivity is temperature dependent, and with bimetallic transition-edge sensors (TES). Of these, the TES detectors are the most studied and appear to have the highest potential for x-ray spectrometry. They operate by holding the temperature of the metallic absorber at the transition edge between the superconducting and normal state. At the transition edge, the conductivity is extremely temperature sensitive. The temperature rise that is induced by an absorbed x ray is detected by measuring the conductivity with a superconducting quantum (SQUID) detector. The transition-edge temperature is typically about 50 mK and is stabilized by thermoelectric feedback. Energy resolution as low as 2 eV and counting rates as high as 500 counts per second (cps) have been reported.^{37,38} Both the energy resolution and the thermal recovery time (therefore the counting rate) are affected by the thermal capacitance of the absorber, which must be as low as possible. Absorber sizes of $300 \times 300 \times 50 \mu\text{m}$ are typical for a very high-resolution detector. Detector arrays are under development to increase the effective area, the count rate capability, and to provide imaging for astrophysical applications.³⁰ Alternatively, polycapillary focusing optics³⁹ have been used to increase the effective area (to $>7 \text{ mm}^2$).

Superconducting Tunneling Junction (STJ) Another type of cryogenic detector involves x-ray-induced breakup of superconducting Cooper pairs, which leads to decreased tunneling current across a superconducting Josephson junction. Although such detectors must also be small and maintained at low temperature to keep the thermally induced current across the junction low, the dependence of the counting rate on the thermal capacitance is relaxed. Such detectors have demonstrated energy resolution $<20 \text{ eV}$ and counting rate $>20,000 \text{ cps}$.³³ Again, arrays or optics have been used to increase the effective area.

60.3 SUMMARY

A list of properties of a number of single-pixel x-ray detectors is provided in Table 1 for comparative purposes. Spatial resolution for typical position-sensitive or imaging detectors is given in Table 2.

TABLE 1 Typical Parameters That Affect Detector Choice, Including Counting Rate Limitations, Availability of Current Mode Operation, and Energy Resolution

Detector	Intensity measurements		
	Pulse Counting (cps)	Current Mode	Energy Resolution
Ionization			
Ionization chamber	1–10 k	Yes	1–10 keV
Proportional counter	10 k		2–20 keV
Geiger counter	50 k		
Semiconductor	0–100 k		
Si(Li)			120 eV
HpGe			140 eV
CdZnTe			250 eV
Pin diode		Yes	
Scintillation detector	100–500 k		20–50%
Cryogenic			
Microcalorimeter detector	100–500 k		2–10 eV
STJ detector	2–20 k		20–50 eV

TABLE 2 Position-Sensitive X-Ray Detectors

Detector	Spatial Resolution (μm)
Film	20–200
Ionization	
Multiwire proportional counter	100–300
Semiconductor	
Arrays	
Segmented electrode	100–500
Charge-coupled detector (CCD)	20–50
Charge injection detector (CID)	20–50
Active pixel	
CMOS	20–50
Amorphous silicon	50–100
Pulse height dispersive	100–300
Scintillation	
Segmented	500–2000
Video	50–100
Multichannel plate readout	20–100
Restimulable phosphor	50–200
Cryogenic	
Array	
Microcalorimeter	200–500
Tunnel junction	300–500
Dispersive	500–1000

60.4 REFERENCES

1. W. J. Price, *Nuclear Radiation Detection*, 2nd ed., McGraw-Hill, New York, 1964.
2. G. G. Knoll, *Radiation Detection and Measurement*, Wiley, New York, 1979.
3. W. R. Leo, *Techniques for Nuclear and Particle Physics Experiments*, Springer-Verlag, Heidelberg, 1987.
4. C. F. G. Delaney and E. C. Finch, *Radiation Detectors*, Oxford University Press, Oxford, 1992.
5. C. J. Buckley, in *X-Ray Science and Technology*, A. G. Michette and C. J. Buckley, (eds.), Institute of Physics Publishing, Bristol, 1993, pp. 207–252.
6. G. Fraser, *X-Ray Detectors in Astronomy*, Cambridge University Press, Cambridge, 1989.
7. S. J. Gurman, “EXAFS Studies in Materials Science,” *J. Mat. Sci.* **17**(6):1541–1570 (1982).
8. E. A. Stearn and S. M. Heald, in “Investigation of Soft Radiation by Proportional Counters,” *Handbook on Synchrotron Radiation*, Vol. 1b, E. E. Koch, (ed.), North Holland, Amsterdam, 1983, p. 955.
9. D. G. Stearns and M. B. Stearns, “Microscopic Methods in Metals,” in *Topics in Current Physics*, U. Gonser, (ed.), 49, Springer, Berlin, 1983, p. 153.
10. S. C. A. Curran, A. L. Cockroft, and J. Angus, *Phil. Mag.* **40**:929 (1949).
11. B. D. Ramsey, J. J. Kolodziejczak, M. A. Fulton, J. A. Mir, and M. C. Weisskopf, “Development of Microstrip Proportional Counters for X-Ray Astronomy,” *Proc. SPIE* **2280**:110–18 (1994).
12. G. Charpak, R. Bouclier, T. Bressani, J. Favier, and C. Zupancic, *Nucl. Instr. and Meth. in Phys. Res.* **62**: 262–268 (1968).
13. B. D. Ramsey, R. A. Austin, and R. Decher, *Sp. Sci. Rev.* **69**:139–204 (1994).
14. R. Hofstadter, *Nucleonics* **4**:2 (1949); *Proc. IRE* **38**:721 (1950).
15. G. L. Miller, W. M. Gibson, and P. F. Donovan, *Ann. Rev. Nucl. Sci.* **12**:189 (1962).
16. W. M. Gibson, G. L. Miller, and P. F. Donovan, in *Alpha, Beta, and Gamma Spectroscopy*, K. Siegbahn (ed.), North Holland, Amsterdam, 1964.

17. E. M. Pell, *J. Appl. Phys.* **31**:291 (1960).
18. J. H. Elliott, *Nucl. Instr. and Meth.* **12**:60 (1961).
19. J. L. Blankenship and C. J. Borkowski, *IRE Trans. on Nucl. Sci.* **NS-9**:213 (1963).
20. J. Kemmer, *Nucl. Instr. and Meth. in Phys. Res. A* **226**(1):89–93 (1984).
21. E. Gatti and P. Rehak, *Nucl. Instr. and Meth.* **225**:608 (1984).
22. A. Q. R. Baron and S. L. Ruby, *Nucl. Instr. and Meth. A* **343** (2-3):517–526 (1994).
23. J. E. Gindley, T. A. Prince, N. Gehrels, J. Tueller, C. J. Hailey, B. D. Ramsey, M. C. Weisskopf, P. Ubertini, and G. K. Skinner, *SPIE Proc.* **2518**:202–210 (1995).
24. F. P. Doty, H. B. Barber, F. L. Augustine, J. F. Butler, B. A. Apotovsky, E. T. Young, and W. Hamilton, *Nucl. Instr. and Meth. In Phys. Res.* **A253**:356–360 (1994).
25. S. Dhawan, *IEEE Trans. Nucl. Sci.* **NS-28**:672 (1981).
26. R. B. Murray, in *Nuclear Instruments and Their Uses*, A. H. Snell (ed.), Wiley, New York, 1962.
27. R. L. Heath, R. Hofstadter, and E. B. Hughes, *Nucl. Instr. and Meth.* **162**:431 (1979).
28. E. F. Eikenberry, et al., *Nucl. Instr. and Meth. A* **501**(1):21 pp. 260–266 (2003).
29. H. Nanto, K. Murayama, T. Usuda, F. Endo, Y. Hirai, S. Tahiguchi, and N. Tagekuchi, *J. Appl. Phys.* **74**:1445–1447 (1993).
30. H. Nanto, Y. Hirai, F. Endo, and M. Ikeda, *SPIE Proc.* **2278**:108–117 (1994).
31. M. Altman, G. Angloher, M. Buhler, T. Hertrich, J. Hohne, M. Huber, J. Jochum, et. al., *Proc. of 8th Int. Workshop on Low Temperature Detectors* (1999).
32. D. A. Wollman, K. D. Irwin, G. C. Hilton, L. L. Dulcie, D. E. Newbury, and J. M. Martinis, *J. Microscopy* **196**–223 (1997).
33. D. McCammon, W. Cui, M. Juda, P. Plucinsky, J. Zhang, R. L. Kelley, S. S. Holt, G. M. Madejski, S. H. Moseley, and A. E. Symkowiak, *Nucl. Phys.* **A527**:821 (1991).
34. D. Lesyna, D. Di Marzio, S. Gottesman, and M. Kesselman, *J. Low Temp. Phys.* **93**:779 (1993).
35. E. Silver, M. LeGros, N. Madden, J. Beeman, and E. Haller, *X-Ray Spectrom.* **25**:115–122 (1996).
36. M. Frank, L. J. Hiller, J. B. Le Grand, C. A. Mears, S. E. Labov, M. A. Lindeman, H. Netel, and D. Chow, *Rev. Sci. Instrum.* **69**:25 (1998).
37. K. D. Irwin, G. C. Hilton, J. M. Martinis, S. Deiker, N. Bergren, S. W. Nam, D. A. Rudman and D. A. Wollman, “A Mo–Cu Superconducting Transition-Edge Microcalorimeter with 4.5 eV Energy Resolution at 6 keV,” *Nucl. Instrum. Meth. A* **444**(1–2):184–187 (2000).
38. D. A. Wollman, G. C. Hilton, K. D. Irwin, N. F. Bergren, D. A. Rudman, D. E. Newbury, and J. M. Martinis, 1999 NCSL Workshop and Symposium.
39. D. A. Wollman, C. Jezewski, G. C. Hilton, Q.-F. Xiao, L. L. Dulcie, and J. M. Martinis, *Microsc. Microanal.* **4**:172–173 (1998).

This page intentionally left blank

ADVANCES IN IMAGING DETECTORS

Aaron Couture

*GE Global Research Center
Niskayuna, New York*

Recent advances have allowed digital x-ray detectors to become widely available commercially, replacing many traditional film based systems. Digital x-ray detectors have also enabled a wide variety of new applications. This chapter summarizes the current state of the art for digital x-ray imaging detectors.

61.1 INTRODUCTION

X-ray detectors have a history of over one hundred years of providing rapid, simple, and low cost images of internal structure for a wide range of applications. Medical and dental x-ray imaging are used heavily in screening (e.g., mammography and dental check-up imaging), diagnostic (e.g., cardiac and angiographic imaging), and surgical (e.g., cardiac catheterization) procedures (see Chap. 31). Over a million medical x-ray procedures per day are performed worldwide. Industrial x-ray imaging is used to penetrate dense materials and provide high-resolution information about the internal composition and structure of manufactured parts. Security applications include airport screening of baggage and personal items, imaging of cargo containers and rail cars, as well as detection of land mines. In the past decade, digital x-ray detectors have begun to replace film screen systems due to decreased cost, improved performance, and workflow. Digital x-ray detectors continue to enable new applications not previously possible: tele-radiology, computer aided detection (CAD), digital subtraction contrast enhancement, and computed tomography (CT) are prominent examples. The cumulative growth of digital systems in the market has also accelerated, with the number of medical systems in the field from less than one thousand in the year 2000 to more than 30,000 in 2007.

Each imaging application brings new and different performance requirements for x-ray imaging detectors. Some of these competing detector requirements include: size, resolution, readout rate, noise, image quality, cost, weight, portability, and power. The following section will summarize a number of key technologies that represent the current state of the art in x-ray imaging detectors. A wide variety of concepts for digital x-ray detectors have been explored both academically and commercially. The technologies detailed in this chapter represent recently developed digital x-ray detectors. Photo-stimulated phosphors (CR plate) detectors, which currently compose a large fraction of the digital x-ray market, are not included here, but are introduced in Chap. 60.

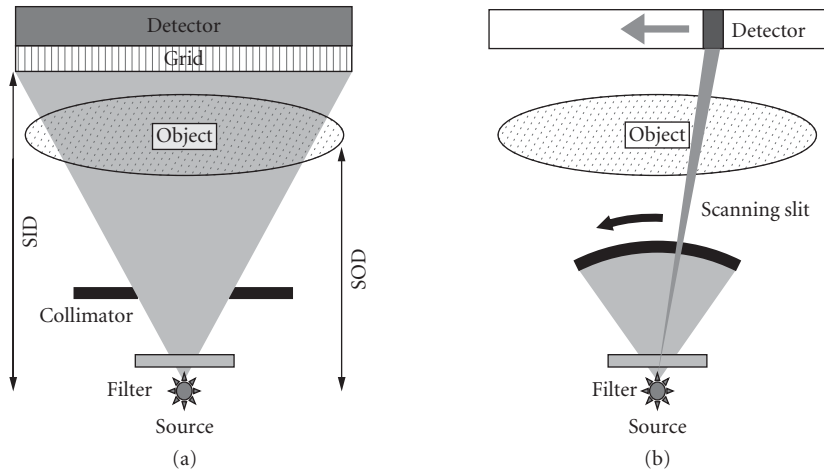


FIGURE 1 (a) Area type x-ray imaging system using a collimator to define the x-ray beam as well as a grid for scatter rejection. Source to object (SOD) and source to image distance (SID) are shown. (b) Slot type x-ray imaging system with fan beam scanning illumination and linear detector.

X-ray imaging is typically performed in one of two geometries (Fig. 1): area detection or slot detection. In area detection, the x-ray source is typically a few millimeters in size and is collimated to expose only the active area of the detector. The magnification of the system is the ratio of the source-to-image distance (SID) to the source-to-object distance (SOD). Especially in thick specimens, scatter can significantly degrade x-ray images. For area detection, a grid is typically inserted directly on the detector surface. Additionally, increasing the object to image distance decreases the scattered radiation hitting the detector; however, the fraction of the object imaged also decreases. In radiographic applications the maximum dimension to be imaged in a single shot is the chest, which sets the typical size for imaging at roughly 40 cm. In slot detection, a collimated fan beam and single slit linear detector are simultaneously scanned across the object during x-ray illumination. Slot detectors have the capability to reject scatter, and also reduce the size and cost of the detector. However, the x-ray tube is subjected to higher power and image acquisition is slower than area detection.

Another classification of x-ray detectors is the mode used for readout, including integrating and continuous. Integrating detectors collect signal from a gated x-ray source prior to triggering the readout of the detector. Detector readout is performed following illumination by the x-ray source. Data conversion electronics can be multiplexed so that a single converter channel is dedicated to an entire column of pixels. Continuous readout detectors output the instantaneous x-ray signal intensity during constant illumination. Depending on the gain, continuous readout detectors may also provide photon counting and energy information (see Chap. 62). Continuous readout requires a single conversion channel dedicated to each pixel. To maintain low power and cost, x-ray imaging detectors typically use integrating readout mode. Finally, x-ray imaging systems can be designed for single shot applications, termed radiographic, as well as for sequenced “video-like” imaging, termed fluoroscopic. For fluoroscopic applications, the fraction of signal from prior frames that contribute to later frames is quantified by the detector lag (0 to 100%).

For virtually every imaging application it is important to minimize x-ray dose. For medical applications x-ray dose can be harmful to both patients and doctors. For industrial and security applications, tube power output is limited, and x-ray dose determines the amount of time required to produce an image. The detective quantum efficiency (DQE) is a metric used widely to quantify the image quality (IQ) of x-ray imaging detectors. Note that this is different from the quantum detector efficiency used for single pixel detectors. The detective quantum efficiency quantifies the ability of a detector to accurately

transfer an x-ray input image into output electronic or digital signal, normalized to input x-ray dose (X). A detector with high DQE can deliver equivalent IQ for lower x-ray dose. It is computed as a function of spatial frequency f ,

$$\text{DQE}(f) = \frac{[S \cdot \text{MTF}(f)]^2}{\text{NPS}(f) \cdot X \cdot C} \quad (1)$$

where C is the x-ray fluence, S is the x-ray conversion efficiency, and NPS is the noise power spectrum, the spatial or temporal noise added to the image. MTF is the modulation transfer function, the ratio of the amplitude of the output image at a spatial frequency f to the input amplitude. MTF is a measure of the spatial resolution.

The spatial resolution of the detector is influenced not only by the pixel pitch, but also by spreading of signal to adjacent pixels. The conversion efficiency of the detector is a function of the efficiency of the absorbing layer to generate electronic signal as well as the fraction of detector area that is sensitive to x rays (fill-factor). In detectors that have electronic noise, the DQE of a detector degrades with lower x-ray dose. Electronic noise is suppressed in detectors that have inherent gain, such as image intensifiers. Flat panel detectors with no active gain have inherent electronic noise related to the transfer of small amounts of charge from the pixels in the array to electronics bonded to the panel. Some x-ray imaging detectors add electronic amplification in order to improve performance at low x-ray doses.

X-ray imaging detectors must also be insensitive to gain hysteresis, sometimes referred to as ghosting. An x-ray detector exposed to both high and low dose conditions must continue to have a uniform response. If the detector is hysteretic, the gain for regions of high exposure can be modified, leading to contrast appearing in subsequent images. Ghosting can lead to image artifacts building up over time. For acceptable quality medical images, the gain hysteresis must be limited to a few percent.

X-ray scatter can degrade the image quality in x-ray images and is especially a problem with thick specimens. The imaging geometry as well as the use of scatter-rejection grids can greatly improve the contrast-to-noise ratio (CNR) under such conditions. To improve the rejection ratio, grids can be manufactured with septa focused on the x-ray source and made from materials with high stopping power. The imperfect x-ray transmission of the grid can degrade the DQE, so high transmission fraction is important in maintaining image quality. Recently, the use of high aspect ratio microlithography (HARM) has been applied to manufacturing focused metal grids.¹ The high aspect ratio as well as dense material could potentially offer enhanced scatter rejection while maintaining good x-ray transmission.

61.2 FLAT-PANEL DETECTORS

Introduction

The most common x-ray imaging detectors available include: flat-panel indirect-conversion detectors, flat-panel direct-conversion detectors, and charge-couple-device (CCD) based detectors. Figure 2 shows a diagram of these three types of detectors, including an electrical schematic of a typical pixel. Each detector includes an absorbing layer responsible for stopping a fraction of incident x rays, a conversion layer resulting in electric charge, and a switching matrix responsible for storing and reading out charge. Flat panel detectors are typically based on amorphous or polysilicon on glass substrates, while CCD detectors are fabricated on single crystal silicon.

Both indirect and direct-conversion flat panel x-ray detectors utilize an array of amorphous silicon (a-Si) thin film transistors (TFT) to form an active array of switches for signal readout. Amorphous silicon has a number of advantages for use in x-ray detectors. First, the liquid crystal display (LCD) industry is based on very similar active matrix a-Si TFT panels. Market pressures have driven displays toward large area, low cost, and low defects. Photolithography, thin film deposition, and etching equipment can be

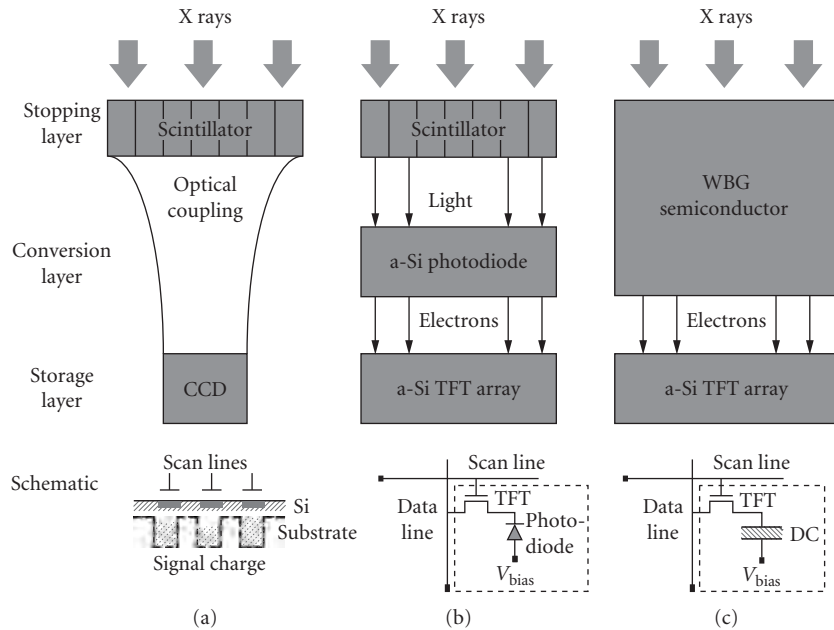


FIGURE 2 (a) Scintillator-based CCD detector with optical coupling; (b) indirect-conversion flat panel detector with scintillator layer optically coupled to photodiode layer, integrated onto a-Si TFT panel; and (c) direct-conversion detector with wide band gap (WBG) semiconductor absorption layer deposited directly onto a-Si TFT panel.

directly incorporated into manufacturing of x-ray detectors. Flat panel x-ray detectors with active area of more than 40 cm and pixel pitches less than 100 μm are available. Second, the a-Si TFT has very low leakage, allowing storage of signal charge for many seconds during x-ray exposure. Last, amorphous silicon is naturally radiation hard.

Figure 2 shows a schematic for pixels in indirect and direct-conversion flat panel detectors. The TFT switching matrix and readout method for flat panel x-ray detectors is the same. Flat panel detectors utilize scan lines that are individually energized to address rows of TFT pixels, as well as data lines that transfer signal charge to readout electronics at the border of the panel. The readout is multiplexed so that all pixels on a single line are converted simultaneously, utilizing a single converter for each data line. Multiplexed readout coupled with a panel design optimized for fast switching has enabled the use of flat panel detectors in fluoroscopic medical applications with 30 Hz frame rates. Readout electronics are designed to satisfy competing requirements for rapid readout, low noise, or low power.

Flat panel detectors also have advantages for compact design and weight. Portable radiography detectors are currently available with thicknesses of a few centimeters and weights under 15 lb. Wireless digital x-ray detectors have recently been demonstrated for radiography.

Flat Panel: Indirect Conversion

Indirect-conversion flat panel x-ray detectors utilize a layer of scintillator material to absorb x rays and emit visible light photons. The scintillator material must be high density in order to efficiently stop x rays, must efficiently convert absorbed x-ray energy into visible photons, and efficiently transfer visible light photons to the photosensing layer. The scintillator material must be a uniform layer across the imaging area, typically hundreds of micrometer thick and up to 40 cm in size. The scintillator

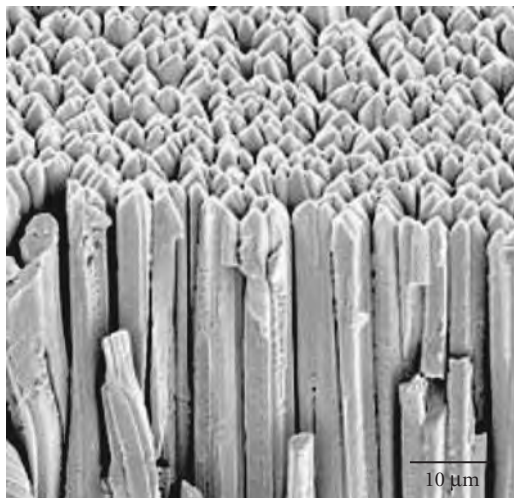


FIGURE 3 Thin film deposited CsI(Tl) scintillator, typically deposited in thicknesses from 100 μm to millimeters. Needle microstructure is responsible for enhancing scintillator spatial resolution.

material is typically a deposited thin film, a plastic sheet, or a solid plate. Thallium doped cesium iodide, CsI(Tl), is widely used as a scintillating material due to its high conversion efficiency, limited after-glow, and capability for thin film deposition, as well as an emission spectrum that matches well with a-Si photodiode absorption. A single x ray can generate thousands of optical photons. The optical photons are emitted isotropically, leading to spreading of the light signal before it is absorbed in the photosensing layer, degrading the spatial resolution and hence the DQE of the detector. Thin film deposited CsI(Tl) can be grown with a structured needle morphology that tends to limit lateral spreading of scintillation light, as shown in Fig. 3. Use of a structured scintillator allows thick layers (500 μm and greater) to be deposited while maintaining good spatial resolution (MTF) of the detector. Thallium doping boosts the photoelectron yield and also shifts the emission peak from 315 to 550 nm; both effects improve the conversion efficiency for a-Si diode photo detectors. In addition, lithographically patterned structuring of thin film deposited scintillator has also been investigated.² The visible light generated by the scintillator is captured by an amorphous silicon photodiode layer and converted to electric charge that can be stored in the pixel using the a-Si TFT array. In addition to requirements related to conversion efficiency and light spreading, which directly affect the DQE of the detector, the scintillator also has requirements related to lag, after-glow, and hysteresis. Typically lag related to the CsI(Tl) can be less than a few percent, and the gain hysteresis is also limited to less than a few percent.

The typical pixel design of an indirect-conversion flat panel x-ray detector is shown in Fig. 2. An amorphous silicon photodiode layer is deposited onto a TFT active matrix, and a common bias is provided to bias the array. Leakage through the photodiode degrades the dynamic range of the detector and can lead to spatial artifacts and temperature sensitivity. However, amorphous silicon diodes have been reported³ with leakage values of $<100 \text{ pA/cm}^2$. Fluoroscopic lag in direct-conversion flat panels can be less than 10 percent.⁴ An etched mesa structure is the most typical photodiode geometry. However, in order to maximize the active area covered by the photodiode array and increase conversion efficiency of the detector, continuous photodiode layers have also been investigated.⁵

The DQE of indirect-conversion flat panel detectors is improved by the x-ray absorption in dense, high brightness scintillators. DQE for mammography detectors at 8.5 mR has been reported in excess of 80 percent at 0 frequency, with greater than 30 percent at 5 lp/mm (5 line pairs per mm is 100 μm

resolution). The DQE for these detectors is degraded mainly by light spreading in the scintillator as well as electronic noise in the TFT readout. Electronic noise sources originating from pixel switching as well as transferring the charge on the data line are significant for indirect-conversion flat panel detectors. Active areas of research include TFT pixel configurations with gain stage at each pixel as well as optimization of panel and converter electronics in order to reduce the electronic noise of the panel.⁶ Additionally, some designs incorporate an additional storage capacitor device to increase the charge that can be stored in each pixel.⁷ Diode switching can also be used in place of the TFT array.⁸ Lightweight and rugged substrates have been investigated, as well as the use of ink-jet printing of organic electronics for the TFT and photodiode.⁹

Flat Panel: Direct Conversion

Direct-conversion detectors, as shown in Fig. 2, combine the x-ray absorbing layer and conversion layer into one material. One advantage of direct-conversion detectors is the reduced manufacturing cost, due to a reduced number of layers on the TFT flat panel. Additionally, since the signal charge created in the conversion layer is subjected to a strong electric field, there is no degradation of the spatial resolution due to spreading of the charge in the direct-conversion material. Last, the fraction of area that is sensitive to x rays is large due to shaping of the electric field in the conversion layer. Direct-conversion detectors are commercially available for medical as well as industrial imaging, typically for single shot imaging systems. Fluoroscopic mode detectors also have been demonstrated.¹⁰

Direct-conversion materials must be optimized for multiple requirements simultaneously: high x-ray absorption (high Z , high density), efficient charge collection, low dark current, good uniformity, low lag, and long term reliability, and stability. The properties of amorphous selenium, mercuric iodide, and lead iodide, three materials that have been investigated for direct-conversion x-ray imaging, are summarized in Table 1. Amorphous selenium is currently the most widely used material in commercially available direct-conversion detectors. The x-ray absorption efficiency scales with the atomic number as Z^4 . Formation energy relates to the energy required to create a charge pair, and is inversely proportional to the number of charges liberated from each absorbed x ray. The mobility lifetime product relates to the ability to remove charge from the conversion layer. The electric field relates to the ability to operate the detector electronics at lower voltages.

Figure 4 shows the thin film layer structure for direct-conversion materials. Lead and mercuric iodide are deposited directly onto the TFT array, using physical vapor deposition (PVD) or screen-printing of semiconductor powder incorporated in a polymer binder (particle in binder, PIB). The bias electrode is deposited in a separate step, and a polymer layer is used to encapsulate the full structure for environmental isolation. Amorphous Se is typically deposited in a layered p-type/intrinsic/n-type (PIN) structure with PVD using lightly doped p-Se and n-Se as blocking contacts that reduce the leakage current, as shown in Fig. 4b. With bias applied, x rays that are absorbed in the direct-conversion material generate charge pairs that are swept to the pixel and bias contacts. Charge is stored in the pixel until readout of the detector, which occurs in the same way for indirect-conversion detectors, as described in the beginning of Sec. 61.2. Leakage current in the direct-conversion material can degrade

TABLE 1 Properties of Typical Direct-Conversion Materials Used in X-Ray Imaging Detectors^{11,12,13}

	HgI ₂	PbI ₂	a-Se
Atomic number (Z)	80, 53	82, 53	34
Band gap (eV)	2.1	2.3	2.2
Charge pair formation energy (eV)	5	5.5	42
Mobility-lifetime product ($\mu\tau$, cm/V ²)	10 ⁻⁵ (h) 10 ⁻⁵ (e ⁻)	10 ⁻⁶ (h) 10 ⁻⁷ (e ⁻)	10 ⁻⁶ (h) 10 ⁻⁶ (e ⁻)
Electric field (V/ μ m)	0.2–1	0.2–1	10

Mobility-lifetime products for holes (h) and electrons (e) can also be dependent on temperature.

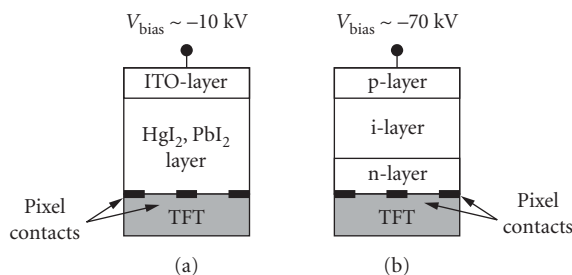


FIGURE 4 Typical configuration and biasing for direct-conversion thin films for (a) mercuric and lead iodide and (b) amorphous selenium. Indium tin oxide (ITO) and p-type/intrinsic/n-type (PIN) top contacts are important in blocking injection of carriers into the semiconductor as well as for removing signal charge.

the dynamic range of the detector and can also generate artifacts if the leakage is not stable with time or temperature. Leakage currents have been reported as low as $700\text{pA}/\text{cm}^2$ for mercuric iodide.¹⁴ Gain hysteresis and ghosting have been well studied during the development of direct-conversion detectors. The zero frequency DQE performance for direct-conversion flat panel detectors is reported to be in excess of 70 percent for amorphous selenium mammography detectors.¹⁵

61.3 CCD DETECTORS

A third class of digital x-ray imaging detectors utilizes a scintillation layer coupled to a high efficiency optical system which transfers the image onto a smaller (2 to 3 cm^2) CCD photo sensor. Applications for CCD detectors vary from medical imaging, mammography, and video rate industrial imaging, to protein crystallography.¹⁶ CCD systems have advantages in cost and low electronic noise related to the photo sensor; however, some degradation in DQE occurs due to inefficiency in the optical system that couples the scintillator to the CCD. A higher reduction ratio used to couple signal to the CCD results in lower efficiency, due to losses in the optical system that are fundamentally limited by solid angle considerations.

Small area prototype flat-panel CCD detectors have been demonstrated with a scintillator deposited directly on the photo sensor.¹⁷ An additional stage of optical amplification can be added in order to compensate for the inefficiency of the optical demagnification.¹⁶ The main elements of CCD x-ray imaging detectors are shown in Fig. 2. They include a scintillating layer for absorbing x rays which outputs visible light that is coupled to an optical system that focuses the image onto the CCD photosensor. The CCD is composed of rows of metal gates patterned on silicon with implanted regions that act as columns. The gates are biased to define the separation of pixels, and photocharges that are generated in the depletion region of the silicon are stored in the charge wells. To read out charge, the biasing voltage on the gates can be changed to shift charge to neighboring rows. Charge transport in CCDs can be highly efficient, resulting in readout noise levels of less than 10 electrons.¹⁶

Figure 5 shows three examples of common configurations for CCD detectors including both area detectors and slot scanners. Figure 5a and b show two tiling methods used to increase the size of the x-ray image, while limiting the demagnification ratio for the detector. Fiber optic tapers used for optical coupling are fused bundles of glass fiber light guides. The bundles are heated and shaped into tapers for demagnification, with near theoretical efficiencies.¹⁶ Tiles must not appear in the final x-ray image, so there must be no change in imaging performance across the boundary of adjacent units. Figure 5c shows the configuration for a CCD array used in a slot scanning geometry. Tiled units with scintillator and fiber optics are constructed into a linear detector that is translated across the imaging plane.

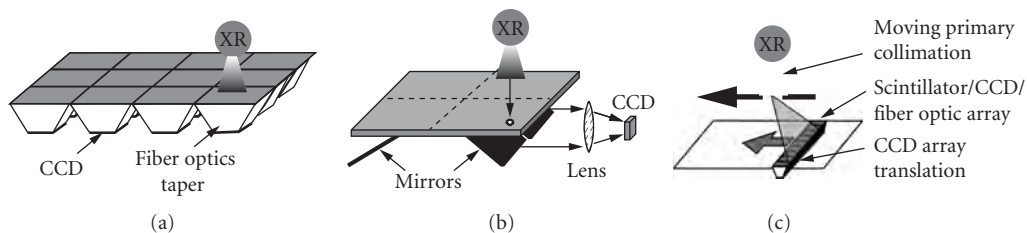


FIGURE 5 Three configurations of scintillator/CCD digital x-ray imaging detectors: (a) tiled detector elements with scintillator coupled to a fiber optic taper and CCD; (b) tiled detector with single scintillator plate plus multiple mirrors optically coupled to CCD; and (c) slot scanning geometry, linear tiled detector elements with scintillator coupled to a fiber optic taper and CCD.

Scintillators for CCD systems can be either optically coupled solid plate or plastic film, or thin film deposited directly on optical elements. The scintillating material has similar requirements and performance to that for indirect-conversion flat panel detectors; however the spectral response of the scintillator must be tailored to the absorption of CCD detectors that peak at red visible wavelengths.

61.4 CONCLUSION

Numerous digital x-ray imaging technologies have created a rapidly expanding medial and industrial x-ray imaging market. This chapter has focused on the most mature technologies; however, rapid market growth is currently driving the development of higher performance and lower cost systems. A number of new technologies are currently under development that have the potential to further improve image quality, increase detector readout rates, and reduce system cost.

61.5 REFERENCES

1. K. Fischer, B. Chadhuri, H. Guckel, and C. Tang, "Fabrication of Two-Dimensional X-Ray Anti-Scatter Grids for Mammography," *Advances in X-ray Optics, Proceedings of SPIE* **4145**:227–234 (August 2000).
2. V. Nagarkar, S. Tipnis, V. Gaysinskiy, S. Miller, A. Karellas, and S. Vedantham, "New Design of a Structured CsI(Tl) Screen for Digital Mammography," *Proceedings of SPIE* **5030**:541–546 (June 2003).
3. S. Tchakarov, P. Cabarrocas, U. Dutta, P. Chatterjee, and B. Equer, "Experimental Study and Modeling of Reverse-Bias Dark Currents in PIN Structures Using Amorphous and Polymorphous Silicon," *Journal of Applied Physics* **94**:7317–7327 (2003).
4. P. Granfors, R. Aufrichtig, G. Possin, B. Giambattista, Z. Huang, J. Liu, and B. Ma, "Performance of a $41 \times 41 \text{ cm}^2$ Amorphous Silicon Flat Panel X-Ray Detector Designed for Angiographic and R&F Imaging Applications," *Medical Physics* **30**:2715–2726 (2003).
5. R. Weisfield, W. Yao, T. Speaker, K. Zhou, R. Colbeth, and C. Proano, "Performance Analysis of a 127-Micron Pixel Large-Area TFT/Photodiode Array with Boosted Fill Factor," *Proceedings of SPIE* **5368**:338–348 (May 2004).
6. L. Antonuk, Y. Li, H. Du, Y. El-Mohri, Q. Zhao, J. Yamamoto, A. Sawant, Y. Wang, and Z. Su, "Investigation of Strategies to Achieve Optimal DQE Performance from Indirect Detection, Active Matrix Flat-Panel Imagers (AMFPIs) through Novel Pixel Amplification Architectures," *Proceedings of SPIE* **5745**:18–31 (February 2005).
7. D. Albagli, S. Han, A. Couture, H. Hudspeth, C. Collazo, and P. Granfors, "Performance of Optimized Amorphous Silicon, Cesium-Iodide Based Large Field-of-View Detector for Mammography," *Proceedings of SPIE* **5745**:1078–1086 (February 2005).

8. C. van Berkel, M. Powell, and S. Deane, "Physics of a-Si:H Switching Diodes," *Journal of Non-Crystalline Solids* **164–166**(2): 653–658 (1993).
9. R. Street, W. Wong, S. Ready, R. Lujan, A. Arias, M. Chabinyc, A. Salleo, R. Apte, and L. Antonuk, "Printed Active-Matrix TFT Arrays for X-Ray Imaging," *Proceedings of SPIE* **5745**:7–17 (February 2005).
10. D. Hunt, O. Tousignant, and J. Rowlands, "Evaluation of the Imaging Properties of an Amorphous Selenium-based Flat Panel Detector for Digital Fluoroscopy," *Medical Physics* **33**:1166–1175 (2004).
11. D. Alexiev, N. Dytlewski, M. I. Reinhard, and L. Mo, "Characterisation of Single-Crystal Mercuric Iodide," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **517**:226–229 (2004).
12. R. Street, M. Mulato, S. Ready, R. Lau, J. Ho, K. VanSchuylenbergh, M. Schieber, et al., "Comparative Study of PbI₂ and HgI₂ as Direct Detector Materials for High Resolution X-ray Image Sensors," *Proceedings of SPIE* **4320**:1–12 (June 2001).
13. G. Belev, and S.O. Kasap, "Amorphous Selenium as an X-Ray Photoconductor," *Journal of Non-Crystalline Solids* **345–346**:484–488 (2004).
14. G. Zentai, L. Partain, R. Pavlyuchkova, C. Proano, G. Virshup, L. Melekhov, A. Zuck, et al., "Mercuric Iodide and Lead Iodide X-Ray Detectors for Radiographic and Fluoroscopic Medical Imaging," *Proceedings of SPIE* **5030**:77–91 (February 2003).
15. J. Jesneck, R. Saunders, E. Samei, J. Zia, and J. Lo, "Detector Evaluation of a Prototype Amorphous Selenium Based Full Field Digital Mammography System," *Proceedings of SPIE* **5745**:478–485 (February 2005).
16. S. Gruner, M. Tate, and E. Eikenberry, "Charge-Coupled Device Area X-Ray Detectors," *Review of Scientific Instruments* **73**:2815–2842 (2002).
17. D. Scheffer, "A Wafer Scale Active Pixel CMOS Image Sensor for Generic X-Ray Radiology," *Proceedings of SPIE* **6510**:651000–651001 (February 2007).

This page intentionally left blank

X-RAY SPECTRAL DETECTION AND IMAGING

Eric Lifshin

*College of Nanoscale Science and Engineering
University at Albany
Albany, New York*

The concept of an x-ray image is usually associated with a radiograph (also see Chap. 31).¹ X rays are transmitted through an object and an image is formed on a detector placed on the side opposite the source. Contrast is based on point-to-point variation in absorption and the resolution is determined by scattering effects and by the pixel size of the detector. Radiographs are collected in parallel, that is, the entire image is formed at one time either traditionally by film or more recently by solid state imagers of various types. The transmission mode is of great value because the penetrating power of x rays makes it possible to see structure within objects be they humans or metallurgical castings. Images containing surface detail of the type created with a reflection optical microscope are difficult, if not impossible, to form because lens systems similar to those used in optical microscopes are generally not available with the exception of zone plates and reflective multilayer optics (see Chaps. 40 and 41). The index of refraction for most materials for x rays is so close to one that glass or other lens materials used for focusing the visible, infrared, or ultraviolet parts of the electromagnetic spectrum will not work for x rays. Thus, images are formed by transmitting x rays through the object of interest onto an image plane. In the most conventional forms of radiography the magnification on the detector plane is unity and any magnification in the image is the result of optically or electronically magnifying the detected image.

Spectral imaging, on the other hand, often involves creating a map of an object based on the spatial distribution of x rays of one or more particular energies. This is usually not done in transmission by the formation of a kind of mono-energetic radiograph, but rather by examining the emission of x rays excited from an area of a material selectively bombarded with some form of ionizing radiation. Furthermore, parallel imaging is not used as in the case for radiography. Instead the ionizing radiation is focused to a point on the object surface and then scanned over that surface in the form of a two-dimensional raster analogous to the electron beam scan on a cathode ray tube (CRT) display such as a conventional television tube. The image is formed in a serial or sequential mode, but if the scanning is fast enough and the persistence of the CRT phosphor long enough, it will look like an image formed in parallel. The focused (or highly collimated) ionizing beam used to cause x-ray emission can consist of electrons, ions, or x rays. In addition to the mode of scanning the beam over the surface of the sample, the beam can also be stationary and the sample physically scanned under a static beam. Either mode of operation forms the basis for a variety of analytical microscopy techniques including scanning electron microscopy (SEM), the analytical electron microscopy (AEM), proton induced x-ray emission (PIXIE), x-ray microfluorescence (XRF), and focused ion beam microscopy (FIB).

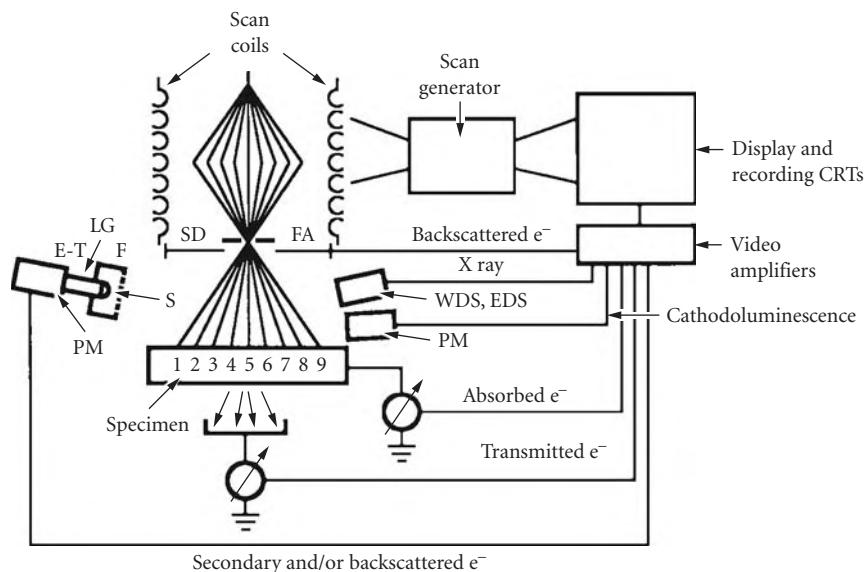


FIGURE 1 Basic SEM operation. An electron beam is scanned over the surface of sample covering the surface plane. Points 1 to 9 shown here are just in one dimension. The scan generator also controls the scan on a CRT display. The brightness of the CRT is determined by the intensity of a selected signal which as shown here could be the secondary electron signal detected with an Everhart-Thornley detector (ET) or various other detectors including EDS and WDS x-ray detectors.

As an example, the basic operation of an SEM, the most popular embodiment of this approach, is shown in Fig. 1. A focused beam of electrons is scanned over the surface of a specimen causing the emission of secondary electrons at each point it strikes. The emitted secondary electron signal is detected and its intensity used to modulate the brightness of a synchronously scanning CRT or other display controlled by the same scan generator. The magnification of the image is simply the ratio of the distance scanned on the display to that scanned on the sample. It is simply varied by changing the scan area on the sample. Variations in the intensity of the secondary electrons emitted from each point forms the basis of the observed image contrast. This point-to-point variation signal can be due to differences in the secondary electron yield arising from differences in surface topography or the material properties of the sample. SEMs typically operate in the magnification range from about $20\times$ up over $1,000,000\times$ with their ultimate resolution determined by how finally the electron beam can be focused, the volume from which the detected secondary electrons originate, and signal to noise ratio associated with the small electron beam currents used. Figure 1 also shows the use of x-ray detectors used to create images based on the point-to-point variation of the x-ray emission of a single energy.

Before discussing x-ray imaging it is first useful to review how x-ray analysis is done from a single point, as for example, one based on a detail observed in an SEM image. Currently the most common spectrometers/detectors used are the energy dispersive detector (EDS) and the wavelength dispersive (WDS) detector. A detailed description of the operation of an EDS detector can be found in Ref. 2 and an introduction is given in Chap. 60. The basic concept is that the x-ray energy of each photon detected is used to create electron-hole pairs in a solid state device. The number of electron-hole pairs produced is proportional to the energy of the x-ray photon detected. The initial electron pulse is converted into a voltage pulse whose size remains proportional to the x-ray energy. The pulses are then sorted by voltage and displayed as a histogram of pulse intensity versus voltage. The use of reference samples makes it easy to adjust the gain such that the x scale is calibrated to give energy rather than voltage.

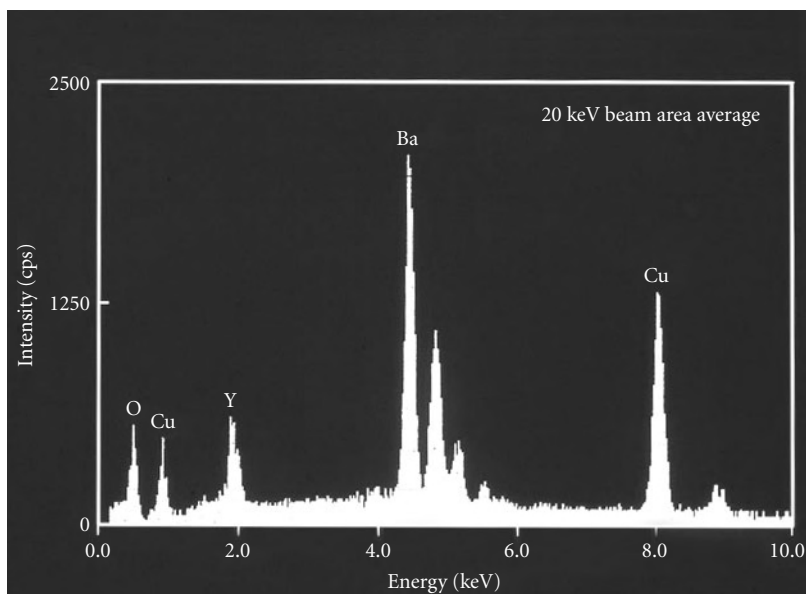


FIGURE 2 EDS spectrum of high temperature superconductor.

Figure 2 gives a typical x-ray spectrum, in this case of a high temperature superconductor, YBCO. Peaks can usually be identified easily as a result of Moseley's law, which relates the energy of a given spectral series to that of the atomic number of the element from which it originates. The characteristic lines observed are the result of the primary electron beam ejecting core shell electrons and the subsequent emission of x-ray photons with an energy equal to the difference between that of the core shell from which the electron was ejected and that of the shell from which the electron drops down to fill the vacancy (see Chap. 29). The range of x-ray energies used in SEM-EDS analysis is typically from about 100 eV up to about 20 KeV, which covers all elements from Be on up in atomic number. The lines used are either K, L, or M lines depending on whether the core shell electron was ejected from the first, second, or third energy level as defined by the principal quantum number of that level being 1, 2, or 3. To ensure an adequate signal intensity the electron beam energy used is typically 2 or 3 times the excitation energy of the core shell electrons associated with a particular element and shell. Since SEMs usually operate at 30 keV or less the lines used in an analysis are generally the K lines for the first third of the periodic table, the L lines for the middle third, and the M lines for the highest atomic number elements. Exceptions occur in cases of serious peak overlaps in the observed spectra.

EDS detectors have been used with SEMs since the late 1960s and since that time performance has improved significantly. In particular the energy resolution as measured at Mn $K\alpha$ has improved from about 500 eV to better than 130 eV. The development of thin windows capable of withstanding an atmosphere difference in pressure between the detector and the specimen environment combined with very low noise electronics has made Be $K\alpha$ (0.109 keV) analysis possible. Very sophisticated user interfaces have been developed to provide for the rapid identification of peaks, background removal (mostly from the x-ray continuum), peak area determination (needed for quantitative analysis), and a range of data display options including x-ray mapping. The past few years have seen another major advance that will revolutionize EDS analysis. It is the development of silicon drift detectors that will eventually replace lithium drifted detectors of the type that has dominated the field for well over 30 years. These detectors³ have all of the features mentioned above, but can increase throughput by 10 to 100 times resulting in count rate capabilities as high as a million counts per second when those signal levels are available. Until recently it was nearly impossible to get count rates greater than about 5000 counts per second without

paying a serious penalty in detector resolution. This was a serious limitation in x-ray mapping for two reasons in particular. First, when doing a point-by-point map the dwell time per pixel may be minimal (on the order of microseconds) depending on the scan rate used and the number of pixels used. Thus the number of counts obtained even from an element present at a high concentration level could be so small that counting statistics would make it very difficult to determine that the signal is above the background level. Second, in EDS the count rate of importance is the total count rate entering the detector. If a minor constituent is to be mapped then its proportion of the total count rate may be so low that it will be totally obscured by the high count rate of the major peaks. The only way to overcome the above difficulties is to take very long scans, but that may lead to problems arising from instrument drift and/or sample contamination. An example of how these difficulties can be overcome with the use of the new silicon drift technology is shown in Fig. 3 where x-ray maps were obtained in 12 s.

In an SEM, characteristic x-ray photon production is a very inefficient process compared to the generation of secondary electrons where the yield and collection of secondary electrons can even be greater than one per incident electron. In fact, the number of x-ray photons collected can be less than one per billion incident electrons. The number of incident electrons is closely related to the electron probe size on the sample and thus for example a 100 nm probe that could contain 10^{10} electrons striking the sample per second may yield only a relatively small number of measured x rays of a given elemental line even from a pure element sample. At 1.0 nm, which is about the smallest size of probes now in use, the yield is considerably less. The spatial resolution of x-ray maps collected in this way is ultimately determined by the x-ray excitation volume, which depends strongly on the electron beam energy, the sample composition, and the spectral line selected. Figure 4 shows a Monte Carlo simulation of electron scattering in copper for different beam energies.³ It clearly shows that the volume of electron scattering can be considerably larger than the electron beam size. Although electron energies are tracked until they are essentially zero, as long as the energy exceeds the ionization energy for a given spectral line, x rays will be produced. In this example that volume will be larger for the lower excitation energy Cu $L\alpha$ line than for the Cu $K\alpha$ line. While reducing the beam energy may help it must be pointed out that the intensity of the emitted line is a strong function of the ratio of the beam energy to

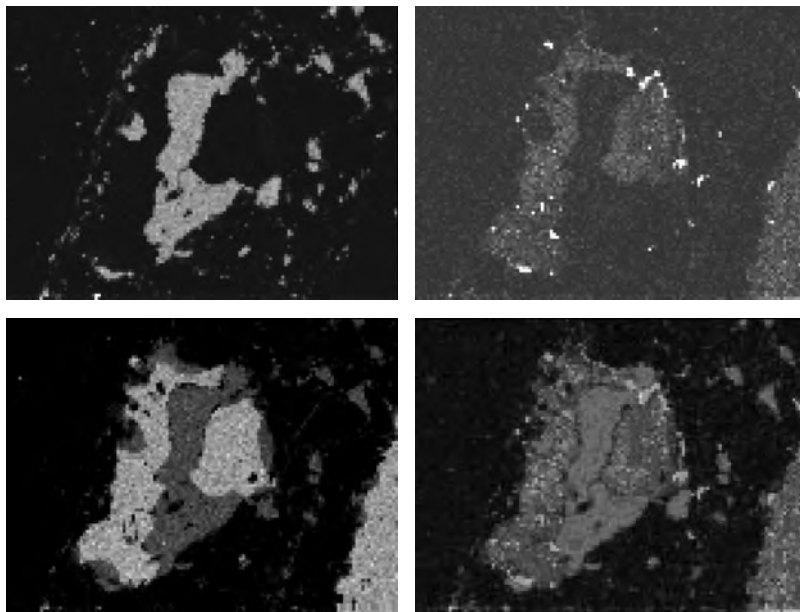


FIGURE 3 A 12-s x-ray spectrum image (128×96 pixels; 1 ms) obtained with a silicon drift detector. (See also color insert.) (Courtesy of Dale Newbury NIST.)

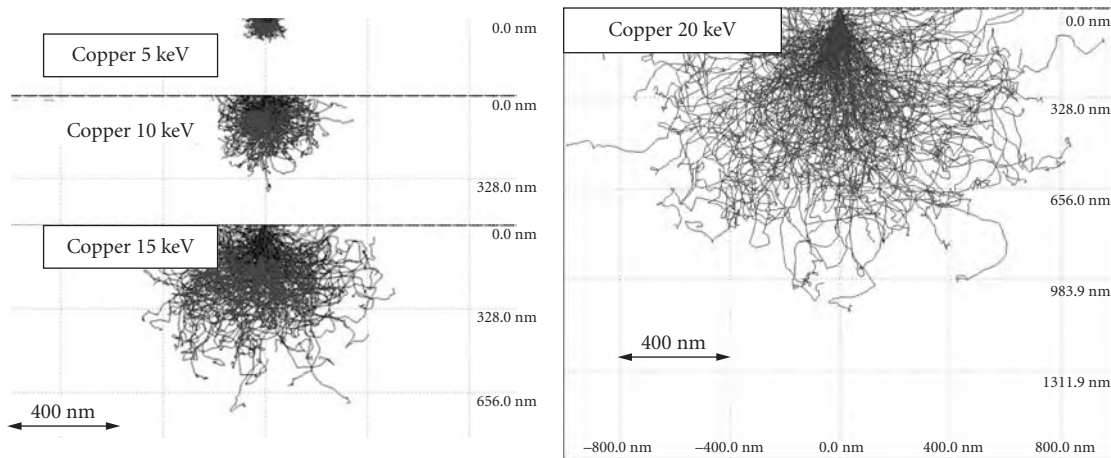


FIGURE 4 Monte Carlo simulations of electron trajectories in copper at various beam energies. Note the significant increase in the scattering volume with increasing energy. (See also color insert.)

the excitation energy, so the resolution approaches its best value as the generated x-ray intensity goes to zero. Traditionally x-ray microanalysis has had a resolution limit of about $1.0\ \mu\text{m}$ for the reasons cited when thick samples are studied. A combination of careful line selection and beam energy can drop this value to about $100\ \text{nm}$ at best. To do better, thin samples must be prepared to limit electron scattering and higher beam energies used. This is the approach of the analytical transmission electron microscope where the limit of spatial resolution for chemical analysis can be around $10\ \text{nm}$.

If larger areas are to be examined and very high spatial resolution is not required, x-ray mapping by x-ray fluorescence can be very effective. It also provides better detection limits since x-ray excited x-ray spectra have much less background due to less of the continuum in the spectrum (the continuum background arises only from scattering of the primary x-ray source since x rays do not generate continuum spectra from the sample). Figure 5 is an example of an x-ray map generated using x-ray excitation and a scanning specimen stage. While the spatial resolution of this type of image can be extended down to about $15\ \mu\text{m}$, it is expected that, as more advanced synchrotrons are implemented, x-ray probes based on collimated beams could be $50\ \text{nm}$ or less. Focused high energy ion beams, particularly protons, can also be used to create x-ray maps, but once again the yield can be low and the resolution will not be much better than $1000\ \text{nm}$. Medium energy focused ion beams, such as the $30\ \text{keV}$ used in an FIB, can be focused to less than $10\ \text{nm}$; however, the x-ray generation cross-sections are low, as are the beam currents, so that no measurable x rays are observed.

As stated previously, most x-ray spectroscopic imaging involves point-by-point (serial) data collection, by either scanning the probing beam or the specimen stage. There has also been some recent work with controlled-drift detectors to create two-dimensional x-ray detectors that provide spectral information from an array of points without scanning. For example, Castoldi et al. use polycapillary optics (see Chap. 53) to excite an array of points on a specimen and then another capillary optic to image the excited x rays on the surface of a two-dimensional detector with a $36\ \text{mm}^2$ detector containing 180 by $180\ \mu\text{m}$ pixels.⁴ In this manner they were able to image the $\text{Cu K}\alpha$ line on a fine pitch printed circuit board used as a sample. While this resolution is considerably less than that of the scanning methods used in the SEM, the method may still be useful for dynamic studies of samples where the sampling time may be very short since the parallel detection process can be quite fast. Another approach to parallel detection has been used in conjunction with CdZnTe and CdTe detectors based on pixelated CMOS signal processing in which both the numbers of x-ray photons and their average energy can be determined. This work has been recently described by Kruger et al.,⁵ but the spatial resolution is also limited due to the pixel sizes currently possible.

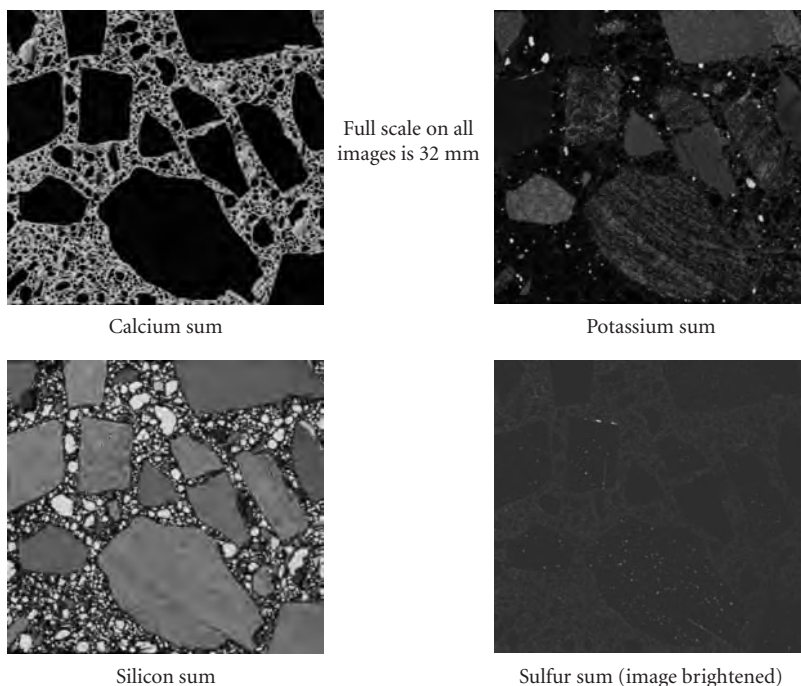


FIGURE 5 Micro-XRF maps. (Courtesy of J. M. Davis, NIST.)

62.1 REFERENCES

1. E. Krestel, *Imaging Systems for Medical Diagnostics*, Siemens Aktiengesellschaft, Berlin, pp. 221–249, (1990).
2. J. I. Goldstein, D. E. Newbury, D. C. Joy, C. E. Lyman, P. Echlin, E. Lifshin, L. Sawyer, and J. Michael, *Scanning Electron Microscopy and X-Ray Microanalysis*, 2nd ed., Kluwer Academic/Plenum Press, New York, NY, pp. 297–323 (2003).
3. D. Drouin, A. R. Couture, D. Joly, X. Tastet, V. Aimez, and R. Gauvin, “CASINO V2.42—a Fast and Easy-To-Use Modeling Tool for Scanning Electron Microscopy and Microanalysis Users” *Scanning* **29**(3):92–101 (2007).
4. A. Castoldi, C. Guazzoni, R. Hartman, and L. Strueder, “Application of Controlled-Drift Detectors to Spectroscopic X-Ray Imaging,” *IEEE Nuclear Science Symposium Conference Record* 1003–1008 (2007).
5. H. Krueger, J. Fink, E. Kraft, N. Wermes, P. Fischer, I. Peric, C. Herrmann, M. Overdick, and W. Ruetten, “CIX—A Detector for Spectral Enhanced X-Ray Imaging by Simultaneous Counting and Integrating,” arXiv.org, e-Print Archive, *Physics*, 1–12, arXiv:0802.2247v1 [physics.ins-det] (2008.)

SUBPART

5.6

NEUTRON OPTICS AND APPLICATIONS

This page intentionally left blank

NEUTRON OPTICS

David Mildner

*NIST Center for Neutron Research
National Institute of Standards and Technology
Gaithersburg, Maryland*

63.1 NEUTRON PHYSICS

The neutron is a subatomic particle with zero charge, a mass of $m = 1.00897$ atomic mass units, a spin of $1/2$, and a magnetic moment of $\mu_n = -1.9132$ nuclear magnetons. These four basic properties make thermal and cold neutrons not only such a rich and useful scientific tool for the investigation of condensed matter, but also a basis for observing many beautiful optical phenomena with remarkable properties,¹ as well as for constructing numerous optical devices.² “Thermal” neutrons are those in thermal equilibrium with their surroundings near room temperature. When a beam of energy, E , given by

$$E = \frac{2\pi^2\hbar^2}{m\lambda^2} \quad (1)$$

is selected from a thermal distribution, the de Broglie wavelength, λ , is comparable to interatomic distances in condensed matter, where $h = 2\pi\hbar$ is Planck’s constant. Consequently neutron *diffraction* in condensed matter is analogous to x-ray diffraction. However, a typical thermal neutron has an energy of 0.025 eV or 4×10^{-21} J, much lower than electromagnetic radiation of comparable wavelength. Consequently, the mass and energy are such that the frequency of the radiation is comparable with the vibrational frequencies found in materials, which makes the measurement of the *inelastic scattering* a useful probe of these vibrations. Hence, neutrons are ideally suited for the study of the atomic structure and dynamics in condensed matter. The magnetic moment of the neutron interacts with those of unpaired electrons in magnetic materials, giving rise to *magnetic diffraction* and *inelastic scattering*. Again, the wavelength and energy of thermal neutrons are ideal for the study of the magnetic structure and dynamics of spin systems. As a result of zero charge, the neutron has only a short-ranged interaction with the nucleus. This means that the amplitude of the interaction is small, so that neutrons penetrate into the bulk of most materials. Moreover, the interaction probability of neutrons varies irregularly with the nuclear isotope, unlike x rays whose amplitudes increase monotonically with the atomic number. Finally, the neutron spin of $1/2$ enables a neutron beam to exist in one of two polarized states. When the neutron scatters from a nucleus of nonzero spin, the strength of the interaction depends on the relative orientation of the neutron and the nuclear spin.

The propagation of neutron de Broglie waves in a potential field is analogous to the propagation of light waves in a medium with a continuously variable refractive index. The potential can be

gravitational, magnetic, or nuclear. For example, slow neutrons follow a parabolic path under the effect of gravity as in classical mechanics. Neutrons in a constant magnetic field experience a torque and undergo precession. In a nonuniform magnetic field they experience a force that depends on the relative orientation of the spin and field vectors. Thus, a nonuniform magnetic field is a birefringent medium for an unpolarized neutron beam, with results analogous to the Stern-Gerlach experiment. Neutrons can be focused by refraction in an inhomogeneous magnetic field provided by a magnetic hexapole. Neutron waves in bulk nonmagnetic materials interact with the atomic nuclei. Generally, the interaction gives rise to an isotropic spherical wave that reradiates the neutron beam incident on the nucleus. The potential of the neutron-nucleus interaction has an imaginary part that represents neutron absorption. A scattering length, b , that depends on a point-like interaction potential, describes the scattering of a beam of neutrons. Generally, b is positive (but not always) and varies with different isotopes, and even with the same isotope, depending on the relative spin orientations of the neutron and isotope.

The similarity of the mathematical descriptions for neutron wave and light propagation gives rise to phenomena that are analogous to those of classical optics. In fact, virtually all the well-known classical optical phenomena that are characteristic of light and x rays have also been demonstrated with neutrons. In geometric optics, there are not only the refraction and reflection of neutrons by materials, but also special properties in magnetic media. In wave optics, there is Bragg diffraction from crystalline materials, and so on, as for x rays, but there are also other phenomena completely analogous to classical optics. Various neutron optics experiments have demonstrated quantum-mechanical phenomena on a macroscopic scale. For example, the perfect crystal neutron interferometer can measure the relative phase between two plane wave states, with the large separation of the beams enabling easy access for material phase-shifting devices and magnetic fields. These provide both quantitative verification of various fundamental quantum-mechanical principles applied to neutrons and accurate measurements of neutron scattering lengths. Finally, neutron optical devices have been developed to transport, collimate, focus, monochromate, filter, polarize, or otherwise manipulate neutron beams for applications in both basic and applied research, such as the study of the microscopic structure and dynamics of materials.

The optical phenomena arise from coherent elastic scattering of neutrons in condensed matter.³⁻⁵ The neutron wave function, $\psi(\mathbf{r})$, can be described by a one-body stationary Schrödinger equation, upon which all neutron optics is based, viz.,

$$[-(\hbar^2/2m)\nabla^2 + V(\mathbf{r})]\psi(\mathbf{r}) = E\psi(\mathbf{r}) \quad (2)$$

in which E is the incident neutron energy. The optical potential $V(\mathbf{r})$ represents the effective interaction of the neutron with the medium. The scattering of a neutron by a single bound nucleus is based on the Born approximation and uses the Fermi pseudopotential, $V(\mathbf{r}) = (2\pi\hbar^2/m)b\delta(\mathbf{r})$, where $\mathbf{r}$ is the neutron position relative to the nucleus, to represent the effective interaction between the neutron and the nucleus. The wave function $\psi(\mathbf{r})$ provides a description of coherent elastic scattering and all neutron optical phenomena. There are other scattering processes (incoherent elastic scattering and inelastic scattering) collectively referred to as *diffuse scattering*. In addition, the incident neutron might be absorbed by the nucleus. Both of these cause attenuation of the coherent wave, $\psi(\mathbf{r})$, in the medium, so that the potential $V(\mathbf{r})$ and the bound scattering length, b , are in general complex.

The interactions of neutrons in bulk nonmagnetic matter are with atomic nuclei, and for neutrons traveling in a bulk medium the potential $V(\mathbf{r})$ may be replaced by a summation of pseudopotentials centered at each nucleus labeled i , given by

$$V(\mathbf{r}) = \sum_i (2\pi\hbar^2/m)b_i\delta(\mathbf{r}-\mathbf{r}_i) \quad (3)$$

This aggregate potential results in a coherent scattered wave that is the sum of the incident plane wave and the superposition of the spherical waves emanating from each nucleus. For a homogeneous system (gas, liquid, or amorphous solid) the optical potential has a constant value $V_0 = (2\pi\hbar^2/m)\rho b$, where ρ is the atom density and b is the average bound coherent scattering length per atom.

The Schrödinger equation is a macroscopic equation that describes coherent elastic scattering and all neutron optical phenomena in terms of the interaction of the neutron wave with a potential barrier.

The general solution in a medium of constant potential, V_0 , can be expressed as a superposition of plane waves where the magnitudes of the wave vectors can be determined by the incident neutron energy,

$$E = \frac{(\hbar\mathbf{k})^2}{2m} = \frac{(\hbar\mathbf{k}')^2}{2m} + V_0 \quad (4)$$

where $\mathbf{k}$ and $\mathbf{k}'$ are the incident and secondary wave vectors. Note that the neutron momentum $\mathbf{p} = \hbar\mathbf{k}$, and the magnitude of $|\mathbf{k}| = 2\pi/\lambda$, where λ is the neutron wavelength. For elastic scattering, the two equations, $|\mathbf{k}| = |\mathbf{k}'|$ (conservation of energy) and $\mathbf{k} = \mathbf{k}' + \mathbf{q}$ (conservation of momentum) combine, where $\mathbf{q}$ is the scattering vector whose magnitude is given by

$$|\mathbf{q}| = (4\pi/\lambda)\sin(\varphi/2) \quad (5)$$

where φ is the scattering angle, that is, the angle between $\mathbf{k}$ and $\mathbf{k}'$, and $\hbar\mathbf{q}$ is the momentum transferred from the neutron to the scattering system. The directions of the various wave vectors and their corresponding amplitudes are determined by requiring $\psi(\mathbf{r})$ and $\nabla\psi(\mathbf{r})$ to be continuous at the boundary between media. Some neutron optical phenomena are well described by the kinematic theory of diffraction. In geometric optics, neutron trajectories obey the same laws of reflection and refraction as in classical optics, though true mirror reflection only occurs for ultracold neutrons. Other phenomena require the dynamical diffraction theory that takes into account the interchange between the transmitted and reflected waves. Goldberger and Seitz⁶ have shown from the theory of dispersion that, with respect to neutron optics, all materials behave like a continuous macroscopic medium with a refractive index. In general, the propagation of de Broglie neutron waves in a potential field $V(\mathbf{r})$ is analogous to the propagation of light waves in a medium with a continuously variable refractive index that is defined by

$$n(\mathbf{r}) = [1 - V(\mathbf{r})/E]^{1/2} \quad (6)$$

In bulk media, $V(\mathbf{r})$ is replaced by V_0 . Sears¹ has given a more rigorous and comprehensive treatment of dispersion theory.

63.2 SCATTERING LENGTHS AND CROSS SECTIONS

The study of the neutron optics within materials requires understanding of scattering lengths and cross sections. This is also necessary for neutron scattering measurements in the study of the structure and dynamics of condensed matter. The scattering of a neutron by a single bound nucleus is described within the Born approximation by the Fermi pseudopotential because $V_0/E \ll 1$. The scattering length is a measure of the strength of the interaction of the neutron with the nucleus, and the intensity of neutron scattering depends on the cross section of the sample. Fortunately, for most nuclei this involves only s wave scattering, and consequently the scattering lengths and cross sections for thermal neutrons are independent of the neutron wave vector $\mathbf{k}$ (or wavelength $\lambda = 2\pi/|\mathbf{k}|$), whereas the absorption cross sections σ_a are inversely proportional to $\mathbf{k}$ (or inversely with velocity v), and therefore increase linearly with λ . For most nuclides, $V(\mathbf{r})$ is a strongly attractive potential and, therefore, the scattering length is positive. Indeed, there are only a few nuclides, such ^1H , ^7Li , ^{48}Ti , and ^{55}Mn , plus a few others, that have negative scattering lengths.

The general theory of neutron scattering lengths and cross sections has been summarized by Sears,^{7,8} and a compilation of recommended values is given in atomic data and nuclear data tables.⁹ The bound scattering length b of a nucleus is in general complex, given by

$$b = b' - ib'' \quad (7)$$

The scattering cross section is given by $\sigma_s = 4\pi\langle|b|^2\rangle$, where the brackets denote a statistical average over neutron and nuclear spin states, and is divided into coherent and incoherent contributions. The coherent scattering depends on the spatial correlations between scattering nuclei and gives rise to variations in scattered intensity as a function of the scattering vector. This, therefore, gives information regarding the atomic structure of the sample under study. The coherent scattering is responsible for the neutron optical effects that include reflection, refraction, diffraction, and interference. The incoherent scattering on the other hand is independent of the scattering vector and arises from the variance of the scattering lengths at each nuclear site. It depends only on the constituent nuclei and the density of the scattering material. In neutron diffraction measurements, including reflectometry and small-angle scattering, it is the incoherent scattering that gives rise to a featureless background, and its subtraction is important for data analysis, while the coherent scattering gives rise to the interference pattern.

The absorption depends on the imaginary part of the scattering length, and the absorption cross section is given by $\sigma_a = (4\pi/k)b''$. In practice, the imaginary part of the scattering length is of the same order of magnitude as the real part only when the absorption is large ($\sigma_a \approx 10^4\sigma_s$). Values of σ_a are tabulated for a neutron velocity of 2200 m s⁻¹ (equivalent to an energy of 25.3 meV, a wave vector $k \approx 3.49 \text{ \AA}^{-1}$, or a wavelength $\lambda \approx 1.798 \text{ \AA}$). Nuclides, such as ³He, ⁶Li, and ¹⁰B shown in Table 1, with strong neutron absorption have an imaginary part to their scattering length, and are the principal converters from neutrons to charged particles that are found in the majority of neutron detectors (see Sec. 63.8). There are also a few nuclides like ¹¹³Cd or ¹⁵⁷Gd that have low energy (n, γ) resonances, these are used for beam collimation and definition, and can also be used as foil detectors.

Furthermore, the absorption is usually large for only one spin state of the compound nucleus. For nuclides such as ¹¹³Cd, ¹⁵⁵Gd, or ¹⁵⁷Gd, the absorption is large only in the $J = I + 1/2$ state (neutron and nuclear spins are parallel), and $b_-'' = 0$. For nuclides such as ³He or ¹⁰B, the absorption is large only in the $J = I - 1/2$ state (neutron and nuclear spins are antiparallel), and $b_+'' = 0$. (This forms the basis of the nuclear-spin polarizing ³He filter; see Sec. 63.7). For a few nuclides such as ⁶Li, the absorption is large for both $J = I \pm 1/2$ states. Finally, the dependence of the scattering length on the particular isotope means that there is a further incoherence for a given element depending on the abundance of the various constituent isotopes.

Consider only elastic scattering and take the potential given in Eq. (3). The differential scattering cross section may be written as the sum of two contributions

$$N \frac{d\sigma}{d\Omega}(\varphi) = \sum_{jj'} \langle b \rangle^2 \exp(i\mathbf{q} \cdot \mathbf{r}_{jj'}) + \sum_{j \neq j'} \langle b^2 \rangle - \langle b \rangle^2 \quad (8)$$

where the first term is the differential coherent scattering cross section, dependent on the distance $\mathbf{r}_{jj'}$ between nuclei labeled j and j' , and summed over all nuclei N within the scattering system,

TABLE 1 Scattering Lengths and Cross Sections for ³He, ⁶Li, and ¹⁰B

	$I(\pi)^*$	$\epsilon^\dagger$ (%)	$b_+^\ddagger$	$b_-^\ddagger$	$b_c^\ddagger$	$b_i^\ddagger$	$\sigma_c^\S$	$\sigma_i^\S$	$\sigma_s^\S$	$\sigma_a^{\S\S}$
³ He	1/2(+)	0.00014	4.30	10.07 -5.93i	5.74 -1.48i	-2.5 +2.57i	4.42	1.53	6.0	5333
⁶ Li	1(+)	7.5	0.67 -0.076i	4.67 -0.63i	2.00 -0.26i	-1.89 +0.26i	0.51	0.46	0.97	940
¹⁰ B	3(+)	20.0	-4.2	5.2 -2.49i	-0.1 -1.07i	-4.7 +1.23i	0.14	3.0	3.1	3835

* $I(\pi)$ is the spin and parity of the nuclear ground state.

[†] ϵ is the natural isotopic abundance.

[‡]Values of scattering length b are in 10^{-13} cm or fm.

[§]Values of cross sections σ are in 10^{-24} cm² or barns.

^{§§} σ_a is the absorption cross section at $\lambda = 1.798 \text{ \AA}$.

Note: b_c , b_i , σ_c , σ_i , and σ_s are defined later in the text, and b_+ and b_- refer to the scattering lengths of the nonzero spin isotopes aligned parallel and antiparallel to the incident neutron spin.

where N is the number of nuclei or atoms, and the second term is the differential incoherent scattering cross section. The coherent scattering length of a mixture of nuclei is the mean scattering length, whereas the incoherent scattering length is the standard deviation of the scattering lengths from that mean. Scattering amplitudes add for coherent scattering, whereas scattering intensities add for incoherent scattering. Incoherent scattering arises from two contributions: one from spin incoherence for those nuclei that have nonzero spin (there are different scattering lengths depending whether the neutron and nuclear spins are parallel or antiparallel), and the other from isotope incoherence. The latter arises because a given element may have different nuclides, each with its own scattering length.

The neutron has a spin $1/2$, and if the nucleus has a nonzero spin I , it may be aligned either parallel or antiparallel to the incident neutron spin s . This gives rise to spin incoherence, and the bound scattering length is spin dependent with

$$b = b_c + \frac{2b_i s \cdot I}{\sqrt{I(I+1)}} \quad (9)$$

where b_c and b_i are the bound coherent and incoherent scattering lengths. The coupling of these spins gives rise to two different scattering lengths b_+ and b_- , each with different statistical weights,

$$w_+ = \frac{I+1}{2I+1}$$

and

$$w_- = \frac{I}{2I+1} \quad (10)$$

The bound coherent scattering length for the nucleus with spin I is therefore

$$b_c = \langle b \rangle = w_+ b_+ + w_- b_- \quad (11)$$

and the bound incoherent scattering length is

$$b_i = [\langle b^2 \rangle - \langle b \rangle^2]^{1/2} = \sqrt{(w_+ w_-)} (b_+ - b_-) \quad (12)$$

An important example is ^1H for which the spin dependent scattering lengths are $b_+ \approx 10.85$ fm and $b_- \approx -47.51$ fm. The result is $b_c \approx -3.74$ fm and $b_i \approx 25.27$ fm. The spin incoherent scattering can be observed using polarized neutrons, either by aligning the nuclear spins within the sample or by polarization analysis. Unless otherwise stated, it is assumed that both the incident neutron beam and the nuclear spins are unpolarized.

Each nuclide has a scattering length given by Eq. (9). The total scattering cross section σ_s is the sum of the coherent and incoherent scattering cross sections, σ_c and σ_i , with

$$\sigma_c = 4\pi |b_c|^2$$

and

$$\sigma_i = 4\pi |b_i|^2 \quad (13)$$

unless $I = 0$ in which case $b_i = 0$ and $\sigma_i = 0$. The total scattering cross section is given by

$$\sigma_s = 4\pi \langle |b|^2 \rangle \quad (14)$$

where the brackets $\langle \dots \rangle$ denote a statistical average over the neutron and nuclear spins. The absorption cross section is given by

$$\sigma_a = (4\pi/k) \langle b'' \rangle \quad (15)$$

where k is the magnitude of the incident neutron wave vector. The absorption cross section is determined by the imaginary part of the coherent scattering length. It is only when neutron and nucleus are both polarized that the imaginary part of the incoherent scattering length contributes to σ_a .

Each element may be composed of different isotopes j of natural abundance ϵ_j (such that $\sum_j \epsilon_j = 1$), each with its own coherent and incoherent scattering lengths, b_{cj} and b_{ij} (except that $b_{ij} = 0$ for $I = 0$). The coherent scattering length for the element is given by

$$b_c = \sum_j \epsilon_j b_{cj} \quad (16)$$

the total scattering cross section is

$$\sigma_s = \sum_j \epsilon_j \sigma_{sj} = 4\pi \sum_j \epsilon_j b_{cj}^2 \quad (17)$$

the bound coherent scattering cross section is

$$\sigma_c = 4\pi \left(\sum_j \epsilon_j b_{cj} \right)^2 = 4\pi |b_c|^2 \quad (18)$$

the bound incoherent scattering cross section is

$$\sigma_i = \sigma_s - \sigma_c = 4\pi \sum_j \epsilon_j b_{cj}^2 - 4\pi \left(\sum_j \epsilon_j b_{cj} \right)^2 = 4\pi |b_i|^2 \quad (19)$$

and the absorption cross section is

$$\sigma_a = \sum_j \epsilon_j \sigma_{aj} = (4\pi/k) \sum_j \epsilon_j b_{cj}'' \quad (20)$$

The incoherent scattering, which is independent of scattering angle φ is composed of contributions from the spin and isotope incoherence, where

$$\sigma_i(\text{spin}) = \sum_j \epsilon_j \sigma_{ij} = 4\pi \sum_j \epsilon_j |b_{ij}|^2 = 4\pi \sum_j \epsilon_j w_{+j} w_{-j} |b_{+j} - b_{-j}|^2 \quad (21)$$

and

$$\sigma_i(\text{isotope}) = 4\pi \sum_j \epsilon_j \epsilon_{j'} |b_{cj} - b_{cj'}|^2 \quad (22)$$

The cross sections for a molecule are determined from tables⁹ of scattering lengths and cross sections that give values of b_c , σ_p , and σ_a for each constituent element (where σ_i is the sum of the spin

and isotope contributions) as well as for each isotope. If f_p is the fraction of atoms p within the molecule, the average coherent scattering length per atom is

$$\langle b \rangle = \sum_p f_p b_{cp} \quad (23)$$

so that the coherent scattering cross section per atom is

$$\sigma_c = 4\pi \langle b \rangle^2 = 4\pi \left(\sum_p f_p b_{cp} \right)^2 \quad (24)$$

whereas the incoherent scattering cross section is the addition of the individual cross sections $\sigma_i = \sum_p f_p \sigma_{ip}$, so that the total scattering cross section per atom is

$$\sigma_s = \sigma_c + \sigma_i = 4\pi \left(\sum_p f_p b_{cp} \right)^2 + \sum_p f_p \sigma_{ip} \quad (25)$$

The absorption cross section per atom is the sum of the individual cross sections

$$\sigma_a = \sum_p f_p \sigma_{ap} \quad (26)$$

For a molecule of molecular mass M and macroscopic mass density ρ_M , the molecular number density is given by $N = N_A \rho_M / M$, where N_A is Avogadro's number, $\approx 0.6023 \times 10^{24}$ molecules/mol. The macroscopic cross sections therefore become

$$\sum_{\text{coh}} = N \sigma_c = 4\pi N \left(\sum_p f_p b_{cp} \right)^2$$

$$\sum_{\text{inc}} = N \sigma_i$$

and

$$\sum_{\text{abs}} = N \sigma_a \quad (27)$$

where the units of the macroscopic cross section are cm^{-1} , provided that the microscopic cross section σ is in units of 10^{-24} cm^2 (or b), and the scattering length b is in 10^{-12} cm .

Scattering Length Density

While the coherent scattering length is an important quantity determined by the neutron-nucleon strong force interaction, it is the scattering length density of the composition that determines the refraction, reflection, and diffraction properties of the material. If there are s atoms per molecule, the scattering length density of the molecule is given by $sN \sum_p f_p b_{cp} = \rho b$, where $\rho = sN = sN_A \rho_M / M$ is the atom number density, and b is the average scattering length per atom. The product ρb is called the scattering length density of a material. The scattered intensity in small-angle scattering is proportional to the square of the difference in the scattering length densities between two regions of different constituents. For example, this contrast ($\rho_A b_A - \rho_B b_B$) might be from inhomogeneities of molecule A (or pores for which $\rho_A b_A = 0$) in a background or solvent of molecules B.

TABLE 2 Neutron Scattering Lengths and Cross Sections for Hydrogen, Oxygen, and Water

	A^*	$I(\pi)^\dagger$	ε^*	$b_c^\ddagger$	$b_i^\ddagger$	$\sigma_c^\S$	$\sigma_i^\S$	$\sigma_s^\S$	$\sigma_a^{*\S}$
H				-3.7391		1.7568	80.26	82.02	0.3326
	1	1/2(+)	99.985	-3.741	25.274	1.7583	80.27	82.03	0.3326
	2	1(+)	0.015	6.671	4.04	5.592	2.05	7.64	0.00052
O				5.803		0.4232	0.0	4.232	0.00019
	16	0(+)	99.762	5.803	0	0.4232	0	4.232	0.0001
	17	5/2(+)	0.038	5.78	0.18	0.420	0.004	4.20	0.236
H ₂ O				-0.56		0.118	56.05	56.17	0.222
D ₂ O				6.38		5.11	0.02	5.13	0.0004

*A is the mass number and ε is the natural isotopic abundance in per cent.

$^\dagger I(\pi)$ is the spin and parity of the nuclear ground state.

‡ Values of scattering length b are in 10^{-13} cm or fm.

§ Values of cross sections σ are in 10^{-24} cm² or barns.

$^{*\S}\sigma_a$ is the absorption cross section at $\lambda = 1.798$ Å.

An important example is the contrast used in small-angle scattering that can be obtained using deuterated solvents for matching the scattering length density to one part of a given system in order to highlight another part. This is available on account of the greatly different scattering properties of hydrogen and deuterium. The values are given in Table 2. Consider light water, H₂O, with $d \approx 1$ g cm⁻³, and molecular mass 18. The number density of molecules $N \approx 0.03346 \times 10^{24}$ cm⁻³. The coherent scattering lengths are $b_H \approx -3.74$ fm and $b_O \approx 5.80$ fm, so that the scattering length density is $(2b_H + b_O)N = \rho b \approx -5.60 \times 10^{-7}$ Å⁻². The average coherent scattering length is $\langle b \rangle = (2/3b_H + 1/3b_O) \approx -0.56$ fm. The coherent scattering cross section for H₂O is $\sigma_c = 3 \times 4\pi \langle b \rangle^2 \approx 0.118$ b per molecule, and $\Sigma_{\text{coh}} \approx 0.0039$ cm⁻¹. This should be compared with the much larger incoherent scattering cross section of light water, 160.52 b/molecule (principally due to the spin incoherence of H, but also the isotope incoherence from the 0.015% of D), and $\Sigma_{\text{inc}} \approx 5.37$ cm⁻¹.

Similarly, consider heavy water, D₂O, which has the number density of molecules $N \approx 0.03329 \times 10^{24}$ cm⁻³. The coherent scattering lengths are $b_D \approx 6.67$ fm and $b_O \approx 5.80$ fm, so that the scattering length density is $(2b_D + b_O)N = \rho b \approx 6.37 \times 10^{-6}$ Å⁻². The average coherent scattering length is $\langle b \rangle = (2/3b_D + 1/3b_O) \approx 6.38$ fm. The average coherent scattering cross section for D₂O is $\sigma_c = 3 \times 4\pi \langle b \rangle^2 \approx 15.35$ b/molecule, and $\Sigma_{\text{coh}} \approx 0.511$ cm⁻¹. This compares with the incoherent scattering cross section of heavy water, 4.10 b/molecule, and $\Sigma_{\text{inc}} \approx 0.137$ cm⁻¹.

The scattering length density may be varied over a wide range using a mixture of deuterated and protonated solvents, and in particular light and heavy water. If x is the volume fraction of heavy water in the mixture, the scattering length density is given by $(1-x)(\rho b)_{\text{H}_2\text{O}} + x(\rho b)_{\text{D}_2\text{O}} = (6.96x - 0.56) \times 10^{-6}$ Å⁻¹. As the deuterated fraction increases, the scattering length density increases linearly, and at 8 percent deuteration the coherent scattering is at a minimum, after which it increases quadratically. On the other hand both the incoherent and the absorption cross sections decrease steadily upon increasing deuteration. The matching of the scattering length density (effectively, the index of refraction) of a material with the liquid in which it is immersed gives rise to a sharp minimum in the diffuse scattering.¹⁰ This technique allows the measurement of the refractive index of a wide range of materials by matching the value of the neutron scattering density, ρb , of the material with a mixture of H₂O and D₂O, which can cover a wide range as shown in Fig. 1. This is analogous to the Christiansen filter in classical optics. The refractive index matching has become an important method for determining the macromolecular structure of biological and polymeric systems using neutron scattering by varying the contrast between different parts of the structure and the supporting medium. This is now the standard method for small-angle neutron scattering from biological macromolecules.¹¹ This technique of refractive index matching is also available for nuclear and magnetic components of the refractive index by matching one of the two spin states to the nonmagnetic material in order to highlight the other.

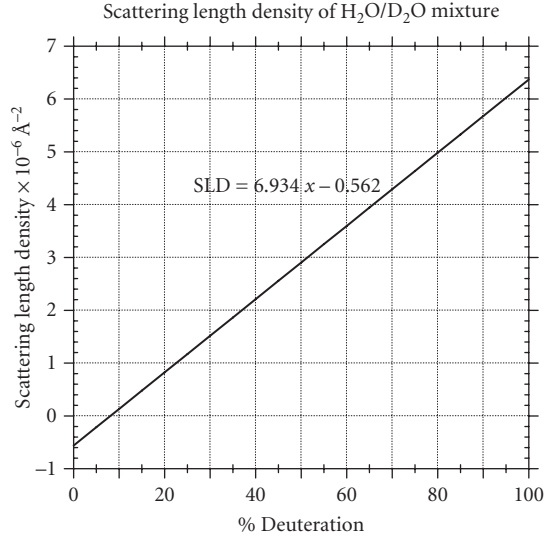


FIGURE 1 The scattering density of water as a function of the fraction x that is deuterated. The scattering lengths of H and D are very different (-0.37×10^{-12} cm and 0.66×10^{-12} cm). The scattering densities of all biological molecules lie between the limits of pure H₂O (-0.56×10^{-12} cm⁻²) and pure D₂O (6.3×10^{-12} cm⁻²) and can therefore be matched by some fraction of deuterated water.

In the case of a mixture of light and heavy water the various scattering properties depend on the fraction x of deuteration. There is H/D exchange between molecules so that there are fractions $(1-x^2)$, $2x(1-x)$, and x^2 of H₂O, HDO and D₂O, respectively. The scattering length density is given by $N[(1-x)b_{\text{H}_2\text{O}} + xb_{\text{D}_2\text{O}}]$. The coherent scattering cross section per molecule is given by $\sigma_{\text{coh}} = 4\pi \times 3[x\langle b_c \rangle_{\text{D}_2\text{O}} + (1-x)\langle b_c \rangle_{\text{H}_2\text{O}}]^2$.

The true incoherent scattering is given by $\sum_p f_p b_{p,\text{inc}}^2 = (1-x)b_{\text{H}_2\text{O},\text{inc}}^2 + xb_{\text{D}_2\text{O},\text{inc}}^2$. None of these quantities are changed by the H/D exchange. However, the compositional scattering is given by $\sum_{pp'} f_p f_{p'} (b_{cp} - b_{cp'})^2 = x(1-x)(b_{c\text{H}_2\text{O}} - b_{c\text{D}_2\text{O}})^2/2$. This is smaller by 50 percent of that if there were no H/D exchange.

Neutron Attenuation

The Lambert-Beer law of attenuation $T = \exp(-\mu\ell)$ relating the transmission of radiation through matter also applies to the attenuation of a neutron beam in a given medium of length ℓ . (This equation is exact for absorption and approximate for scattered neutrons, but converges to the exact value as the detector is moved further downstream from the sample.) The attenuation coefficient μ is usually written using the macroscopic cross section $\Sigma = \rho\sigma_T$, where $\sigma_T = \sigma_a + \sigma_s$ is the total cross section. The total transmission through an object that is composed of isotopes i , each with a number density ρ_i , thickness t_i , and total cross section σ_{Ti} , is given by $T = \exp[-\sum(\rho_i\sigma_{Ti}t_i)]$. Because the cross sections vary irregularly between elements in the periodic table, neutron radiography can give good contrast between neighboring elements, unlike x-ray radiography. Certain heavy elements also have

large resonances. Moreover, neutrons have an increased sensitivity relative to x rays for light elements, such as hydrogen, carbon, nitrogen, and oxygen. In particular, the anomalously large scattering cross section for hydrogen makes neutron radiography important for the location of hydrogenous regions within materials, and complementary to x-ray radiography for a number of applications. The technique has been important in studying low temperature hydrogen fuel cells.

Neutron radiography¹² is used in both industrial and fundamental research for nondestructive testing, with tomographic and image reconstruction methods. The spatial resolution of a neutron image is defined as the minimum spatial separation required for resolving two point-like objects that have been blurred by the imaging system. The spatial resolution of the image due to the geometry of the neutron beam arrangement is determined by the aperture diameter D and the aperture-detector separation distance L . A value of $L/D \approx 600$ is nearly optimal when using a standard radiography camera (resolution $\approx 250 \mu\text{m}$). Increasing L/D reduces the neutron intensity and requires longer acquisition times. For higher resolution neutron radiography detectors (resolution $\approx 25 \mu\text{m}$) much higher values of L/D (≥ 1200) are necessary.

63.3 NEUTRON SOURCES

There are several types of sources of thermal neutron beams that have adequate fluxes for useful measurements.^{13,14} (Though note that the resulting neutron beams are characterized by relatively low neutron current densities compared to beams on x-ray synchrotron sources.) In each case these sources produce fast neutrons that must be slowed down or moderated to thermal energies with velocities useful for neutron spectroscopy. The moderator of area $\approx 100 \text{ cm}^2$ becomes the effective source that produces a neutron density with a quasi-Maxwellian-Boltzmann velocity distribution approximately in thermal equilibrium with the moderator. The performance of a neutron source is usually characterized by its brilliance defined by $d^4\Phi/d\Omega d\lambda dA dt$ in units of neutrons $\text{ster}^{-1} \text{ \AA}^{-1} \text{ cm}^{-2} \text{ s}^{-1}$. A fully moderated thermal neutron source produces beams having a Maxwell-Boltzmann velocity distribution, such that the neutron beam has a wavelength spectrum that varies with the neutron wavelength as

$$I(\lambda) \propto \lambda^{-5} \exp[-(\lambda_T/\lambda)^2] \quad (28)$$

where λ_T is the wavelength of a neutron with an energy $k_B T$

$$\lambda_T = h(2mk_B T)^{-1/2} \quad (29)$$

and where k_B is the Boltzmann constant and T is the temperature of the moderator. Fig. 2 shows the fully moderated wavelength spectrum for $T = 293.6 \text{ K}$.

Thermal nuclear reactors produce neutrons using the fission reaction, and are generally steady-state sources, though there are also pulsed reactors. Each fission event releases a huge amount of energy (200 MeV) in the form of kinetic energy of the fission fragments, gamma, rays, and several fast neutrons. The fission fragments are the major source of heating within the reactor core, and the heat removal is the limitation to the accessible power at research reactors and therefore to the available neutron beam currents. Furthermore, the gammas and fast neutrons require careful shielding, and to produce a useful thermal neutron beam, the fast neutrons must be moderated by over five orders of magnitude from around 2 MeV to $\approx 25 \text{ meV}$ by some hydrogenous material (usually light or heavy water).

Pulsed sources generally require a particle accelerator (though they may operate in a quasi-continuous mode). An electron linac produces bremsstrahlung photons when the electrons are stopped in a heavy metal target, and these excite resonance interactions to produce evaporation neutrons. (This can also be achieved from the (γ, n) reaction in light nuclei such D or Be.) Alternatively,

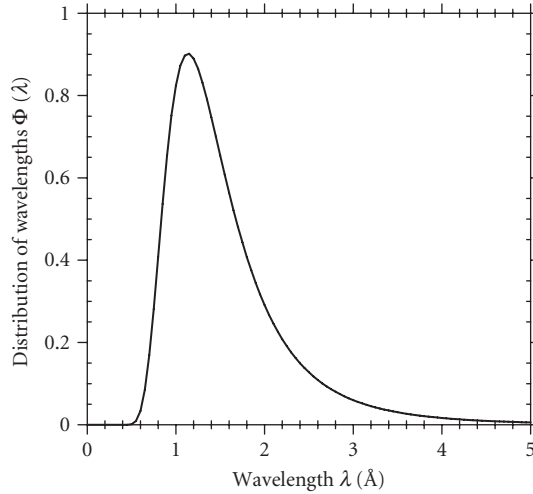


FIGURE 2 The normalized neutron flux distribution of a completely moderated Maxwellian spectrum, $\Phi(\lambda) = 2\Phi_0 (\lambda_T^4 / \lambda^5) \exp[-\lambda_T^2 / \lambda^2]$, corresponding to a temperature T of 293.6 K, or $\lambda_T = 1.798$ Å, or energy 25.3 MeV, or velocity 2200 m s^{-1} . The mean wavelength is $\sqrt{(\pi/2)} \lambda_T = 2.25$ Å, and the most probable wavelength is $\sqrt{(2/5)} \lambda_T = 1.137$ Å (corresponding to an energy of 63.2 MeV). Note that there are always many unwanted fast and epithermal neutrons, whatever the source, that need to be removed from the beam.

beams of high energy ($\approx 100 \text{ MeV}$) H^- ions produced in a linear accelerator are stripped of their electrons as they are injected into the proton synchrotron ring. After further acceleration to much higher energies ($\approx 800 \text{ MeV}$) they smash into the heavy metal target to produce many spallation products including high-energy neutrons. As before, these neutrons must be moderated to thermal energies by some hydrogenous material to provide neutron beams suitable for experimental work.

Whatever source is used, an important consideration is the background and the acceptable signal-to-noise for any neutron measurement on account of high energy gamma rays and fast neutrons ($E > 1 \text{ keV}$) produced by the source. While thermal neutrons can be stopped by materials with a high absorption cross section (such as cadmium, gadolinium, boron, and lithium), the shielding from fast neutrons requires that they first be slowed down to thermal energies before the capture cross section in one of those materials becomes appreciable. A suitable combination of various materials is required; heavy metals such as iron can enable neutron inelastic collisions, hydrogenous materials such as concrete can moderate the neutron spectrum, and finally absorbing material such as boron can capture the moderated neutrons. Heavy concrete having all three items is often used, but detailed analysis of the shielding¹⁵ requires careful computer simulations for optimum results.

While shielding from the direct gammas from the source (either the reactor core or the spallation target) is often provided by a combination of lead, steel, and concrete, the experimental arrangement must take into account the capture gammas resulting from thermal neutrons absorbed in materials to produce both prompt and long-term activation. Both cadmium and gadolinium emit high-energy gammas upon neutron capture, and so must be used judiciously. Boron also produces a 0.478 MeV gamma, so that lithium is the preferred shielding material in the measurement area. On the other hand gadolinium makes excellent beam definition as a knife-edge aperture, provided other

material shields the capture gammas. Other aperture materials need to be thicker edged and can cause extra diffuse scatter. (Note that both boron and lithium, together with helium, are also neutron detector materials on account of their high absorption cross sections.)

Neutrons produced in a reactor come into thermal equilibrium with the moderator, and have an energy spectrum with a Maxwellian distribution characteristic of the temperature of the moderator. (In the case of a pulsed source thermal equilibrium is only approximate because time-of-flight measurements require a thin moderator to obtain adequate time resolution.) As indicated in Eq. (29), the lower the temperature of the moderator the lower the average neutron energy and the longer the available wavelengths. Consequently, it is best to have a cold moderator (often hydrogen or a mixture of hydrogen and deuterium) to produce large amounts of the long-wavelength neutrons for high-resolution measurements. Indeed, the greater availability of cold neutrons has increased the number of beam lines and experiments that involve neutron optics.

Neutron beams produced at both sources must be transported from the moderator to the experimental hall, either through beam tubes, or better, through neutron guides using total reflection. This allows both thermal and subthermal beams to be brought with little loss to a region of much lower source background. Straight guides look directly at the source, so that either curved guides or neutron filters must be used to remove these source gamma rays and fast neutrons from the beam. The various filters that are used are either Bragg cut-off filters (such as the liquid nitrogen-cooled polycrystalline beryllium filter), or single-crystal filters (such as the sapphire that can be used at room temperature, or magnesium fluoride or bismuth at cryogenic temperatures).

The determination of which source might be more appropriate is dependent on the particular measurement required.¹⁶ Though for some experiments the wide energy spectrum of the polychromatic beam after filtration is sometimes used, generally a monochromatic beam is used for neutron spectroscopy, whereas high-resolution powder diffraction is best performed using the entire spectrum on a pulsed source. However, the most appropriate beam to use depends on the particular measurement envisaged, though most neutron optics research has been performed on cold sources at reactors.

The steady state reactor source can use either a crystal monochromator or a velocity selector to determine the wavelength. The resolution of the measurement can be tailored to the particular problem by suitable choices of collimation and neutron wavelength, and is symmetric and approximately Gaussian. The monochromator crystal selects a narrow wavelength band typically in the range $\Delta\lambda/\lambda \approx 10^{-2}$ from the white beam according to Bragg's law. The velocity selector is usually composed of helical channels formed by absorbing blades that rotate about an axis parallel to the neutron beam. The mean neutron velocity is determined by the rotation speed, and tilting the rotation axis by a small angle enables a change in the coarse wavelength resolution $\Delta\lambda/\lambda \approx 0.1$ to 0.2. Disc choppers composed of absorbing material with slits rotating in phase at high speed on an axis parallel to the beam and placed a distance apart can also monochromate the beam. The rotation speed and the slit widths determine the wavelength resolution, which may be increased by using counter-rotating choppers. Finally a Fermi chopper with narrow curved slits that rotates about an axis perpendicular to the beam can also monochromate the beam.

The pulsed source measurement requires special considerations because it is by necessity performed in time-of-flight mode synchronized to the source frequency. The instrument collects the data by measuring the time interval t taken for the neutron pulse to travel a distance L from the moderator to its detection point. This allows a determination of the neutron energy or wavelength by

$$\lambda = (h/m)(t/L) \quad (30)$$

Since the relative uncertainty in the length of the flight path is small, the time uncertainty $\Delta t/t$ dominates the resolution of the pulsed source instrument. Hence the resolution of the measurement depends on the pulse width Δt produced at the thin (≈ 50 mm) moderator, and is asymmetric and wavelength dependent. Consequently, the neutron moderation time must be kept to a minimum, while maximizing the beam intensities at long wavelengths emanating from the moderator. Heterogeneous poisoning through a layer of neutron absorbing material within the moderator keeps the moderation

properties while reducing the pulse width at low energies. Cooling the moderator to low temperatures shifts the spectrum to long wavelengths while keeping the pulse width narrow. Finally, grooved or fluted moderators can enable greater beam currents by increasing the effective volume for moderation and yet decreasing the effective width for pulse emission at long wavelengths.

63.4 NEUTRON OPTICAL DEVICES

Neutrons stream from the source to the sample through evacuated tubes that might include a number of optical devices. A simple example is the neutron guide that uses grazing incidence reflective optics to transport the neutron beam. Because neutron sources are very weak compared with x-ray synchrotron sources, efforts are made to increase the neutron current density at the sample by various focusing methods. For example, the simple pinhole collimation used in small-angle scattering can be enhanced by multiple confocal pinhole collimations that converge at a point on the two-dimensional area detector. Though the overall current on the sample has increased, the resolution of the measurement is the same as for pinhole collimation. Focusing can be achieved using various optical devices. These elements use the optical characteristics of thermal and cold neutrons, such as diffractive optics (monochromators to select specific wavelengths), reflective optics using grazing incidence (guides transport the beam, mirrors deflect the beam, and various reflection devices can focus the beam), or refractive optics (prisms also refract the beam and concave lenses focus the beam). Many of these optical devices can also be adapted specifically for polarized neutrons. In addition, filters use neutron absorption to limit the range of transmitted wavelengths. However, note that the results are limited by Liouville's theorem that shows that the phase space density of neutrons cannot be increased using energy conserving methods; indeed, it can only decrease, by the use of imperfect devices that absorb, scatter, or reflect fractions of the neutron beam. Consequently, there are limitations on the ability to focus the neutron beam onto the sample, because any increase in neutron current density is attained at the expense of the beam divergence at the sample. If, however, increased beam divergence can be tolerated, even in one dimension, this can be useful. There are other modes of focusing available. Focusing of neutrons in momentum space becomes important for diffraction and scattering applications. There is also time focusing for improving the resolution for time-of-flight measurements¹³ on instruments when the source, sample, analyzer (if any), and detector have extended areas. In this case the orientation of the various components is arranged so that neutrons with different velocities arrive simultaneously at the detector, having had the same momentum change, despite the broad range of velocities within the incident beam.

Neutron Collimation

Neutrons may be collimated by defining the beam path with pinhole or slit apertures in a material of high absorption cross section, such as cadmium or gadolinium, that is essentially black to thermal neutrons. The aperture sizes and the distance between them define the beam direction and the divergence. The transmission may be calculated using phase space acceptance diagrams indicating the position and angle in each of the two orthogonal transverse directions. The maximum angle transmitted relative to the beam centerline is given by the collimation angle. Often collimation is required in only one dimension (that of the scattering plane of the measurement). Soller collimators, composed of a number of long thin blades of neutron absorbing material spaced equally apart, allow large transmission, but with narrow divergence in the one plane. These are often used to define the wavelength resolution of single-crystal monochromator instruments. Mini collimators composed of thin single-crystal wafers coated with gadolinium are also used for the purpose. (Oscillating radial collimators with radial blades have also been used in conjunction with position-sensitive detectors to reduce the background from the sample environment.) The transmission function of an ideal collimator is triangular in shape. The transmission may be increased by the use of

highly polished reflecting surfaces, such that transmission is uniform in position and angle up to θ_c , the critical angle of the surface material, provided the device, now a guide, is sufficiently long. Furthermore, neutron benders have been constructed from stretched mylar blades, coated with copper and separated by spacers.

Neutron Guides

The principle of total external reflection from smooth surfaces at small grazing angles enables neutron guide tubes¹⁷ to be used as channels to transport high-intensity, thermal, or cold neutron beams over relatively long distances (≈ 100 m) with only small losses in intensity. They can supply beams for multiple instruments for neutron scattering experiments at locations of low background with significant improvement in signal-to-noise. Gamma rays and fast neutrons emitted from the source decrease with the square of the distance from the source, whereas the guide maintains the same beam current, neglecting reflectivity losses. These neutron guides are common at research reactors, particularly those with cold sources because their efficiency increases with wavelength. Neutron guides are analogous to fiber optics or light pipes in ordinary optics. A neutron entering the guide tube with an angle of incidence at the wall that is less than the critical angle $\theta_c = \lambda \sqrt{(\rho b)/\pi} = \lambda / \lambda_c$ for its particular wavelength is transported along the hollow tube by multiple total reflections.

The guides are usually made of highly polished boron-glass plates of rectangular cross section that are plated with nickel, which has the highest critical angle per unit wavelength, $\theta_c/\lambda \approx 17.3$ mrad nm⁻¹ or about 1° nm⁻¹, of all elements. Nickel ($b \approx 10.3$ fm) has a scattering length density $\rho b \approx 9.37 \times 10^{-6}$ Å⁻² with $\lambda_c = (\pi/\rho b)^{1/2} \approx 58$ nm and is the reference to which other guides are compared. Isotopically pure ⁵⁸Ni ($b \approx 14.4$ fm) has a greater scattering length density of 1.31×10^{-5} Å⁻² and $\lambda_c \approx 49$ nm, and has a critical angle per unit wavelength, $\theta_c/\lambda \approx 20.4$ mrad nm⁻¹. The use of ⁵⁸Ni therefore results in a 40 percent increase in transmission. Further increases in transmission may be obtained by using glass coated with supermirror layers on top of nickel. The superposition of many diffraction peaks from alternating layers of thin films having a large contrast in scattering length density acts as a two-dimensional crystal to extend the total reflection angle beyond that of nickel. The best contrast is obtained with nickel-titanium layers because Ti is one of the few elements that have a negative scattering length ($b \approx -3.44$ fm). The contrast is given by $\rho b = (\rho b)_{\text{Ni}} - (\rho b)_{\text{Ti}} \approx (9.37 + 1.95) \times 10^{-6} = 1.13 \times 10^{-5}$ Å⁻². The multilayer consists of a large number of thin layers with a range of d-spacings up to $d = \sqrt{(\pi/4\rho b)}$ such that there is a continuous diffraction using Bragg's law $\theta = \lambda/2d$ at small angles. The resulting supermirror is characterized by the ratio of the effective critical scattering vector $q_c (= 4\pi\theta_c/\lambda)$ to that of natural nickel. Large area coatings with a ratio of 3.6 and with adequate reflectivity are routinely available for use in optical devices and higher values are under development.¹⁸ In practice, there is some loss in reflectivity beyond the nickel total reflection angle, and it drops off considerably at highest values of q_c . This is caused by a number of factors: the roughness of the substrate surface, the interdiffusion between layers, and the stresses induced within the material. Sears¹⁹ has reviewed the theory of multilayer neutron monochromators, and Anderson²⁰ has reviewed the development of supermirrors.

Guides are composed of many sections that might typically be 750 mm long, with cross sections 200 mm high and 60 mm wide. Neglecting any reflection losses, the solid angle of the beam emerging from the exit of the guide is $4\theta_c^2$, providing that the guide entrance is uniformly illuminated in space and angle. The effective collimation achieved by these rectangular neutron guides corresponds to $2\theta_c$ in each orthogonal direction, and therefore depends on wavelength. Straight guides have the maximum transmission (proportional to the square of the wavelength) and require filters to obtain good transmission characteristics without the unwanted radiation. However, many guides are curved to take the beam away from the direct streaming of fast neutrons and gammas, and have good transmission for slow neutrons above a characteristic wavelength. Fast neutrons and gamma rays pass through the glass wall and are eliminated from the beam, so that guides can transport highly collimated beams of thermal neutrons, devoid of gammas and fast neutrons, to regions of low-background far from the source. If the guide sections are

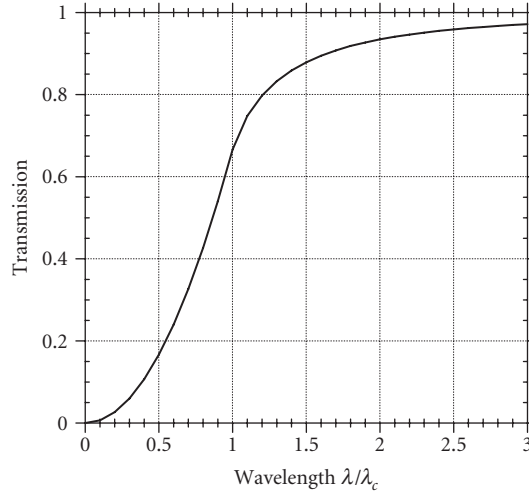


FIGURE 3 The transmission efficiency through a curved guide or neutron bender of length at least $\sqrt{(8wR)}$ as a function of wavelength λ , relative to the straight guide, assuming that the reflectivity of the surfaces of guide coating is 100 percent. For $\lambda < \lambda_c$, $T = (2/3)(\lambda/\lambda_c)^2$, and for $\lambda > \lambda_c$, $T = (2/3)(\lambda/\lambda_c)^2 \{1 - [1 - (\lambda/\lambda_c)^{-2}]^{3/2}\}$.

placed along a polygonal approximation to a curved guide with a gentle curvature, the transmission is defined by a characteristic wavelength $\lambda_c = (\rho b/\pi)^{-1/2} \psi_c$, where the characteristic angle of the curved guide $\psi_c = (2w/R)^{1/2}$, with w and R being the width and the radius of curvature of the guide. For wavelengths less than λ_c , transmission only occurs by garland reflections along the outer (concave) wall. Fig. 3 shows that at λ_c , the transmission is only 2/3 that for the straight guide. For wavelengths greater than λ_c , zig-zag trajectories are also allowed and the transmission increases asymptotically toward that of the straight guide. Consequently, the curved guide acts as a filter provided that the length of the guide is greater than $\sqrt{(8wR)}$ to avoid the direct line-of-sight with the source.

Guides are also found on pulsed source instruments. They maintain the incident beam current density over long distances to enable more instruments to surround the source. However, the pulse frequency, f , limits the wavelength range to $\Delta\lambda = (h/m)(1/fL)$ available on an instrument when the detectors are at a distance L from the source, on account of the frame overlap of the slowest neutrons by the fastest neutrons from the subsequent pulse. Disc choppers and filters often provide the means to overcome this wavelength limitation on a beam line.

The current density of a neutron beam at the sample may be increased by the use of a converging guide. The gain of the converging guide depends on three factors: (1) the convergence of the guide, (2) the ratio of the critical angles on the converging and straight guides, and (3) the neutron wavelength. The greatest gains are achieved when the critical angle of the converging guide is much greater than the divergence of the incoming beam defined by the critical angle of the straight guide. Even with supermirror coatings, intensity gains for the two-dimensional converging guide are limited to perhaps an order of magnitude on account of the increase in absolute angle that a neutron trajectory makes with the device axis upon successive reflections.²¹

More recently, the use of "ballistic" guides reduces losses from nonperfect reflection. Neutrons enter a guide that initially diverges and then converges to its exit. The central part of the guide has a coating of lower q_c (and therefore higher reflectivity) than the entrance and exit of the guide. Neutron trajectories have a smaller number of reflections with higher reflectivity, resulting in a greater neutron

current density than the equivalent straight guide. The details of the transmission of these ballistic guides as a function of wavelength require computer simulation.²²

Neutron Filters

The use of long wavelength neutrons at reactor and pulsed sources often requires a filter to remove epithermal neutrons (energies greater than ≈ 1 eV), and large perfect single crystals can produce a beam relatively free of fast neutron background. They are also used to reduce the higher-order reflections from crystal monochromators. A useful neutron filter material²³ must have wavelength dependent cross sections such that the total cross section is low at thermal energies of interest, but large at epithermal and higher energies. The efficiency of the filter depends on the magnitude of the various cross sections. The absorption cross section is usually linearly dependent on the neutron wavelength and is always independent of temperature. The coherent Bragg scattering cross section depends on the neutron wavelength and the crystal temperature, orientation, and perfection, and can be reduced by suitable orientations and by using highly perfect crystals. The incoherent elastic scattering cross section is usually small and is independent of the wavelength. The inelastic or phonon scattering cross section also varies with wavelength and depends on the crystal temperature T . At low temperatures it varies as $T^{7/2}$ and becomes linear with T at high temperatures. At energies well below $k\Theta_D$, where Θ_D is the Debye temperature of the crystal, the single phonon cross section is also linear dependent on the wavelength. At higher energies ($E > k\Theta_D$), the multiphonon cross section increases with energy and temperature, and rises to the free atom cross section at much higher energies.

Typical single-crystal filters include silicon, quartz (SiO_2), sapphire (Al_2O_3), magnesium oxide, and magnesium fluoride (Table 3). These filters show a minimum in the total cross section around (0.2 to 0.3 nm), with a small linear increase at longer wavelength, and a sharp increase at shorter wavelengths. These materials all have low absorption and incoherent scattering cross sections. MgO and MgF_2 have relatively low Debye temperatures and the cross sections in the thermal region are reduced considerably at cryogenic temperatures. On the other hand there is little to be gained by lowering the temperature of Al_2O_3 from room temperature on account of its high Debye temperature ($\Theta_D = 1032$ K). Polycrystalline beryllium cooled to 77 K is frequently used as a Bragg cutoff filter for neutrons with wavelengths less than 0.4 nm (energies above 5 MeV), since no Bragg scattering is possible for sufficiently long wavelengths such that $\lambda \geq 2d_{\text{max}}$, where d_{max} is the largest plane spacing. Pyrolytic graphite crystal oriented with the c axis along the incident beam direction allows certain energy windows to exist for which there is no Bragg reflections available to scatter the neutrons. Finally, resonance filters that have high absorption resonance cross sections can eliminate specific wavelengths from the beam to remove $\lambda/2$ contamination from the scattered beam.

TABLE 3 Possible Filter Materials

	M^*	ρ_M^*	$N^\dagger$	$\sum_i b_i^{\ddagger}$	$\rho b^{\S}$	$\Sigma_{fa}^{\P}$
SiO_2	60.08	2.65	0.0266	1.576	4.19	0.253
Al_2O_3	101.96	3.98	0.0234	2.431	5.71	0.328
MgO	40.31	3.58	0.0535	1.118	5.98	0.384
MgF_2	62.32	3.18	0.0307	1.668	5.13	0.334

* M is the molecular mass in g/mol and ρ_M is density in g cm^{-3} .

$^\dagger N$ is the molecule number density in 10^{24} cm^{-3} .

$^\ddagger \sum_i b_i$ is the molecular coherent scattering length in 10^{-12} cm .

$^\S \rho b = N \sum_i b_i$ is the scattering length density in $10^{-6} \text{ \AA}^{-2}$.

$^\P \Sigma_{fa}$ is the free atom macroscopic cross section in cm^{-1} at high energies.

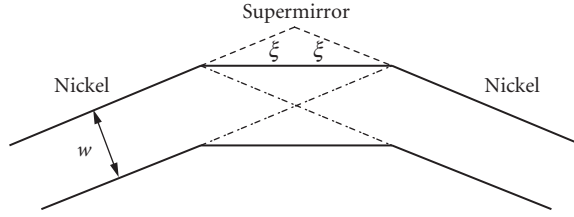


FIGURE 4 A schematic drawing of a neutron optical filter in which the initial long guide of nickel, of width w , is followed at an angle, ξ , by a short length of supermirror guide, and then another nickel guide at a further angle, ξ . If the angle ξ is made equal to the critical θ_c of nickel at the shortest wavelength of interest, then the supermirror must have a critical angle of at least $2\theta_c$.

A different concept is the neutron optical filter,²⁴ in which the beam may be deflected by a small angle with a high transmission above a cutoff wavelength, without the spatial asymmetry that is introduced by the curved guide. This is illustrated in Fig. 4. An initial long nickel guide is followed by a short length of supermirror guide, offset by a small angle, ξ , after which the nickel guide continues at a further small angle, ξ . The length of the intermediate supermirror guide is such that there is no direct line of sight through the entire system. The advantage of such a geometry is that it has the property that a parallel beam is transported unchanged, unlike the case of a curved guide. The supermirror critical angle must be at least twice that of nickel. A modification is to have the central section tapered and with a different critical angle. Consequently, this type of filter can be designed for particular experimental arrangements.

63.5 REFRACTION AND REFLECTION

All isotopes have a scattering length b that characterizes the neutron-nucleus interaction (albeit some have a complex value).⁷ Consequently, all materials have an index of refraction that depends on the scattering length of the isotopes that compose the material. The basic difference is that, as for x rays, the index of refraction for neutrons has a value of $n < 1$, whereas $n > 1$ for light. Moreover, the neutron wavelengths are about three orders of magnitude smaller than for visible light. This means that in general much larger instruments are required to observe the optical phenomena with the same spatial resolution. In addition, the most intense thermal neutron sources are orders of magnitude weaker than conventional light and x-ray sources. This means that the phenomena are generally not as easy to observe, nor is the resolution as good. However, successful neutron optical devices have been developed and are now commonplace in instrumentation for scientific research studies, particularly those that use cold- or long-wavelength neutrons.

The index of refraction of a medium is defined by $n = k'/k = \lambda/\lambda'$, and for slow neutrons is given by

$$n^2 = 1 - \frac{\lambda^2 \rho b}{\pi} \quad (31)$$

where $\lambda = 2\pi/|k|$ is the incident neutron wavelength, and ρ is the average atom density of the material. Since $V/E \ll 1$, Eq. (31) may be approximated by $n = 1 - \lambda^2 \rho b / 2\pi$. For thermal neutrons with wavelength $\lambda > 0.1$ nm, the refractive index for most materials differs from unity by a few parts in 10^6 . Consequently, the various optical features are similar to those in x-ray optics, and various analogous features have been demonstrated. However, the small deviation of the refractive index from unity means

that focusing through either refraction or reflection is weak. Note that at the boundary between two media, A and B, the relative index of refraction is given by

$$n^2 = 1 - (\lambda^2/\pi)(\rho_A b_A - \rho_B b_B) \quad (32)$$

where $\rho_A b_A$ is the scattering length density of medium A.

A critical wavelength is defined by $\lambda_c = (\pi/\rho b)^{1/2}$. If b is real and positive, the index of refraction is real if $\lambda < \lambda_c$ and imaginary if $\lambda > \lambda_c$. The phenomenon of total reflection for all angles of incidence occurs when n is purely imaginary, that is when $\lambda > \lambda_c$. Typically for materials $\lambda_c \approx 140$ nm, which corresponds to a neutron energy of 4×10^{-8} eV or a temperature of 0.5 mK. It is therefore only for ultracold neutrons²⁵ that true mirror reflection can occur. Mirror reflections for ultracold neutrons can be used to confine neutrons in material cavities or bottles. Smooth, clean surfaces are needed for good storage, though this is limited by β decay of the free neutron (with a half-life of ≈ 887 s). In practice, the storage time is shorter because of losses caused by inelastic scattering from surface impurities. Ultracold neutrons are useful for precision measurements of the free neutron lifetime, the neutron magnetic dipole moment and the search for its electrical dipole moment.

Note that the usual form of the refractive index is given by Eq. (31). This neglects local field effects (the nucleus scatters not only the incident wave but also waves scattered by all other nuclei). It also neglects the attenuation of the amplitude due to diffuse scattering of the coherent wave in the medium. Sears¹ has given a more rigorous and comprehensive treatment of dispersion theory. Fluctuations in the scattering length density may be due to both number density fluctuations as well as to the effects of spin and isotope disorder that produce fluctuations in the scattering length. Coherent scattering gives rise to the neutron optical effects that include reflection, refraction, diffraction, and so on, and depend on the orientation of the system relative to the incident beam, whereas diffuse incoherent scattering is distributed isotropically.

Refraction and Mirror Reflection

The total external reflection of neutrons from material surfaces has been demonstrated, analogous to the total external reflection of x rays and the total internal reflection of light. Because the index of refraction is close to unity, total reflection is possible only at glancing incident angles, θ , on a material surface when $\cos \theta > n$. This is the neutron analogue to Snell's law for light (the continuity of the tangential component of the wavevector). The measurement of the critical angle for various surfaces using monochromatic neutrons can be used for the determination of scattering lengths with a relative uncertainty of ≈ 0.2 percent. The reflectivity, R , is given by Fresnel's law,

$$R = \left| 1 - [1 - (\theta_c/\theta)^2]^{1/2} \right| / \left| 1 + [1 - (\theta_c/\theta)^2]^{1/2} \right| \quad (33)$$

where $\theta_c = \lambda/\lambda_c = \lambda(\rho b/\pi)^{1/2}$, for nonmagnetic materials. If $b > 0$, $R = 1$ for $\theta < \theta_c$, the critical angle for total reflection. Because $(1 - n^2) \approx 10^{-5}$ for thermal neutrons, $\theta_c \approx 3$ mrad $\approx 0.2^\circ$. Hence, the measurement of the reflectivity can be used to determine both the magnitude and sign of the bound coherent scattering length.

The mean optical potential of neutrons in most materials is comparable in magnitude with the gravitational potential energy corresponding to a height difference of order of 1 m. This is the principle of the gravity refractometer,²⁶ in which neutrons from a well-collimated initial horizontal beam fall under gravity in a long flight path onto the horizontal surface of a liquid. All neutrons achieve a critical vertical velocity at a height $h_0 = (2\pi\hbar^2/m^2g)\rho b$ and penetrate the surface of the liquid; otherwise, they will be totally reflected by the mirror and be detected in the counter. Consequently, h_0 is a measure of the quantity ρb , and hence of the scattering length, b . This experiment, illustrated in Fig. 5 demonstrates the unique particle-wave properties of neutron optics, with the motion obeying classical physics and the refraction obeying quantum physics. The significant point is that the

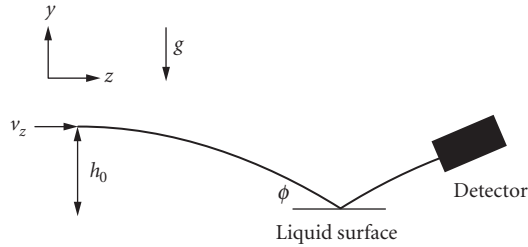


FIGURE 5 A schematic diagram of the neutron gravity spectrometer. A neutron of initial horizontal velocity v_z and vertical velocity $v_y = 0$ falls a vertical distance h and strikes the liquid surface at a grazing angle of $\phi = \sqrt{(2gh)}/v_z$ after traveling a horizontal distance $v_z\sqrt{(2h/g)}$. The critical height for total reflection is $h_0 = (2\pi\hbar^2/m^2g)\rho b$, independent of v_x or λ .

measurement is independent of the neutron wavelength. This allows the use of a more intense neutron beam to achieve a measurement with high statistical accuracy, resulting a most accurate method to determine b for a liquid.

Grazing-Angle Reflection

Focusing systems based on reflective optics do not suffer from the chromatic aberrations of refractive optics. For example, the advantage of an ellipsoidal mirror is that, neglecting gravity, all the trajectories emanating from one focus (the source) reach the other focus regardless of the neutron wavelength. However, single bounce mirror systems suffer from high aberrations. The difficulty with grazing incidence optics is that a near parallel beam incident is required and so that the optical element must be placed far enough from the source to ensure small incident beam divergence. Kirkpatrick-Baez neutron mirrors²⁷ using two successive reflections in orthogonal directions can efficiently focus neutron beams into small areas with a maximum divergence that is limited by the mirror critical angle. The size of the focal spot is primarily determined by geometrical demagnification of the source and by figure errors in the mirror shape. Approximately two orders of magnitude in neutron current density increase can be achieved within a spot of diameter $\approx 100\text{ }\mu\text{m}$ using crossed mirrors.

Grazing angle reflective optics based on two-bounce Wolter geometries are used extensively in x-ray astronomy because they minimize optical aberrations for off-axis trajectories (see Chap. 64). They can be also designed to focus cold and thermal neutron beams using consecutive reflections from parabolic and hyperbolic surfaces. The tilt angle of the hyperbolic section is 3 times larger than the tilt angle of the parabolic section to preserve the grazing angle of the trajectory. The mirrors are fabricated using an electroformed nickel replication process. They have a cylindrical form with different diameters but with the same focal length, such that they can be nested to increase the system throughput. Nested replicated optics can improve the focused beam intensity by increasing the incident area of the quasi-parallel beam.

Polycapillary Optics

Polycapillary optics (see Chap. 53) provides a better method for increasing the neutron current density at a point by using many narrow guides. This allows the possibility of a greater curvature than for macroguides for a given wavelength because the transmission characteristics of a curved guide

depend on the ratio of the transverse to longitudinal dimensions of the guide. These miniature versions of neutron guides can be used to transport, bend, and even focus thermal neutron beams. Glass polycapillary fibers with thousands of hollow channels recently developed for transporting and focusing x rays can also be used for cold (≈ 0.4 nm) neutron beams. The advantage of the narrow (≈ 10 μm) channels is that the neutron beam can be bent more sharply than for the wide (≈ 50 μm) guides. The transmission properties of the capillaries depend on the internal diameter and bending radius of the channels, the glass composition, and the surface smoothness.

The focusing of neutrons comes at a price. Liouville's theorem requires that an increase in neutron current density is obtained with a necessary increase in beam divergence. This means that real-space focusing has limited applications in condensed matter research using neutron scattering for which angular divergence is important, whereas neutron absorption techniques in analytical and materials research do not require a monochromatic neutron beam or a small angular divergence. Consequently, a high-intensity beam focused by a neutron capillary lens onto a small sample area can be useful for improving both the detection limits of individual elements and the spatial resolution of the measurement. Polycapillary fiber lenses have been produced and tested for analytical applications.²⁸ The gain depends on the divergence and the energy spectrum of the incident beam. Within the typical focal spot size of 0.5 mm, limited by the outer diameter of the fibers, the gain can be nearly two orders of magnitude. For the monolithic lens, a fused version of the polycapillary lens, the focal spot size is less than 200 μm with a gain of about 10.

Prisms and Lenses

The prism deflection of neutrons has been observed.²⁹ Because the refractive power, $(n - 1)$, is negative for most materials, prism deflection is in the opposite direction compared with ordinary light optics. Deflection angles of a few seconds of arc can be measured with high accuracy, allowing the measurements of refractive indices to 1 part in 10^4 . Because $(n - 1)$ is proportional to λ^2 , prism deflection of neutrons is highly dispersive. It has been used in a fused-quartz converging lens (which is concave in shape) and the focusing of 20 nm neutrons has been demonstrated.³⁰ More recently,³¹ the focusing of shorter (9 to 20) nm wavelength neutrons has been demonstrated using a series of many closely spaced symmetric biconcave MgF_2 lenses originally produced for infrared optics, with gains of approximately 15 and focal lengths of (1 to 6) m, which is well matched to small-angle neutron scattering. Magnesium fluoride is preferred to quartz because it has a higher scattering length density ($(\rho b)_{\text{MgF}_2} = 5.13 \times 10^{-6} \text{ \AA}^{-2}$, $(\rho b)_{\text{SiO}_2} = 4.19 \times 10^{-6} \text{ \AA}^{-2}$). This compound refractive lens is analogous to that for x rays (see Chap. 37).

Focusing now plays an increasing role in small-angle neutron scattering instruments in order to attain lower values of scattering vector without reducing the current density on the sample, by decreasing the extent of the beam penumbra at the detector. Pinhole collimation using the optimum equal-flight-path configuration, with the source aperture size twice that of the sample, results in a beam profile that is conical at the detector. The intensity on sample is proportional to the area of both apertures. A focusing optic consisting of a series of biconcave lenses enables the direct beam profile at the detector to be both narrower and uniform. When the detector is placed at the image position of the source with respect to the optic the beam profile at the detector is uniform, and its size is independent of the size of the sample aperture.

The focal length for a set of N biconcave lenses of radius of curvature $|r|$ is given by

$$f = |r|/2N(1 - n) \quad (34)$$

where the index of refraction is given by $n = 1 - \rho b \lambda^2 / 2\pi$. Hence the focal length of the biconcave lens is $f = \pi |r| / (N \rho b \lambda^2)$. Since f varies inversely with the square of the wavelength, refractive lenses are strongly chromatic, with large chromatic aberrations. For perfect focusing with the lens placed immediately before the sample, the usual lens equation is valid with $1/L_1 + 1/L_2 = 1/f$, where L_1 and L_2 are the incident (source to sample) and secondary (sample to detector) flight paths, respectively.

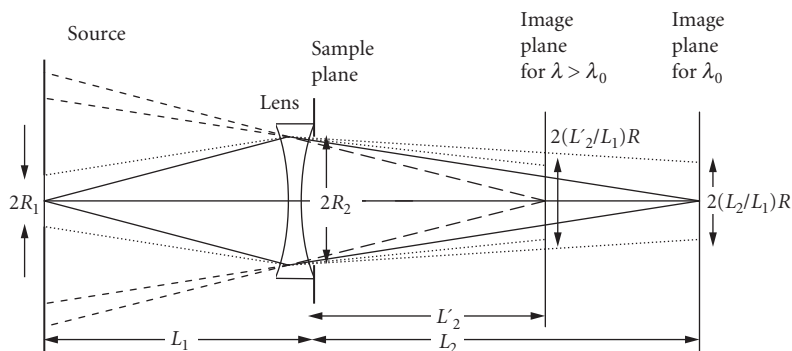


FIGURE 6 A schematic diagram of a focusing lens arrangement with source of radius R_1 at a distance L_1 from the sample with an aperture of radius R_2 . The focusing lens with a focal length f_0 is placed in front of the sample such that the source is imaged (continuous lines) with a radius $(L_2/L_1) R_1$ at a distance L_2 from the lens for a wavelength λ_0 . For another wavelength $\lambda (> \lambda_0)$ the source is imaged (long dashed lines) with a radius $(L'_2/L_1) R_1$ at a distance $L'_2 (< L_2)$ such that $1/L'_2 - 1/L_2 = 1/f_0[(\lambda/\lambda_0)^2 - 1]$. (See also color insert.)

However, the chromatic aberration determines that this is only true at one wavelength. If f_0 is the focal length of the lens at a wavelength λ_0 such that the system is focused at that wavelength, then the focal length of the lens at some other wavelength λ is given by $f = f_0(\lambda_0/\lambda)^2$. This is illustrated in Fig. 6. Since the index of refraction of material lenses is very close to unity, the focusing is weak. The radius of curvature of the lens must be small, and a large number of lenses means that the part of the beam travels through a large thickness of material. If a Fresnel lens is used the thickness can be reduced considerably, decreasing the attenuation. This is particularly useful for cold neutrons.

The use of both long wavelength neutrons and long flight paths results in the transmitted beam falling under the influence of gravity by an amount that is wavelength dependent. That is, the neutron's vertical position changes by an amount $y = -(g/2)(m\lambda/h)^2 L$ over a distance L , where g is the acceleration due to gravity. Since the incident beam has some divergence, this smearing results in an oval shape to the neutron beam spot at the detector. A prism can compensate for the chromatic aberration caused by gravity because the refractive index for neutrons has a wavelength dependence ($\sim \lambda^2$) similar to that for gravity. An "antigravity device" has been demonstrated using single-crystal prisms such that the transmitted beam is restored to the instrument centerline defined by the source and sample aperture centers. Others have studied both magnetic prisms and the compound refractive prism made of single crystal elements.

63.6 DIFFRACTION AND INTERFERENCE

The Bragg diffraction by single crystals, powders, amorphous materials, and liquids is analogous to that for electromagnetic waves, with formulas similar to those for x-ray diffraction except the magnitude is determined by the coherent scattering lengths. The diffraction pattern is simply a function of scattering angle and is proportional to the square of the Fourier transform of the structure of the diffracting sample. The small-angle neutron scattering from inhomogeneous materials also has expressions similar to those for x rays. The availability of the (partial) substitution of hydrogen by deuterium allows the ability to change the relative scattering powers for different features within the structure.

Most neutron scattering instruments use a large crystal monochromator to select neutrons of a particular wavelength λ from the polychromatic beam from the source to be incident onto the sample.

The mean wavelength reflected from the crystal depends on the Bragg angle θ and the monochromator crystal plane spacing d according to Bragg's law, $v\lambda = 2d \sin \theta$, where v is the order of diffraction. Usually the first order ($v = 1$) is desired because the reflection is most intense. A neutron filter may be desired to remove the higher-order contamination. Monochromators are characterized by their peak reflectivity and their mosaic width obtained by a rocking curve, rotating the crystal between two perfect crystals. The intrinsic angular reflectivity width for a particular wavelength of perfect crystals, called the Darwin width, is far too narrow for their use as neutron monochromators in diffractometers, and though the peak reflectivity is very high, the integrated reflectivity is low. It is necessary to deform the crystals in order to introduce dislocations to form many slightly misaligned crystals, called a mosaic crystal. Alternatively, the perfect crystal may be bent such that the crystal orientation varies across the face of the crystal, such that the crystal lattice spacing has a gradient across its depth (see Chap. 39 for x-ray monochromators). In both cases, the mosaic of the crystal is described by the width of the Gaussian angular crystal plane distribution obtained in a rocking curve measurement. The reflectivity may be reduced by extinction and absorption, and sometimes by simultaneous parasitic scattering. However, perfect crystals are used for some applications, such as high-resolution backscattering instruments, neutron interferometry, and ultra small-angle scattering.

The wavelength resolution of the diffracted beam depends on the angular collimation both before and after the monochromator and on the mosaic spread of the crystal. For maximum reflected intensity at a given resolution it is usual to make the collimations comparable, and the mosaic spread large. Typical crystal mosaic spreads are between 0.2° and 0.5° . The collimation out of the scattering plane should be as relaxed as possible to increase the reflected intensity. Typical neutron monochromators are graphite, silicon, germanium, copper, and beryllium. Monochromators may be used in reflection (Bragg) geometry or in Laue (transmission) geometry. Neutron monochromator efficiencies are higher in reflection geometry than in transmission geometry.

Exact analytic general solutions^{32,33} of the Darwin equations have been derived that describe the multiple Bragg reflection of x rays or neutrons in a mosaic slab crystal. Both the intensity reflection coefficient R and the transmission coefficient T , where $R + T = 1$, can be expressed in compact form for both the Bragg case (reflection geometry) and the Laue case (transmission geometry). For the Bragg case

$$R = b / \{ [a(a+2b)]^{1/2} \coth[a(a+2b)]^{1/2} + (a+b) \}$$

$$T = \frac{[a(a+2b)]^{1/2}}{[a(a+2b)]^{1/2} \cosh[a(a+2b)]^{1/2} + (a+b) \sinh[a(a+2b)]^{1/2}} \quad (35)$$

And for the Laue case

$$R = 1/2 \exp(-a) [1 - \exp(-2b)]$$

$$T = 1/2 \exp(-a) [1 + \exp(-2b)] \quad (36)$$

The dimensionless quantities are defined by $a = \mu t / \sin(\theta + \alpha)$ and $b = \sigma t / \sin(\theta + \alpha)$, where the angles θ and α , respectively, are the Bragg angle and the angle between the reflecting planes and the crystal surface, and t is the thickness of the monochromator crystal. The linear coefficients σ and μ , respectively, are the Bragg reflection coefficient, and the attenuation coefficient for all other processes other than Bragg reflection, including absorption, incoherent scattering, and coherent inelastic scattering. The quantity $\mu = \rho(\sigma_a + \sigma_i)$, where σ_a and σ_i are the absorption and incoherent scattering cross sections and ρ is the atom number density. The quantity σ depends on the incident neutron wavelength λ and the Bragg angle θ and is nonzero only if Bragg's law is satisfied. It is given by $\sigma = Q_c W(\theta - \theta_{hkl})$, where $W(\theta - \theta_{hkl})$ is a normalized rocking curve, and is sharply peaked at $\theta = \theta_{hkl}$, and its shape is characteristic of the mosaic structure of the crystal with Miller indices hkl . Here $Q_c = \lambda^3 |F_{hkl}|^2 / v_0^2 \sin 2\theta_{hkl}$, where F_{hkl} is the structure factor for the unit cell of volume v_0 . This enables the calculation of the reflectivity of an absorbing crystal of finite thickness in situations

where the Bragg planes make an arbitrary angle with the surface of the crystal. The symmetric case is the usual monochromator geometry ($\alpha = 0$) that results in a reflected intensity with an asymmetric intensity distribution across the beam. The asymmetric case is known as a Fankuchen-cut crystal ($\alpha \neq 0$), and for $\alpha > 0$ the beam is compressed and the diffracted beam is more intense.

A considerable gain in intensity for a neutron spectrometer is obtained with an improved monochromator with an anisotropic mosaic structure by focusing in both the horizontal and the vertical directions, to form a double focusing monochromator³⁴ using many crystals assembled together, with little degradation of instrumental resolution. Vertical focusing increases the intensity for a diffraction measurement without affecting the resolution. This is achieved by placing the crystals along an arc of radius of curvature R_V in the vertical plane given by

$$R_V = \frac{2L_1L_2 \sin \theta_B}{L_1 + L_2} \quad (37)$$

where L_1 and L_2 are the distances of the effective source and the sample from the monochromator, or the distances from the sample and the detector if the crystal is used as an analyzer. Horizontal focusing involves Bragg crystal optics (see Chap. 22) and requires the bending of long single crystals in a suitable mechanical device, or placing smaller crystals along an arc of radius of curvature R_H in the horizontal plane given by

$$R_H = \frac{2L_1L_2}{(L_1 + L_2) \sin \theta_B} \quad (38)$$

Diffraction Effects

Various important measurements have illustrated the wave nature of neutrons. Shull³⁵ has demonstrated Fraunhofer slit diffraction using a monochromatic beam of neutrons diffracted from a perfect single crystal of silicon through a slit that was analyzed by a similar crystal. The diffraction broadening of the transmitted beam was shown for several very narrow slit widths to agree with the classical formula $(\sin x/x)^{1/2}$ for diffraction from a single slit. Later experiments³⁶ with greater precision performed for single slit and double slit geometries have showed excellent agreement with wave diffraction theory. The diffraction of thermal neutrons by ruled gratings at grazing incidence has also been demonstrated.³⁷

The Fresnel diffraction of an opaque straight edge³⁸ has been found to be in excellent agreement with theory. The focusing of slow neutrons by zone plates (phase gratings of variable spacings) has been demonstrated.^{30,39} Conventional zone plates, analogous to those for x rays (see Chap. 40), use constructive interference of diffracted waves through open zones, with absorption of radiation in dark zones. Neutron zone plates use phase reversing in alternate zones and can achieve greater focusing efficiency. Zone plates can be produced by photolithography on silicon wafer substrates from a computer-drawn pattern. The resolution is limited by the inherent chromatic aberration of the zone plate. Cylindrical zone plates have been used as a condensing lens.⁴⁰

Interference

There have been various demonstrations of the interference of coherent beams using thermal, cold, and ultracold neutrons.^{41,42} Two beam interferometers allow the determination of the relative phases of two coherent wavefronts and the investigation of different influences, such as force fields and material media, on the phase of a wave. This is again analogous with classical optics. When neutrons are incident on perfect crystals under Bragg conditions, there is a dynamic interchange of flux between the incident beam and the Bragg diffracted beam inside the crystal medium. The radiation density traveling in either the Bragg or the forward direction is a periodic function of depth

in the crystal. This results in extinction effects, with an associated primary extinction length of order (1 to 100) μm . The dynamical theory of diffraction leads to anomalous transmission effects, and there have been numerous experiments using perfect crystals to observe these special effects of neutron wave propagation in periodic media. For example, Shull⁴³ has observed the Pendellösung interference fringe structures in the diffraction from single-crystal silicon.

The superposition of two spatially separated parts of one coherent wavefront produces interference patterns from which the relative phases of the two parts can be deduced. The effective source for the original wavefront is a beam with a spatial coherence produced by passage through a narrow slit from which a partially coherent cylindrical wave emerges. This beam illuminates a pair of closely spaced narrow slits, and the diffracted radiation gives rise to interference fringes in the plane of observation,³⁶ analogous to Young's slits in classical optics.

A similar experiment⁴⁴ has shown the interference of two spatially separated parts of one wavefront using the boundary between two domains of opposite magnetization in a crystal of Fe-3%Si. In traversing the ferromagnetic foil on either side of a domain wall in the sample, the spin of the neutron precesses in opposite directions. The two parts of the wavefunction acquire a relative phase shift that shows up in the interference pattern in the plane of observation. The expected pattern is the superposition of two Fresnel straight-edge patterns that for zero phase shift combine to give a uniform distribution. For π phase shift, destructive interference occurs in the center of the pattern, giving a deep minimum. The relative phase shift depends on the angle of precession of the neutron spin, which is proportional to the thickness of the region of the magnetic field traversed. The results of the experiment demonstrate that a phase shift of π (destructive interference) occurs when the spin is rotated by odd multiples of 2π , verifying the distinctive behavior of spin 1/2 particles under rotation, that the neutron is a spinor.

The coherent splitting of the amplitude of a wave by means of partial reflection from thin films gives rise to interference effects in visible light. For neutrons this partial reflection from the front and back surfaces of various thin films occurs at glancing angles close to critical reflection. The resultant interference phenomena give valuable information about the structure of the interfaces involved.⁴⁵ Various neutron optical devices based on thinfilm interference have been developed, such as bandpass monochromators, supermirrors that are highly efficient neutron polarizers,⁴⁶ reflectivity and polarizing characteristics of Fe-Ge multilayers,⁴⁷ and devices involving multiple-beam interference effects, analogous to classical optics (see Chap. 41 on multilayers for x rays). Multiple-beam interference has been demonstrated⁴⁷ with ultracold neutrons in which thin films of copper and gold are used in the construction of a resonant structure analogous to a Fabry-Perot cavity.

Perfect Crystal Interferometers

Bonse and Hart⁴⁸ have pioneered the x-ray interferometer that uses Bragg reflection by crystal flats to split and recombine two coherent beams. The crystal flats are cut from a monolithic piece of large, highly perfect single-crystal silicon leaving a central backbone. This ensures that all the crystal planes in each crystal flat are perfectly aligned, since each remains part of the original monolithic crystal. This idea has been applied to neutron interferometry, and involves triple Laue transmission geometry optics.⁴⁹ This design is analogous to the Mach-Zehnder interferometer of classical optics. The interferometer illustrated in Fig. 7 comprises three identical, perfect crystal flats cut perpendicular to a set of strongly reflecting planes, with distances between flats usually a few centimeters. There are of order 10^9 oscillations of the neutron wave function in each beam path, so that there are stringent requirements on the microphonic and thermal stability of the device. This requires vibrational isolation of the interferometer from the reactor hall, plus thermal isolation. The three crystal slabs must be aligned within the Darwin width of a few seconds of arc for thermal neutrons in silicon. The precise definition of the wavelength is accomplished by the Bragg diffraction of the interferometer. Placing an object of a given thickness in one of the beams leads to a phase shift or change in the interference contrast. The path difference must be small compared with the coherence length of the beam that depends on the degree of beam monochromatization. There have been many applications of neutron interferometry in fundamental physics, such as gravitationally induced quantum interference and tests of nonstandard quantum mechanics.^{41,42}

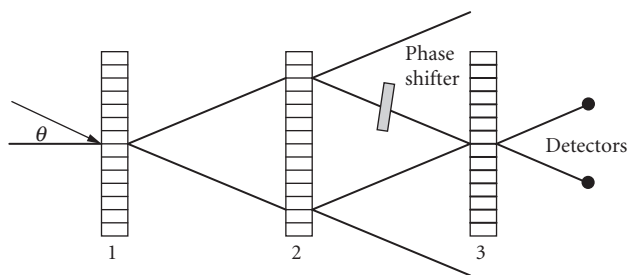


FIGURE 7 A schematic diagram of the triple Laue neutron interferometer cut from a monolithic perfect crystal of silicon in which the first slab splits the incident beam into two coherent parts, with the second slab acting as a mirror to bring the beams together at the third slab. Interference fringes are observed in the detectors by rotating the phase shifter in one of the paths.

The analog of the classical Rayleigh interferometer is a two-crystal monolithic device cut from a large silicon crystal.⁵⁰ The incident beam is restricted by a narrow entrance slit and is Bragg-reflected by the first silicon crystal slab. Inside the crystal, the diffracted neutron beam fills the entire Borrmann triangle, and the beam that leaves the back face of the crystal is broad, having a width that is dependent on the crystal thickness. There is a conjugate focal point on the exit face of the second crystal, from which neutrons along the complementary interfering rays leave the device in the incident and diffracted beam directions. If the central rays are blocked by a cadmium absorber, the parallelogram ray diagram is analogous to the three-crystal interferometer. Another variant uses Bragg reflection geometry rather than Laue transmission geometry. The incident beam reflected from the front surface of the first crystal is brought back to interfere with the beam originating from the back face of this crystal by using a second parallel Bragg reflecting crystal. The advantage is that the spatial definition of the outgoing interfering beam is as sharp as the beam incident on the interferometer.

Perfect crystal neutron interferometry is a convenient and most precise technique for measuring coherent scattering lengths solid, liquids, and gases. It has the advantage of not being limited to liquids as the gravity reflectometer. The period of the intensity oscillations as a function of the orientation of a slab of material that is rotated about an axis perpendicular to the plane of the interferometer enables the determination of the refractive index and hence the scattering length of the slab material. There is also a differential technique using a phase shifter in one of the beams. Neutron interferometry measures the difference in the average neutron-nuclear potential of the sample traversed by the two beams. Intensity oscillations are observed by varying the atom density, with more rapid interference oscillations for larger scattering lengths. The sensitivity of the technique allows a measurement of the hydrogen content in various transition metals at a level of about 0.05 atom fraction.⁵¹

63.7 POLARIZATION TECHNIQUES

Neutron Polarization

It is the spin of $1/2$ of the neutron that results in neutron optics having phenomena quite different from x-ray optics, and enables neutrons beams to be polarized. The neutron spin, $\mathbf{s}$, interacts with the nuclear spin, $\mathbf{I}$, and the optical potential, $V(\mathbf{r})$, and the bound scattering length, b , are both spin-dependent for $I \neq 0$. The total spin, $\mathbf{J} = \mathbf{I} + \mathbf{s}$, is a constant of the motion with eigenvalues $J = I \pm 1/2$. The interaction potential of the magnetic dipole moment μ_n of the neutron in a magnetic field $\mathbf{B}$ is

given by $V_{\text{mag}} = -\mu_n \mathbf{B}$. When a neutron beam traverses a magnetic field, the field defines a quantization axis such that the neutron spin is either “up” or “down” relative to the field direction. That is, the spins of neutrons may be oriented such that their projections are either parallel (+) or antiparallel (–) to the magnetic field. Because there are two possible spin states, a beam with a fraction, ϕ_+ , of its neutrons in the spin “up” state and a fraction, ϕ_- , in the “down” state has a magnitude of its polarization, $\mathbf{P}$, in the field direction given by

$$|\mathbf{P}| = \phi_+ - \phi_- = 2\phi_+ - 1 = 1 - 2\phi_- \quad (39)$$

The total optical potential in a medium containing a magnetic field is the sum of the nuclear and magnetic potentials. Consequently, the index of refraction may be written

$$n^2 = 1 - \lambda^2 [\rho b / \pi \pm m \mu_n B / (2\pi^2 \hbar^2)] \quad (40)$$

This is often written as $n^2 = 1 - \lambda^2 \rho(b \pm p)/\pi$, where p is consider a magnetic scattering length, the sign of which depends on the neutron spin orientation relative to magnetic field, $\mathbf{B}$, in the medium.

This polarization property of a neutron beam gives many special capabilities to neutron scattering and neutron optics.⁵² The refractive index depends on the orientation of the spin with respect to the magnetization direction, and therefore there are two critical glancing angles of reflection, with polarizing mirrors having a greater critical angle for the + state than for the – state. This property of reflection of neutrons from magnetized mirrors is widely used for obtaining polarized neutrons. From Eq. (40), if $\rho b < m \mu_n B / (2\pi^2 \hbar^2)$, neutrons with spins antiparallel to the magnetization are not reflected, and only those neutrons with parallel spin orientation are reflected. However, values of p are less than b , so that devices work with an angular range such that reflection occurs for + spins and transmits for the – spins. Supermirror polarizers⁵³ can be used in transmission or reflection geometry. In transmission geometry the required (–) spin state beam is transmitted through the supermirror and its substrate and the unwanted (+) spin state is reflected from the beam and absorbed elsewhere. In reflection geometry the required (+) spin state beam is reflected, and the unwanted (–) spin state is transmitted through the supermirror and absorbed in the substrate.

Polarizing mirrors are generally ferromagnetic films deposited on silicon substrates and $m = 3$ Fe/Si polarizing mirrors are common. Polarized beams are obtained using total reflection from magnetized iron and cobalt mirrors, with the refractive index of the ferromagnetic material in one spin state matched to that of the nonmagnetic material. Polarizing mirrors produce high polarization and good transmission, but are restricted to low incident angles and wavelengths longer than 2 Å. In addition, thin (≈ 100 nm) evaporated films of magnetic alloys have been used. Two-dimensional multilayer structures have alternate layers of a (magnetized) magnetic material whose refractive index for the (–) spin neutrons is matched to the second (nonmagnetic) layer, for example, in Fe-Ge multilayers. Supermirrors can produce a broadband-polarized beam using evaporated multilayer films deposited to provide a lattice spacing gradient. The reflectivity profile may be extended in angle beyond the simple mirror critical glancing angle through thin-film interference.

A ferromagnetic single crystal may simultaneously polarize and monochromate a neutron beam. A magnetic field applied perpendicular to the scattering vector saturates the magnetic moment along the field direction. Diffraction occurs corresponding to plane spacings with an intensity that is dependent on both the nuclear and the magnetic structure factors. The sign of the latter depends on the direction of the neutron spin relative to the magnetic field. Ideally, if the nuclear and magnetic structure factors are equal in magnitude, so that for one spin state the intensity of diffraction is zero and the other spin state has twice the structure for diffraction. The naturally occurring ferromagnets are the 3d elements Fe, Ni, and Co, but none are good single crystal polarizers. The nuclear scattering lengths for iron and nickel are too large for their weak magnetic moments, while cobalt is a strong neutron-absorbing element. Alloys are found to have a better match of the structure factors, and $\text{Co}_{0.92}\text{Fe}_{0.08}$ (220 reflection with $d = 1.6$ Å) and magnetite Fe_3O_4 are used as polarizing monochromators, but Heusler alloy Cu_2MnAl (111 reflection with $d = 3.43$ Å) is often

preferred, with a reflectivity of 25 to 30 percent for the reflected spin state and a neutron polarization of 95 percent.

There are various polarizing filters that absorb one spin state while transmitting the other spin state. The filter thickness is a compromise between the desired polarization P that increases and the transmission T that decreases with filter thickness. High polarization ($P > 0.96$) with reasonable transmission ($T > 0.2$) requires that the spin-independent cross section is small compared to the spin dependent absorption cross section. Resonance absorption polarization filters depend on the spin dependent absorption cross section of polarized nuclei at their nuclear resonant energy. Cooling ^{149}Sm or ^{151}Eu in a magnetic field can provide efficient polarizing filters with reasonable nuclear polarization over a wide energy range. A better filter is obtained with nuclear-spin polarized ^3He that takes advantage of the huge absorption cross section of ^3He that comes from only the $(-)$ spin state, so that it preferably absorbs neutrons with spins antiparallel to the ^3He spin, while transmitting the $(+)$ spin state.

Neutron capture by a ^3He nucleus results in an excited ^4He nuclear state that subsequently decays into a proton and a triton. The cross section for capture of thermal neutrons with spin antiparallel to the ^3He nuclear spin is 10666 b , whereas that for neutrons with spin parallel is essentially zero. Theoretically with a sufficient gas density of 100 percent polarized ^3He , all neutrons with antiparallel spin are absorbed, while nearly all neutrons with parallel spin are transmitted, resulting in 100 percent neutron polarization and 50 percent transmission. For an initially unpolarized neutron beam that passes through a cell filled with ^3He with less than 100 percent polarization, the transmitted beam has a polarization P_n given by

$$P_n(\lambda) = (n_+ - n_-)/(n_+ + n_-) = \tanh[\sigma(\lambda)\rho_{\text{He}}LP_{\text{He}}] \quad (41)$$

and a transmission T_n given by

$$T_n(\lambda) = T_0 \cosh[\sigma(\lambda)\rho_{\text{He}}LP_{\text{He}}] \quad (42)$$

where n_+ and n_- are the numbers of neutrons with parallel and antiparallel spin, $\sigma(\lambda)$ is the absorption cross section for unpolarized neutrons of wavelength λ , ρ_{He} is the number density of ^3He , L is the length of the cell, and P_{He} is the nuclear polarization of the ^3He . For $P_{\text{He}} = 0$, T_0 is the transmission through an evacuated cell, neglecting the transmission loss of the glass windows. ^3He can be polarized either by spin exchange⁵⁴ with optically pumped Rb or metastability-exchange optical pumping methods,⁵⁵ and ^3He nuclear polarizations near 75 percent have been attained. The characteristics of the filter depend on its opacity, which is defined by the product of the cell length, the ^3He gas pressure, and the neutron wavelength (proportional to the absorption cross section). Figure 8 shows the neutron polarization P_n and the transmission T_n for a ^3He nuclear polarization $P_{\text{He}} = 0.75$ as a function of the cell opacity. This indicates that there is a trade-off between neutron polarization and transmission, and a figure-of-merit is usually given by $P_n^2 T_n$. Note that shorter wavelength neutrons require a thicker target to produce the same polarization.

A spin flipper is a device that can reverse the direction of polarization of a beam. A thin flat coil with a relatively strong field inside is placed perpendicular to the neutron beam. Within the coil the beam polarization direction is no longer parallel to the magnetic field and it begins to precess around the field. The particular field strength of the flipper and its direction determine whether the neutrons precess exactly by π or $\pi/2$ by the time the beam has passed through the flipper.

Larmor Precession

When a polarized neutron beam enters a homogeneous magnetic field B that is perpendicular to the neutron magnetic moment, the direction of the neutron spin precesses in B with a Larmor frequency given by $\omega_L = 2|\mu_n B|/\hbar = \gamma B$, where γ is the gyromagnetic ratio of the neutron. This phenomenon is utilized by the spin echo spectrometer,⁵⁶ in which the incoming and outgoing velocities of the neutron

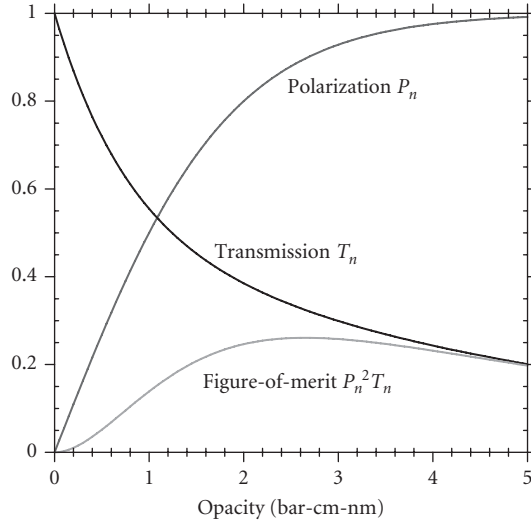


FIGURE 8 Neutron polarization P_n , transmission T_n , and figure-of-merit $P_n^2 T_n$ as a function of opacity, the product of the ^3He pressure in bar, assuming a room temperature cell, the cell length L in cm, and the wavelength λ in nm, for a nuclear polarization $P_{\text{He}} = 0.75$ of the ^3He .

are measured, so that small changes in velocity that may take place on scattering can be inferred from the shift in phase in the Larmor precession angle. This is effectively a time-of-flight method in which polarized neutrons precess in the magnetic field to provide its own clock. Such a technique enables energy resolution in the neV region, orders of magnitude smaller than for conventional instruments.

Cold neutrons from a neutron guide are coarsely monochromatized by a velocity selector and polarized by supermirrors. The spins are made horizontal by a $\pi/2$ spin flipper. The polarized neutron beam enters the first magnetic field B_1 of length L_1 and the spins precess many times. The total precession angle of a neutron spin is $\phi_1 = \gamma B_1 L_1 / v_1$, where v_1 is the neutron velocity. After scattering the spins are reversed in direction by a π flipper, and then precess in the reverse direction. The total precession angle of a neutron spin in the second magnetic field B_2 of length L_2 is $\phi_2 = \gamma B_2 L_2 / v_2$, where v_2 is the final neutron velocity. The spins are flipped again by $\pi/2$, and the polarization of the neutron beam is analyzed. If the scattering is elastic, $v_1 = v_2$, and if $B_1 L_1 = B_2 L_2$, the final beam has the same polarization. If there is a change in energy $\hbar\omega$ upon scattering where ω is small, the total change in the precession angle is given by $\phi_{\text{tot}} = \gamma B L / v^3 (\hbar\omega/m)$.

The analogue to the Stern-Gerlach effect arises from the interaction between the neutron magnetic moment μ_n and an inhomogeneous magnetic field that bends the neutron trajectory by imparting an acceleration

$$d^2\mathbf{r}/dt^2 = \mp |\mu/m| |\nabla| \mathbf{B}| \quad (43)$$

A superconducting hexapole magnet of typical length ≈ 2 m can simultaneously focus and polarize a neutron beam. If the incident neutron beam is unpolarized, the trajectory of those neutrons with spins oriented parallel to the direction of the magnetic field bend toward the magnet axis, and those antiparallel bend away from the axis. As a result half the neutrons are spin polarized about the local magnetic field and are focused a distance that is inversely proportional to λ . Because the effect index of refraction of the magnetic field lens is near unity, the focusing is weak and depends on the strength of the hexapole magnetic field.

63.8 NEUTRON DETECTION

Neutrons are uncharged and do not directly ionize atoms, and consequently they easily penetrate most materials. They cannot be detected directly, and their detection requires a nuclear reaction within a converter that results in the emission of a charged particle or ionizing radiation that can more readily interact with other materials.^{13,57} (Fast neutrons may be detected by proton recoil or after thermalization.) The efficiency of absorption of neutrons incident normally on an absorber of thickness ℓ is given by $1 - \exp[-\rho\sigma_a\ell]$, where ρ is the atom number density of the absorbing nuclei and σ_a the absorption cross section. The three most common reactions used for the detection of slow neutrons all have cross sections that increase linearly with wavelength (see Table 1). All three reactions are exothermic with Q values much greater than the energy of slow/thermal neutrons (≈ 25 meV) so that there is no energy discrimination. (Note that energy or wavelength determination of a neutron beam requires a crystal spectrometer or a mechanical monochromator.) The choice of the reaction depends on its ability to discriminate against gammas. The reaction Q value determines the energy released, and the greater the energy, the easier it is to discriminate against gammas.

1. The $^{10}\text{B}(n,\alpha)$ reaction has a thermal (velocity of 2200 m s^{-1} , or wavelength of $\lambda \approx 0.179 \text{ nm}$) absorption cross section of 3835 b , and a Q value of 2.792 MeV to decay directly the ground state (6 % branch ratio), or a Q of 2.310 MeV for decay to an excited state with the subsequent emission of a 478-keV gamma (94%) with a half life of about 10^{-3} s . The reaction products are an alpha particle with an energy of either 1.77 MeV (6%) or 1.47 MeV (94%), and a ^7Li ion with an energy 1.01 MeV (6%) or 0.84 MeV (94%). This reaction is used in BF_3 gas proportional counters or in B-loaded scintillators. The natural isotopic abundance of ^{10}B is 19.8 percent.
2. The $^6\text{Li}(n,\alpha)$ reaction with a thermal absorption cross section of 940 b and a Q value of 4.78 MeV emits no gamma and is commonly used in neutron scintillator detectors. The reaction products are an alpha particle with an energy of 2.05 MeV and a ^3H ion with an energy of 2.73 MeV . While the absorption cross section is lower for this reaction than for ^{10}B , the Q value of the reaction is higher and therefore more ionized charge is produced. The natural isotopic abundance of ^6Li is 7.4 percent.
3. The $^3\text{He}(n,p)$ reaction, with an unpolarized thermal absorption cross section of 5333 b and a Q value of 0.764 MeV , is used in ^3He gas-proportional counters. The reaction products are a proton with an energy of 0.573 MeV and a ^3H ion (triton) with an energy of 0.191 MeV . Though the absorption cross section is much higher than the other two isotopes, it is relatively expensive. The lower Q value can be compensated by high electron multiplication in ^3He proportional counters.

In all three cases, the reactions products are emitted in opposite directions randomly oriented in space. These reactions are also used in the one- and two-dimensional detectors, and often form the basis of other detection methods, such as imaging detectors and neutron imaging plates. (They are also key reactions used in neutron depth profiling.) In general cadmium and gadolinium, or other rare earth elements, that have large thermal absorption neutron cross sections are not considered good detector materials, because they result in the copious production of gammas.

Neutron detectors may be operated in either current mode or in pulse mode. In the former, the detector events are integrated either over a short time to provide a beam monitor, or over a longer time as in a charge coupled device camera for beam alignment, or in a photographic plate or an imaging plate detector used for neutron radiography or sometimes in neutron diffraction. These detectors can operate with high intensity and can have high spatial resolution. Generally they have a large dynamic range, but may suffer from low efficiency, gamma sensitivity, and high noise. In the pulse mode individual neutrons are detected in a gas counter or a scintillator detector, and sometimes a converter foil; these are used in most neutron spectroscopy measurements. These detectors generally have high efficiency, low noise, and can discriminate against gammas, though they may be limited in their count rate and have limited spatial resolution. Discrete detector and multiwire proportional

counters (MWPCs) with coincidence encoding require little signal processing and can operate at relatively high counting rates. Signals from each detection element can be processed individually, allowing high data rates. Resolution can be improved if the charge (or scintillation light) from each neutron detection event is distributed over several detection elements.

Gas Detectors

The gas detector may be operated in either a current or a pulse mode. When the charged particles from the reaction pass through a gas, the primary mode of interaction is ionization and the excitation of gas molecules along the particle track. The average energy lost by the incident particle per ion pair formed depends on the particular gas, the type of radiation and its energy. Typical values are about (30 to 35) eV per ion pair. For instance a 1 MeV particle that is fully stopped within the gas produces slightly greater than 3×10^4 ion pairs. The operation of the ionization chamber is based on the collection of all charge created by direct ionization within the gas by the application of an electric field. Two metallic electrodes surround the fill gas of ^3He or $^{10}\text{BF}_3$. The two emitted particles slow down in the gas and the charges of the electron-ion pairs produced along their tracks are collected on the plates using an electric field strength of ≈ 100 V/cm. The ^3He reaction produces 0.764 MeV kinetic energy in the charged particles and results in about 2.5×10^4 electron-ion pairs. The ionization chamber is usually used in current mode and serves as a low efficiency beam monitors on most neutron beam lines.

By increasing the electric field within the gas, the original ion pairs created within the gas are accelerated toward the electrodes, leading to additional ionization of the gas molecules by the accelerated electrons. This secondary ionization leads to an avalanche of charge, and the density of the electron cloud grows exponentially with distance as the avalanche progresses. It terminates when the central anode wire has collected all the free electrons produced. This is the ubiquitous proportional counter, with the total charge reaching the anode proportional to the number of primary electrons; and the gas multiplication or gain can be as high as $\approx 10^5$. Individual neutron events are easily detected. The charge collection time is about 10 μs , so that the proportional counter is count rate limited to about 3×10^4 s $^{-1}$. The fill gas must not have appreciable electron attachment, and must quench any remaining charge to prevent the migration of the heavier ions. This is generally performed by inert gases and the best is P10 gas (90% argon and 10% CH_4 by volume). Partial pressure in the ^3He counter is adjusted for the required neutron absorption efficiency, and a pressure of (10 to 20) bar can be used. When the reaction takes place near the edge of the detector housing, some of the charge is lost by the "wall effect" such that the full charge avalanche does not take place. However, a heavy gas with a high stopping power to reduce the charged particle range may ameliorate this. The differential pulse height spectrum is such that a lower level discriminator may be used to eliminate the gamma ray background.

The gas proportional counter is noiseless and has high charge amplification with good gamma discrimination and can stand high radiation fields. They are usually cylindrical, though detectors with elliptical cross sections are also available. The neutron sensitive fill gas is usually ^3He . It may be made position-sensitive by having an amplifier at both ends of the anode wire. In the charge division method, the collected charge is divided in proportion to the resistance to ground at each end along the wire. The positional sensitivity depends on the highly resistive anode wire. The other method is the observation of the rise time from preamplifiers at either end of the resistive anode wire. Typical spatial resolution may be as high as 1.5 mm. The multiwire gas proportional counter MWPC is a two-dimensional detector with a large number of parallel anode wires in a plane mounted between two cathodes composed of strips in orthogonal directions. The multiple electrodes are enclosed within a large single gas chamber. As before the gas multiplication of electrons is directed toward the anode wire, and positive charge is induced on the cathodes. These enable the positional determination in the two directions using coincidence encoding. The positional resolution is greater than 1 mm. The charge collection time is about 1 μs , so that the two-dimensional proportional counter is limited to about 3×10^5 events s $^{-1}$. At high detection rates, the space charge reduces the electric field around the wire and hence the charge gain. The microstrip detector⁵⁸ uses a photolithography technique, and

replaces the usual multiwire anode plane to reduce space-charge effects and gives higher counting rates and better spatial resolution.

Neutron Scintillators

All gas detectors require some form of gamma discrimination, and only glass scintillators lead to tolerable separation between neutron and gamma signals. The energy from the charged particle raises electrons within the material to excited states that decay back to the ground state by the emission of photons. The scintillator must be transparent to its own emitted radiation. Often the scintillating material must be mixed with a compound that contains the absorbing nuclei, such as a lithium compound dispersed in a matrix of ZnS(Ag). A ZnS(Cu, Ag, Au) scintillator is superior because it gives green light that is better for CCD (charge-coupled device) detection. ^6Li -based scintillators are favored over ^{10}B -based scintillators, because the greater energy released by the ^6Li reaction results in greater light production. Fused $\text{B}_2\text{O}_3/\text{ZnS}$ has less effective gamma discrimination. The crystals often contain small amounts of an activating impurity to provide the necessary electronic states to produce suitable scintillating properties. The light produced in the scintillator may be directly coupled to a light detection device such as a photomultiplier tube, though this may also be a channel-plate amplifier, a CCD camera, or photographic film. Alternatively, the scintillator can be coupled through light guide or optical fibers, or through a system of lenses and mirrors. A lithium-loaded ZnS scintillator with an image intensifier is often used to monitor the profile of the neutron beam with the output of observed on a screen.

The range of the ionizing particles is only a few microns within the scintillator material, enabling much greater spatial resolution. A thin detector limits the gamma sensitivity and the same time limits the detection efficiency. Typical scintillators include ZnS/ ^6LiF activated with Ag, ^6LiI crystal activated with Eu, and ^6Li glass activated with Ce, with the first mostly used, having $\approx 1.6 \times 10^5$ photons produced per absorbed neutron. Only a fraction of these photons are collected in the photomultiplier or other light-sensitive device, so that the quantum efficiency of a typical photocathode may be ≈ 20 percent. Photomultiplier tubes also have to suffer from dark current, which can be decreased by cooling the tube. The spread in pulse amplitudes from a scintillator detector is typically 10 to 20 percent and is the limitation on the spatial resolution of scintillator position-sensitive detectors (PSDs).

Scintillating glass fibers that are loaded with both ^6Li and Ce can be operated with a spatial resolution of 100 μm . The triton emitted from the reaction can excite a Ce^{3+} ion, and the subsequent deexcitation of the electron to the ground state results in the emission of visible light that is transmitted through the fiber and collected by the photomultiplier tube. These fibers can be arranged to form a large area position-sensitive detector. The principal disadvantage of these glass fibers is the sensitivity to gammas.

Photographic film may have a large dynamic range with high spatial resolution of $\approx 100 \mu\text{m}$, but as an integrating device is sensitive to gammas. The Anger camera consists of a scintillator that is coupled by optical fibers to a number of photomultipliers. The positional encoding of the neutron event is determined by centroiding the output signal, with a precision much smaller than the size of the individual detector element, with a resolution of several mm.

Solid State Detectors

The deposition of the particle energy in a semiconductor results in the formation of electron-hole pairs. The energy required to form such pairs is ≈ 3 eV. FWHM pulse-height resolutions are generally better than 0.5 percent. Semiconductor diodes have a depletion layer to form the active region of the detector. The depletion layer thickness is adjusted by the bias voltage on the diode, with thicknesses up to several mm in high purity Si or Ge. The electrons and holes are collected with no amplification, analogous to the gas ionization chamber, and B or Li are common additions in semiconductor devices as the neutron absorber. The ranges of the primary charged particles is short, so that the position resolution should be good. The detectors are also sensitive to gammas.

The cross sections for neutron induced fission reactions in ^{233}U , ^{235}U , and ^{239}Pu are relatively large at low neutron energies. The reaction Q values are extremely large, ≈ 200 MeV, with approximately 160 MeV being released as kinetic energy of the fission fragments. These produce a large charge in the ionization chamber.

When the neutron-sensitive material is a solid film, the density of the material is so high that the path lengths in the film are very short, and the probability of the charged particle escaping and reaching the detection medium is reduced. The maximum detection efficiency is obtained with an optimum foil thickness that depends on the range of the charged particle in the foil, and this depends on whether the foil detection is in transmission or backscattering geometry. A problem with foils is that only one charged particle is detected and even this loses varying amounts of its energy before entering the charged particle detector. Hence the pulse height spectrum is broad, and gamma discrimination is difficult.

Gadolinium foil can also be used in conjunction with a microchannel plate (MCP) detector where the MCP acts as an electron multiplier. Greater neutron detection efficiency is obtained when the neutron absorbing material (such as ^6Li or ^{10}B) is imbedded within the glass of the MCP. The reaction products create secondary electrons that are attracted by a positive bias and cause an electron avalanche that travels down the channel. This results in thousands of electrons emerging from the MCP due to a single neutron. The position of the electronic pulses may be determined directed using a two-dimensional wire grid or indirectly on a phosphor screen to produce a proportional light image. Typical MCPs consists of ^{10}B -doped glass with thousands of $5\text{ }\mu\text{m}$ diameter individual channels. Spatial resolution approaching $10\text{ }\mu\text{m}$ is possible with good detection efficiency. Sometimes these MCPs are arranged in a chevron configuration to increase the electron gain. Also microsphere plate detectors operate on principles similar to MCPs, and spatial resolutions better than $250\text{ }\mu\text{m}$ have been achieved. These devices are essentially insensitive to gammas.

Imaging Plates

Photographic film can be used to detect the light produced in a neutron-sensitive scintillator screen or in conjunction with a neutron-sensitive foil such as Gd to detect the charge reaction directly. The path length of the charge particles is quite short in the high-density photographic emulsion, enabling good spatial resolution with photographic film. The best spatial resolution of $\approx 20\text{ }\mu\text{m}$ is achieved with good quality film in contact with a Gd foil enriched with ^{157}Gd . A resolution of 0.5 mm is obtained with a standard neutron polaroid camera using a scintillator screen of $\text{ZnS(Ag)}^6\text{LiF}$. The optical densities of the photographic films can be digitized with optical scanners. These integrating detectors have no data rate limitations, but can provide no time information and therefore cannot be used in time-of-flight applications.

The concept of storage photostimulable phosphors or imaging plates involves radiation trapping in a phosphor-coated detector plate and the release of the trapped energy as light when the plate is stimulated under scanning in a reader with a fine laser beam of typical spot size of $50\text{ }\mu\text{m}$. A photomultiplier tube that detects the light is coordinated with the scanned positions to create a digital image of the radiation field. The light intensity is linear with the received radiation level over a wide dynamic range ($> 10^5$). The information on the detector plate can be erased after scanning so that it can be reused. It is far more sensitive than film autoradiography. The detectors have a spatial resolution ranging from 25 to $200\text{ }\mu\text{m}$ with a single detector plate of size up to $350 \times 420\text{ mm}$.

Reusable photostimulable phosphors suitable to neutron detection incorporate converter material such as Gd_2O_3 or LiF within the phosphor plate itself. Yet the inherent sensitivity of the imaging plate to gamma radiation produce high background interference. In addition, residual radioactivity from neutron activated materials, particularly from europium, and the depletion of neutron-sensitive materials in the imaging plate after repeated usage may require frequent calibration. An alternative is a two-step transfer process within which a foil is placed in the neutron field and then is transfer later to the imaging plate to record the radioactivity distribution in the converter and reproduce the neutron field information. Dysprosium is the most efficient neutron converter, decays with a half life of 2.35 h almost entirely by beta emission, and the activity is linear for over 5 orders of magnitude.

63.9 REFERENCES

1. V. F. Sears, *Neutron Optics: An Introduction to the Theory of Neutron Optical Phenomena and Their Applications*, Oxford University Press, Oxford, 1989.
2. M. Utsuro (ed.), "Advance in Neutrons Optics and Related Research Facilities," *J. Phys. Soc. Japan* **65** (suppl. A): (1996).
3. W. Marshall and S. W. Lovesey, *Theory of Thermal Neutron Scattering: The Use of Neutrons for the Investigation of Condensed Matter*, Clarendon Press, Oxford, 1971.
4. G. L. Squires, *Thermal Neutron Scattering*, Cambridge University Press, Cambridge, 1978.
5. S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter*, Oxford University Press, Oxford, 1984.
6. M. L. Goldberger and F. Seitz, "Theory of the Refraction and the Diffraction of Neutrons by Crystals," *Phys. Rev.* **71**:294–310 (1947).
7. V. F. Sears, "Neutron Scattering Lengths and Cross Sections," in *Methods of Experimental Physics: Neutron Scattering*, K. Sköld and D. L. Price (eds.), vol. 23A, Academic Press, San Diego, CA, 1986, pp. 521–550.
8. V. F. Sears, "Neutron Scattering Lengths and Cross Sections," *Neutron News* **3**(3):26–37 (1992).
9. L. Koester, H. Rauch, and E. Seymann, "Neutron Scattering Lengths: A Survey of Experimental Data and Methods," *Atomic Data and Nuclear Data Tables* **49**:65–120 (1991).
10. L. Koester and H. Ungerer, "Coherent Scattering Amplitudes Measured by Small Angle Scattering of Neutrons," *Z. Phys.* **219**:300–310 (1969).
11. H. B. Stuhmann, "Molecular Biology," in *Methods of Experimental Physics: Neutron Scattering*, K. Sköld and D. L. Price (eds.), vol. 23C, Academic Press, San Diego, CA, 1986, pp. 367–403.
12. *Neutron Radiography*, Proceedings of the 8th World Conference on Neutron Radiography, Gaithersburg, MD, M. Arif and R.G. Downing, (eds.), DEStech Publications Inc., Lancaster, PA, 2008.
13. K. M. Beckurts and K. Wirtz, *Neutron Physics*, Springer-Verlag, Berlin, 1964.
14. J. M. Carpenter and W. B. Yelon, "Neutron Sources," in *Methods of Experimental Physics: Neutron Scattering*, K. Sköld and D. L. Price (eds.), vol. 23A, Academic Press, San Diego, CA, 1986, pp. 99–196.
15. D. H. Kim, C. S. Gil, J.-D. Kim, and J. Chang, "Generation and Validation of a Shielding Library Based on ENDF/B-VI.8," *Radiation Protection Dosimetry* **115**:232–237 (2005).
16. C. G. Windsor, "Experimental Techniques," in *Methods of Experimental Physics: Neutron Scattering*, K. Sköld and D. L. Price (eds.), vol. 23A, Academic Press, San Diego, CA, 1986, pp. 197–257.
17. H. Maier-Leibnitz and T. Springer, "The Use of Neutron Optical Devices on Beam-Hole Experiments," *React. Sci. Tech. (J. Nucl. Energy A/B)* **17**:217–225 (1963).
18. P. Böni, "New Concepts for Neutron Instrumentation," *Nucl. Instrum. Meth. A* **586**:1–8 (2008).
19. V. F. Sears, "Theory of Multilayer Neutron Monochromators," *Acta Cryst.* **A39**:601–608 (1983).
20. I. Anderson, "From 1 to m: The Development of Supermirrors," *SPIE Proc.* **1738**:118–124, *Neutron Optical Devices and Applications*, C. F. Majkrzak and J. L. Wood (eds.) (1992).
21. J. R. D. Copley and C. F. Majkrzak, "Calculations and Measurement of the Performance of Converging Neutron Guides," *SPIE Proc.* **983**:93–104, *Thin-Films, Neutron Optical Devices: Mirrors, Supermirrors, Multilayer Monochromators, Polarizers, and Beam Guides* (C. Majkrzak, ed.) (1988).
22. C. Schanzer, P. Böni, U. Filges, and T. Hils, "Advanced Geometries for Ballistic Neutron Guides," *Nucl. Instrum. Meth. A* **529**:63–68 (2001).
23. A. K. Freund, "Cross-sections of Materials used as Neutron Monochromators and Filters," *Nucl. Instrum. Meth.* **213**:495–501 (1983).
24. J. B. Hayter, "Neutron Optics at the Advanced Neutron Source," *SPIE Proc.* **1738**:2–7, *Neutron Optical Devices and Applications*, C. F. Majkrzak and J. L. Wood, (eds.) (1992).
25. A. Steyerl, "Neutron Physics," in *Springer Tracts in Modern Physics*, vol. 80, pp. 57–130, Springer-Verlag, Berlin, 1977.
26. L. Koester, "Absolutmessung der Kohärenten Streulänge von Quecksilber mit den Neutronen-Schwerkraft Refraktometer am FRM," *Z. Phys.* **182**:326–328 (1965).
27. G. E. Ice, C. R. Hubbard, B. C. Larson, J. W. L. Pang, J. D. Budai, S. Spooner, and S. Vogel, "Kirkpatrick-Baez Microfocusing Optics for Thermal Neutrons," *Nucl. Instrum. Meth. A* **539**:312–320 (2005).

28. Q. F. Xiao, H. Chen, V. A. Sharov, D. F. R. Mildner, R. G. Downing, N. Gao, and D. M. Gibson, "Neutron Focusing Optic for Submillimeter Materials Analysis," *Rev. Sci. Instrum.* **65**:3399–3402 (1994).
29. C. S. Schneider and C. G. Schull, "Forward Magnetic Scattering Amplitude of Iron for Thermal Neutrons," *Phys. Rev.* **B3**:830–835 (1971).
30. R. Gähler, J. Kalus, and W. Mampe, "An Optical Instrument for the Search of a Neutron Charge," *J. Phys.* **E13**:546–548 (1980).
31. M. R. Eskildsen, P. L. Gammel, E. D. Isaacs, C. Detlefs, K. Mortensen, and D. J. Bishop, "Compound Refractive Optics for the Imaging and Focusing of Low-Energy Neutrons," *Nature* **391**:563–566 (1998).
32. V.F. Sears, "Bragg Reflection in Mosaic Crystals I. General Solution of the Darwin Equations," *Acta Cryst.* **A53**:35–45, and "Bragg Reflection in Mosaic Crystals II. Neutron Monochromator Properties," *Acta Cryst.* **A53**:46–54 (1997).
33. H.-C. Hu, "Universal Treatment of X-ray and Neutron Diffraction in Crystals I. Theory," *Acta Cryst.* **A53**:484–492, and "Universal Treatment of X-ray and Neutron Diffraction in Crystals II. Extinction," *Acta Cryst.* **A53**:493–504 (1997).
34. C. Broholm, "Proposal for a Doubly Focusing Cold Neutron Spectrometer at NIST," *Nucl. Instrum. Meth. A* **369**:169–179 (1996).
35. C. G. Schull, "Single-Slit Diffraction of Neutrons," *Phys. Rev.* **179**:752–754 (1969).
36. A. Zeilinger, R. Gähler, C. G. Schull, and W. Treimer, "Experimental Status and Recent Results of Neutron Interference Optics," *AIP Conf. Proc.* **89**:93 (1981).
37. H. Kurz and H. Rauch, "Diffraction of Thermal Neutrons by a Ruled Grating," *Z. Phys.* **220**:419–426 (1969).
38. R. Gähler, A. G. Klein, and A. Zeilinger, "Neutron Optical Tests of Nonlinear Wave Mechanics," *Phys. Rev.* **A23**:1611–1617 (1981).
39. P. D. Kearney, A. G. Klein, G. I. Opat, and R. A. Gähler, "Imaging and Focusing of Neutrons by a Zone Plate," *Nature* **287**:313–314 (1980).
40. A. G. Klein, P. D. Kearney, G. I. Opat, and R. A. Gähler, "Focusing of Slow Neutrons with Cylindrical Zone Plates," *Phys. Lett.* **83A**:71–73 (1981).
41. S. A. Werner and A. G. Klein, "Neutron Optics," in *Methods of Experimental Physics: Neutron Scattering*, K. Sköld and D. L. Price (eds.), vol. 23A, Academic Press, San Diego, CA, 1986, pp. 259–337.
42. H. Rauch and S. A. Werner, *Neutron Interferometry: Lessons in Experimental Quantum Mechanics*, Oxford University Press, Oxford, 2000.
43. C. G. Shull, "Observation of Pendellösung Fringe Structure in Neutron Diffraction," *Phys. Rev. Lett.* **21**:1585–1589 (1968).
44. A. G. Klein and G. I. Opat, "Observation of 2π Rotations by Fresnel Diffraction of Neutrons," *Phys. Rev. Lett.* **37**:238–240 (1976).
45. R. R. Highfield, R. K. Thomas, P. G. Cummins, D. P. Gregory, J. Mingins, J. B. Hayter, and O. Schärpf, "Critical Reflection of Neutrons from Langmuir-Blodgett Films on Glass," *Thin Solids Films* **99**:165–172 (1983).
46. F. Mezei, "Novel Polarized Neutron Devices: Supermirror and Spin Component Amplifier," *Commun. Phys.* **1**:81 (1976); F. Mezei and P. A. Dagleish, "Corrigendum and First Experimental Evidence on Neutron Supermirrors," *Commun. Phys.* **4**:41 (1977).
47. K.-A. Steinhauser, A. Steyerl, H. Scheckhofer, and S. S. Malik, "Observation of Quasibound States of the Neutron in Matter," *Phys. Rev. Lett.* **44**:1306–1309 (1981).
48. U. Bonse and M. Hart, "An X-Ray Interferometer," *Appl. Phys. Lett.* **6**:155–156 (1965).
49. H. Rauch, W. Treimer, and U. Bonse, "Test of a Single Crystal Neutron Interferometer," *Phys. Lett.* **47A**:369–371 (1974).
50. A. Zeilinger, C. G. Schull, M. A. Horne, and G. Squires, "Two-Crystal Neutron Interferometry," in *Neutron Interferometry*, U. Bonse and H. Rauch (eds.), Oxford University Press, London, 1979, pp. 48–59.
51. H. Rauch, E. Seidl, A. Zeilinger, W. Bausspiess, and U. Bonse, "Hydrogen Detection in Metals by Neutron Interferometry," *J. Appl. Phys.* **49**:2731–2734 (1978).
52. W. G. Williams, *Polarized Neutrons*, Oxford University Press, Oxford, 1988.
53. P. Böni, D. Clemens, M. S. Kumar, and C. Pappas, "Applications of Remanent Supermirror Polarizers," *Physica B* **267–268**:320–327 (2007).

54. W. C. Chen, G. Armstrong, Y. Chen, B. Collett, R. Erwin, T. R. Gentile, G. L. Jones, J. W. Lynn, S. McKenney, and J. E. Steinberg, " ^3He Spin Filters for a Thermal Neutron Triple Axis Spectrometer," *Physica B* **397**:168–171 (2007).
55. E. Lelièvre-Berna, "Mid-Term Report on the NM13 Neutron Spin Filter Project," *Physica B* **397**:162–167 (2007).
56. F. Mezei, *Neutron Spin Echo*, Springer-Verlag, Berlin, 1980.
57. G. F. Knoll, *Radiation Detection and Measurement*, 2nd ed., chap. 14, John Wiley & Sons, New York, 1989.
58. A. Oed, "Micro Pattern Structures for Gas Detectors," *Nucl. Instrum. Meth. A* **471**:109–114 (2001).

This page intentionally left blank

GRAZING-INCIDENCE NEUTRON OPTICS

Mikhail Gubarev and Brian Ramsey

*NASA/Marshall Space Flight Center
Huntsville, Alabama*

64.1 INTRODUCTION

Due to their unique sensitivity to light elements, neutrons are an important tool for progress in material and life sciences, medical research, and therapy as well as for fundamental physics. However, the availability of useful neutron beams is limited as neutron sources typically have relatively low brilliance (compared, e.g., to x-ray sources.) Focusing and concentrating of neutrons, through the use of suitable optics, offers a way to significantly increase neutron current density at the samples under investigations. Such a capability becomes particularly important for experiments with very small samples. In addition, neutron focusing can also improve the signal to noise ratio for such applications as small angle neutron scattering (SANS) analysis and neutron crystallography, when the neutron beam can be focused onto the detector.

This chapter is concerned with focusing neutron optics, that is, the optical elements capable of producing an image of the source through a limited but determined number of reflections. Other optical elements, such as capillary optics and neutron guides, utilize a stochastic number of reflections to transport or concentrate neutrons. In these, neutrons undergo multiple reflections from guide or capillary walls to emerge in a new direction. These elements are described in Chaps. 53 and 63.

64.2 TOTAL EXTERNAL REFLECTION

The optical properties of materials with respect to neutrons are characterized by their refractive indices¹ which are a function of the neutron wavelength

$$n = 1 - \delta + i\beta \quad (1)$$

with

$$\delta = \frac{\lambda^2 N_d b}{2\pi}$$

and

$$\beta = \frac{\lambda N_d \sigma_a}{4\pi}$$

where N_d is the atomic number density, b is the coherent scattering length, and σ_a is the absorption cross-section. For most materials (with the exception of those containing Li, B, Cd, Sm, or Gd) the absorption cross-section is almost zero and Eq. (1) can be reduced to

$$n = 1 - \delta$$

The parameter δ is strongly dependent on neutron wavelength and in the case of thermal neutrons, is about 10^{-5} for most elements and their isotopes. So, as is the case for x rays, the refractive index for neutrons is slightly less than unity. Therefore, thermal and cold neutrons can be reflected from smooth surfaces at shallow “grazing-incidence” angles (total external reflection) or be refracted at the boundaries of different materials. Both reflective and refractive optics can be used to focus and concentrate neutrons. The advantage of neutron reflective optics, reviewed here, is that they are achromatic compared to refractive optics.

Total external reflection occurs when neutrons are incident on a surface at angles below the critical angle ϑ_{cr} , given by

$$\vartheta_{cr} = \sqrt{2\delta} \quad (2)$$

so that at $10 \text{ \AA}$, for example, the critical angle for reflection off nickel is approximately 1° .

Because the neutron absorption cross section is negligible for most materials, the reflectivity, which can be calculated by the Fresnel equations (see, e.g., Chap. 63), is almost 100 percent below the critical angle. This permits the development of efficient grazing-incidence neutron optics.

64.3 DIFFRACTIVE SCATTERING AND MIRROR SURFACE ROUGHNESS REQUIREMENTS

Mirror surfaces must be quite smooth as diffractive scattering off surface irregularities produces the wings of a mirror’s point-spread function (PSF). The grating equation provides the relationship between the diffractive scattering angle θ for neutrons with wavelength λ incident, at an angle α , onto a surface with roughness at spatial frequency f

$$\theta \sin \alpha = f \lambda \quad (3)$$

The neutron flux excluded from an angular aperture with radius θ , the total integrated scatter (TIS) is given by

$$\text{TIS}(>\theta) = \int_{\theta'} \psi_{sc} 2\pi \vartheta' d\vartheta' = \left(\frac{4\pi \sin \alpha}{\lambda} \right)^2 \int_{\theta \sin \alpha / \lambda} \text{PSD}(f') df' = (4\pi \sin \alpha)^2 \left(\frac{\sigma}{\lambda} \right)^2 \quad (4)$$

where ψ_{sc} is the single reflection point spread function, PSD is the power spectrum density function of the surface, and σ is the surface roughness [the rms value, calculated for the spatial frequencies

$f > \theta(\sin\alpha/\lambda)$]. The formula describes the single reflection case and for double reflection, as for many optical configurations, the single reflection point spread function has to be multiplied by a factor of two.

At a given neutron wavelength and surface roughness, the excluded fraction of neutrons increases with grazing angle, so the worst case occurs at the critical angle of the surface material. For a nickel surface the critical angle is given by

$$\vartheta_{\text{cr}}(\text{rad}) = 1.73 \times 10^{-3} \times \lambda(\text{\AA}) \quad (5)$$

So, the worst case TIS for a nickel surface (single reflection) is

$$\text{TIS}_s(>\theta_{\text{cr}}) \approx 5 \times 10^{-4} \sigma_s^2 \quad (6)$$

for spatial wavelengths $1/f < 60/\theta$, where σ_s is the surface roughness (the rms value in $\text{\AA}$). In practice, optical surfaces can be polished to the level of a few $\text{\AA}$, so neutron scattering due to the surface microroughness can be kept to acceptable levels. Effects calculated above, though, will be doubled for 2-bounce mirror systems.

64.4 IMAGING FOCUSING OPTICS

Optical elements such as elliptical, toroidal, Kirpatrick-Baez (KB) or Wolter mirrors (see Chap. 44), and “tapered” (elliptical or parabolic) reflectors (see Chap. 52) utilize single or double reflections to image neutron sources and to produce fine focusing. The shallow critical angles for neutrons lead to mirror designs with large mirror lengths and small apertures. With low neutron currents, especially for thermal neutron sources, this dictates the use of nested systems to improve throughput. Fabrication of large super-polished high-figure-quality surfaces is a challenging task in general as figure errors and the roughness of the mirror surface determine the optical performance.

The specific properties of neutrons and the typical size of the neutron sources pose additional challenges (over similar x-ray optics) for development of high performance optics for beam control. Neutron sources are typically extended and also less brilliant compared to x-ray sources. Thus on-axis optical schemes, originally developed for point-like x-ray sources, may not deliver optimum optical performance. The major portion of an extended neutron source would be off the mirror’s optical axis, and this results in higher optical aberrations compared to an on-axis point source if classical focusing schemes are used. These aberrations can be minimized to some extent by collimating a neutron source down for applications when the neutron current is not so critical and by optimizing the figure of the mirror to boost the off axis optical performance. In addition, the effects of gravity, which need not be considered for x rays, are significant in neutron applications. The dependence of velocity on wavelength makes a grazing incidence neutron optic gravitationally chromatic.

Optical Elements: Configurations

Depending on the experimental task and resources available, the optimal optical configuration for a specific neutron application may include single, double, or more reflections.

Optical schemes using single reflection for neutron focusing can be divided into two subgroups based on neutron source configuration. An elliptical scheme is applicable for divergent point-like sources. In this case the source is placed at the first ellipsoid focal point and the neutrons are focused at the second focal point of the ellipsoid. A parabolic scheme would be applicable for quasi-parallel neutron beams, where a single-bounce optical element would be placed after straight neutron guides. The schematics of neutron focusing based on these optical configurations are shown in Figs. 1 and 2.

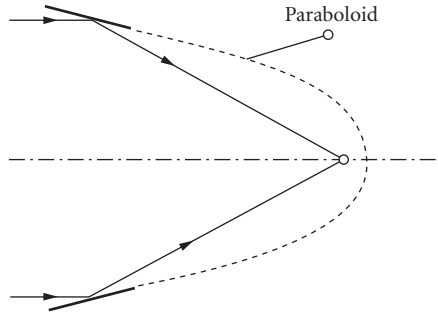


FIGURE 1 A single parabolic reflector used to focus a quasi-parallel input neutron beam.

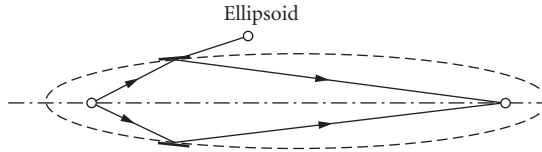


FIGURE 2 A single ellipsoidal reflector used to focus a divergent point source of neutrons.

A significant problem with any grazing-incidence focusing system is image degradation due to astigmatism and spherical and coma aberrations. The use of true ellipsoidal (for finite source) and paraboloidal (for infinite source) mirrors would eliminate the astigmatism and spherical aberration components, but in practice the figure of single-bounce grazing incidence mirrors is commonly approximated to ease mirror fabrication. Astigmatism, the optical aberration due to the difference in the meridional and sagittal focal distances of a mirror, can be eliminated when a toroidal mirror is used for neutron focusing. The mirror is astigmatism-free, if the sagittal and meridional radii of a toroidal mirror are set such that the corresponding focal distances are equal. This requirement is satisfied if the ratio between the sagittal and meridional radii of the mirror is

$$\frac{R_s}{R_m} = \sin^2 \alpha$$

where R_s and R_m are the sagittal and meridional radii of the mirror curvature and α is the grazing angle. Toroidal neutron mirrors have been used to focus neutrons at the ILL Spin-Echo Spectrometer² and at the Jülich SANS instrument and reflectometer.^{3,4} (The Jülich toroidal mirror has now been transferred to the FRM neutron source in Garching.⁵) Elliptical and parabolic tapered neutron guides have also been used as a single-bounce systems.^{6–10}

The principal disadvantage of single-bounce systems is off-axis aberrations, especially coma. The off-axis coma is due to the difference in angular magnification for neutrons reflected from different areas of the mirror located on the same mirror meridian and it cannot be corrected using only one reflection. Image deterioration caused by this optical aberration is directly proportional to the mirror length. The aberration might in practice be severe because neutron sources are extended and the quasi-parallel neutron beams formed by straight neutron guides have nonzero divergence, so in both these cases many neutrons are nonaxial. To limit the focal spot blur of a single-bounce optical system due to off-axis coma one can therefore decrease the mirror length and/or collimate the neutron source down. However, both lead to reduced throughput, but this can be compensated in some cases

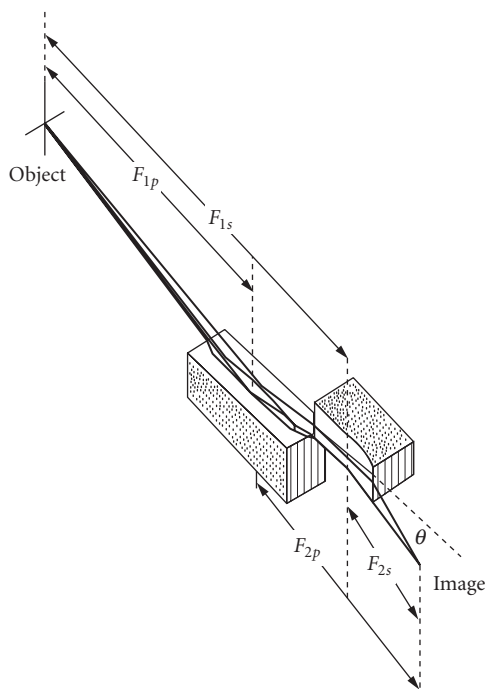


FIGURE 3 The Kirkpatrick-Baez mirror configuration uses a single reflection from each of two orthogonal mirrors. (Courtesy of Gene Ice, Oak Ridge National Laboratory.)

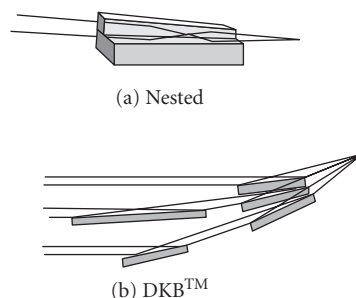


FIGURE 4 Two KB optical configurations for a neutron microscope. (Courtesy of Gene Ice, Oak Ridge National Laboratory.)

by the use of a nested optical system. A distinct advantage of single-bounce systems, however, is the ease of construction and alignment.

Another way to simplify the fabrication of grazing incidence mirrors is to use an optical scheme with two or more consecutive reflections from separate surfaces orthogonal to each other. Such a scheme was first described by Kirkpatrick and Baez (KB) for spherical mirrors¹¹ and is shown schematically in Fig. 3. The KB optical scheme, which has elliptical and parabolic mirrors for focusing beams from finite and infinite sources, respectively, eliminates astigmatism and spherical aberrations. However, because the reflections appear in orthogonal planes the KB schemes suffers from off-axis coma the same way as the single-bounce optical systems. The KB mirrors are relatively easy to fabricate but are relatively difficult to align due to sensitivity to the tilt errors. However, if bendable mirrors are used the misalignment can be often corrected by refocusing-rebending.

The KB systems provide good optical performance in small apertures if short mirrors are used, making the system attractive if neutrons need to be focused onto small samples. To increase the aperture, nested systems are needed and the KB systems lends itself readily to this with the mirrors stacked together in two orthogonal planes. A few nesting schemes have been proposed for KB microscopes and two examples are shown in Fig. 4. In one method, shown in Fig. 4a, mirrors are nested against each other so that the first reflection can be in either the horizontal or vertical plane. Another method, termed deflected KB (DKB), utilizes flat deflection mirrors to “steer” the incoming flux into the elliptical focusing mirrors, enabling a larger collecting area. A one-dimensional schematic illustrating focusing in one plane is depicted in Fig. 4b with the nested elliptical mirrors shown at the right. Inner rays enter directly, but outer rays are first deflected by the plane mirrors on the left. A second system, orthogonal to this, would provide focusing in the other dimension. Greater neutron current gains are

possible with more deflections. However, the increase in effective aperture due to reflection from an additional mirror have to be traded against the increase in image blur due to the neutron scattering from additional mirror surface's microroughness. In cases where the KB optical system is intended for neutron imaging, the difference in magnifications in orthogonal planes, proportional to the distance between the two orthogonal mirrors, needs to be taken into account.

The KB scheme has been successfully used for the neutron focusing.¹² The smallest focal spot achieved to date was using bendable mirrors with the figure adjusted to approximate an ellipse. The focal spot for neutrons with wavelength of 0.1 nm was measured to be $89 \times 90 \mu\text{m}$ (FWHM).¹³

The astigmatism of grazing incidence mirrors can be easily corrected by the use of optical geometries with an even number of reflections from the confocal surfaces of revolution. The on-axis coma is eliminated while the off-axis coma can be also be reduced for large apertures. With these aberrations either eliminated or reduced, the image degradation due to the obscured aperture diffraction would need to be taken into account.¹⁴ Since each reflection from a nonideal surface, in a multibounce system, results in additional scattering due to microroughness, the two bounce systems have been developed the most. Such systems, first discussed by Wolter,¹⁵ are used extensively in x-ray astronomy and can also be designed for use with cold and thermal neutron beams. In Wolter geometries two mirrors with surface figures of second order such as a paraboloid, hyperboloid, or ellipsoid are arranged coaxially. A schematic of two of the most used Wolter mirror configurations is shown in Fig. 5.

Wolter geometries with paraboloid, hyperboloid, and ellipsoid mirrors are optimized to provide superior optical performance on-axis and are capable of delivering better optical performance compared to single-bounce and KB systems. It has been previously shown in the literature that the application of a polynomial approximation to the surface of grazing incidence mirror shells^{16,17} as well as defocusing¹⁸ enhances the global performance of the mirror over the entire field-of-view. Nested Wolter-1 geometry optics can greatly improve the focused beam intensity by increasing the incident beam area while keeping the optical aberrations low. For reference, a picture of an x-ray optic module with 12 nested mirrors is shown in Fig. 6.

The focusing capabilities of neutron imaging optics based on Wolter geometries has been successfully demonstrated using a beam of cold neutrons.¹⁹ A test mirror originally designed as a 1/10-scale

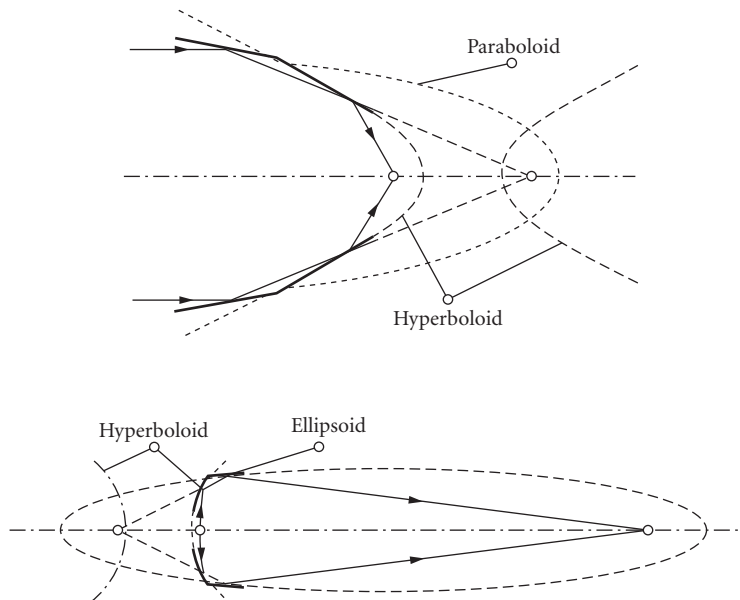


FIGURE 5 Commonly used Wolter-1 mirror configurations. (See also color insert.)



FIGURE 6 An x-ray optic module with 12 Wolter-I nested mirrors.

version of the innermost mirror of NASA's Chandra X-Ray Observatory²⁰ was fabricated using an electroformed nickel replication process, wherein a thin nickel mirror shell is electroformed off a superpolished and figured mandrel.²¹ In the experiment the area of the optic illuminated by the neutron beam was estimated to be 17.7 mm^2 . A focal-spot size, defined mostly by the divergence of the incoming beam, was measured to be 1.15 mm (FWHM).

Optical Elements: Materials

The largest possible critical angle observed for an isotope is that of Ni^{58} , and so this is the material of choice for grazing incidence optics. Typically, optical elements are fabricated from another more traditional material such as zerodur, glass, or silicon, and then the mirror surface is coated with Ni^{58} . In the case of the nickel replication technique, natural nickel can be used to produce the mirror shell with a small loss of critical angle. In some applications it is desirable to use nonmagnetic materials to preserve neutron polarization and for these Cu^{65} with an aluminum-layer overcoat to prevent oxidation can be used.³

Multilayer coatings, typically alternating layers of nickel and titanium, can be deposited on the surface of a curved neutron mirror to increase the acceptance angle of the mirror. Such a neutron "supermirror" efficiently reflect neutrons with wavelengths λ when the mirror angle is less than a critical angle,¹³ given by

$$\vartheta_{\text{cr}} (\text{rad}) = m \times 1.73 \times 10^{-3} \times \lambda (\text{\AA}) \quad (7)$$

where m is the ratio of the maximum supermirror angle compared to the maximum efficient scattering angle of a nickel surface. Typical practical values of m are 3 to 4.

64.5 REFERENCES

1. V. F. Sears, *Neutron Optics*, Oxford University Press, Oxford, 1989, p. 64.
2. C. Hayes, C. Lartigue, A. Kolmar, J. R. D. Copley, B. Alefeld, F. Mezei, D. Richter, and T. Springer, "The Focusing Mirror at the ILL Spin-Echo Spectrometer IN15: Experimental Results," *Journal of the Physical Society of Japan* **65**(suppl. A):312–315 (1996).

3. B. Alefeld, C. Hayes, F. Mezei, D. Richter, and T. Springer, "High-Resolution Focusing SANS with Toroidal Neutron Mirror," *Physica B* **234–236**:1052–1054 (1997).
4. B. Alefeld, L. Dohmen, D. Richter, and Th. Bruckkel, "X-Ray Space Technology for Focusing Small-Angle Neutron Scattering and Neutron Reflectometry," *Physica B* **283**:330–332 (2000).
5. Institut Fur Festkorperforschung: http://www.fz-juelich.de/iff/pics_pdfs/ism/2007/PNI_Kentzinger.pdf, accessed May 2009.
6. S. Yamada, T. Shinohara, H. Sasao, T. Oku, J. Suzuki, H. Matsue, and H. M. Shimizu, "Development of a Multichannel Parabolic Guide for Thermal Neutron Beam Focusing," *Physica B* **385–386**:1243–1246 (2006).
7. N. Kardjilov, P. Böni, A. Hilger, M. Strobl, and W. Treimer, "Characterization of a Focusing Parabolic Guide Using Neutron Radiography Method," *Nuclear Instruments and Methods in Physics Research A* **542**: 248–252 (2005).
8. T. Hils, P. Boeni, and J. Stahn, "Focusing Parabolic Guide for Very Small Samples," *Physica B* **350**:166–168 (2004).
9. S. Mühlbauer, M. Stadlbauer, P. Böni, C. Schanzer, J. Stahn, and U. Filges, "Performance of an Elliptically Tapered Neutron Guide," *Physica B* **385–386**:1247–1249 (2006).
10. S. Mühlbauer, P. G. Niklowitz, M. Stadlbauer, R. Georgii, P. Link, J. Stahn, and P. Böni, "Elliptic Neutron Guides—Focusing on Tiny Samples," *Nuclear Instruments and Methods in Physics Research A* **586**:77–80 (2008).
11. P. Kirkpatrick and A. Baez, "Formation of Optical Images by X Rays," *Journal of Optical Society of America* **38**: 766 (1948).
12. G. E. Ice, C. R. Hubbard, B. C. Larson, J. W. L. Pang, J. D. Budai, S. Spooner, and S. C. Vogel, "Kirkpatrick-Baez Microfocusing Optics for Thermal Neutrons," *Nuclear Instruments and Methods in Physics Research A* **539**:312–320 (2005).
13. G. E. Ice, C. R. Hubbard, B. C. Larson, J. W. L. Pang, J. D. Budai, S. Spooner, S. C. Vogel, R. B. Rogge, J. H. Fox, and R. L. Donabarger, "High-Performance Kirkpatrick-Baez Supermirrors for Neutron Milli- and Micro-Beams," *Materials Science and Engineering, A* **437**:120–125 (2006).
14. P. L. Thompson and J. E. Harvey, "Aplanatic Wolter Type 1 Telescope Design: Is there a Practical Advantage?," *Proceedings of SPIE* **3444**:526–542 (1998).
15. H. Wolter, "Spiegelsysteme streifend einfallend als abbildende optiken für Röntgenstrahlen," *Annalen der Physik* **6**(10):94 (1952).
16. P. Conconi, G. Pareschi, S. Campana, G. Chincarini, and G. Tagliaferri, "Wide-Field X-Ray Imaging for Future Missions, Including XEUS," *Proceedings of SPIE* **5168**:334–345 (2004).
17. C. J. Burrows, R. Burg, and R. Giacconi, "Optimal Grazing-Incidence Optics and Its Application to Wide-Field X-Ray Imaging," *Astrophysical Journal* **392**:760 (1992).
18. L. P. VanSpeybroeck and R. C. Chase, "Design Parameters of Paraboloid-Hyperboloid Telescopes for X-Ray Astronomy," *Applied Optics* **11**(2):440–445 (1972).
19. M. V. Gubarev, B. D. Ramsey, D. E. Engelhaupt, J. M. Burgess, and D. F. R. Mildner, "An Evaluation of Grazing-Incidence Optics for Neutron Imaging," *Nuclear Instruments and Methods in Physics Research B* **265**(2):626–630 (2007).
20. Chandra X-Ray Center: http://cxc.harvard.edu/cdo/about_chandra/overview_cxo.html, accessed May 2009.
21. B. D. Ramsey, "Replicated Nickel Optics for the Hard-X-Ray Region," *Experimental Astronomy* **20**(1):85–92 (2006).

INDEX

Index note: The *f* after a page number refers to a figure, the *n* to a note, and the *t* to a table.

- Abbe's sine condition, 45.3
- Aberrations:
- of gratings and monochromators, 38.7
 - in grazing incidence optics, 45.1–45.8, 45.2*f*
 - in grazing incidence telescopes, 44.6–44.12, 44.7*f*, 44.9*f*–44.11*f*
 - in grazing-incidence neutron optics, 64.4, 64.6
 - and imaging through atmospheric turbulence, 4.17–4.20, 4.17*f*–4.18*f*, 4.19*t*, 4.20*t*, 4.27–4.30, 4.27*f*
 - Zernike modes of, 5.11, 5.11*t*, 5.12*f*
- Absolute method (scatterometer calibration), 1.15
- Absorption:
- and atmospheric optics, 3.4–3.5, 3.4*f*
 - of Cr^{3+} , 2.19–2.21, 2.19*f*, 2.20*f*
 - electro-, 13.55–13.56, 13.56*f*, 13.58
 - excited state, 14.6
 - measurements of, 2.2–2.13, 2.4*f*, 2.6*f*, 2.8*f*, 2.10*f*, 2.12*f*
 - molecular, 3.12–3.15, 3.13*f*, 3.22–3.23, 3.22*f*, 3.23*f*
 - in neutron optics, 63.6
 - photo-, 36.1
 - in photonic crystal fibers, 11.20–11.21, 11.20*f*
 - and reabsorption of spectra, 56.8
 - and x-ray optics, 26.7
- ac Kerr effect, 7.11
- Achromaticity, of fiber-based couplers, 16.3–16.4
- Acoustically rotated tangential phase matching, 6.26*f*, 6.27
- Acousto-optic devices, 6.3–6.45
- and acousto-optic diffraction, 6.4, 6.9
 - Bragg cells (wideband), 6.30, 6.31*t*
 - deflectors, 6.22–6.30, 6.24*f*, 6.26*f*, 6.28*f*, 6.29*t*
 - figures of merit for, 6.16–6.17
 - materials for, 6.16–6.22
- Acousto-optic devices (*Cont.*):
- modulators, 6.31–6.35, 6.34*t*
 - acousto-optic frequency shifters, 6.35
 - and Bragg diffraction, 6.4, 6.6, 6.7, 6.14
 - image (scophony), 6.34–6.35
 - principle of operation, 6.32, 6.32*f*
 - optical birefringence in, 6.17
 - propagation and attenuation in, 6.17
 - theory of acousto-optic interaction, 6.5–6.16
 - elasto-optic and roto-optic effects, 6.5–6.6
 - of finite geometry, 6.14–6.16
 - frequency characteristics, 6.12–6.14, 6.13*f*, 6.14*f*
 - phase matching, 6.9–6.12, 6.10*f*
 - plane wave analysis, 6.6–6.9
 - tunable filters as [*see* Acousto-optic tunable filters (AOTFs)]
 - wideband AO Bragg cells, 6.30, 6.31*t*
- Acousto-optic diffraction, 6.4, 6.9
- Acousto-optic figures of merit, 6.16–6.17
- Acousto-optic frequency shifters (AOFS), 6.35
- Acousto-optic interaction, theory of, 6.5–6.16
- elasto-optic and roto-optic effects, 6.5–6.6
 - of finite geometry, 6.14–6.16
 - frequency characteristics, 6.12–6.14, 6.13*f*, 6.14*f*
 - phase matching, 6.9–6.12, 6.10*f*
 - plane wave analysis, 6.6–6.9
- Acousto-optic modulators, 6.23*t*, 6.31–6.35, 6.34*t*
- acousto-optic frequency shifters, 6.35
 - image (scophony), 6.34–6.35
 - principle of operation, 6.32, 6.32*f*
- Acousto-optic tunable filters (AOTFs), 6.23*t*
- collinear beam, 6.43, 6.43*f*, 6.45*t*
 - long-infrared, 6.42

- Acousto-optic tunable filters (AOTFs) (*Cont.*):
 and longitudinal spatial modulation, 6.12
 mid-infrared, 6.42
 noncritical phase-matching, 6.37, 6.39–6.42, 6.38f, 6.43t
 principle of operation, 6.36–6.39, 6.37f, 6.38f
 ultraviolet, 6.42
- Activated phosphor detectors, 60.7–60.8
- Active mode locking, 20.17
- Active optics, 50.1, 50.2
- Adaptive microstructured optical arrays (MOAs), 50.7–50.8, 50.8f
- Adaptive optics (AO), 5.1–5.46, 50.1
 concepts and components of, 5.2–5.5, 5.3f
 designing systems of, 5.38–5.46
 requirements, 5.39, 5.40t
 results for 3.5-m telescope systems, 5.43–5.45, 5.43f–5.45f
 results for 10-m telescope systems, 5.45–5.46, 5.45f–5.46f
 subaperture size, 5.40–5.43, 5.42f
 as enabling technology, 5.2
 hardware and software implementation for, 5.21–5.38
 higher-order wavefront sensing techniques, 5.36–5.37
 laser beacons, 5.27–5.34, 5.28f–5.31f, 5.33f
 real-time processors, 5.34–5.35, 5.34f, 5.35f
 Shack-Hartmann technique, 5.23–5.27, 5.23f, 5.25f, 5.26f
 tracking, 5.21–5.23, 5.22f
 wavefront correctors, 5.37–5.38, 5.38f
 and imaging through atmospheric turbulence, 4.35–4.36
 and turbulence, 5.5–5.21
 anisoplanatism, 5.19
 atmospheric tilt and Strehl ratio, 5.14, 5.14f–5.15f
 Fried's coherence diameter and spatial scale, 5.9–5.13, 5.11t, 5.12f, 5.13f
 higher-order phase fluctuations, 5.18–5.19
 on imaging and spectroscopy, 5.19–5.21, 5.19f, 5.20f
 Kolmogorov model, 5.5–5.6
 tracking requirements, 5.15–5.18, 5.17f
 variation of n and C^2n parameters, 5.6–5.8, 5.7f, 5.8f
- Adaptive x-ray optics, 50.1–50.8
 hard vs. soft x-ray telescopes, 50.2
 history of, 50.1, 50.2f
 for synchrotron and lab-based sources, 50.2–50.7, 50.4f, 50.6f, 50.8f
- Add/drop multiplexers, optical, 21.2, 21.8, 21.8f, 21.9f, 21.12
- ADONIS system, 5.35
- Advanced X-Ray Astrophysical Facility (AXAF), 44.4, 44.6f, 44.10
- Aerosols, 3.6, 3.7, 3.10–3.11, 3.10f, 3.11f
- Air cladding, for photonic crystal fibers, 25.2, 25.21
- Air-filled photonic crystal fibers, 25.20–25.21
- Airy disc, 37.6
- Airy distribution, of thin lenses, 40.3
- Akhieser loss, 6.17
- Algorithms, singular-value-decomposition, 5.25, 5.26
- All solid-core photonic crystal fibers, 25.21
- All-fiber monolithic systems (fiber lasers), 25.9–25.11, 25.16–25.18, 25.16f
- All-optical switching, in OTDM communication networks, 20.22, 20.23f
- All-solid bandgap guiding fibers, 11.15, 11.16, 11.17f
- Aluminum oxide (Al_2O_3), 2.19, 2.19f, 2.21, 2.22
- Amplification:
 chirped pulse, 25.2, 25.32, 25.33
 lump, 21.44
 by SOAs, 19.1f–19.2f, 19.2, 19.22–19.27
 CWDM systems, 19.27
 DWDM systems, 19.25–19.27, 19.25f, 19.26f
 single-channel systems, 19.22–19.24, 19.22f–19.24f
- Amplified spontaneous emission (ASE), 9.13, 10.11, 14.3, 14.6
- Amplified spontaneous emission (ASE) noise, 19.3, 19.4f, 19.9, 19.18, 19.24, 19.35
- Amplified spontaneous emission (ASE) power, 19.20
- Amplifiers:
 in communication systems, 9.13–9.14
 erbium-doped fiber, 14.4–14.7, 14.5f, 21.38–21.41, 21.38f–21.42f
 erbium/ytterbium-doped fiber, 14.7–14.8
 field-effect transistor, 13.70
 linear optical, 19.14, 19.27
 optical fiber, 14.1–14.11, 14.2t

- Amplifiers (*Cont.*):
 parametric, 11.23, 11.24, 14.10–14.11
 praseodymium-doped fiber, 14.7
 Raman, 14.8–14.9, 14.8*f*, 14.10*f*,
 21.42–21.44, 21.42*f*–21.44*f*
 rare-earth-doped, 14.2–14.4, 14.3*f*
 semiconductor optical [see Semiconductor
 optical amplifiers (SOAs)]
 semiconductor vs. fiber, 9.13–9.14
 slab-coupled optical waveguide, 19.21
 in WDM networks, 21.37–21.44, 21.37*f*
 ytterbium-doped fiber, 14.7
- Amplitude modulation, 7.22–7.24, 7.23*f*, 7.24*f*
- Amplitude modulation index, 7.22
- Amplitude zone plates, 40.4–40.5, 40.5*f*
- Amplitude-modulated (AM) acoustic waves, 6.32
- Amplitude-modulated (AM) systems, optical
 fibers in, 9.16
- Amplitude-shift-keying (ASK), 21.28, 21.29
- Analog modulation, 6.32
- Analog to digital conversion, 20.4, 20.8, 20.8*f*
- Analog transmission, on optical fibers,
 9.15–9.17, 9.16*f*
- Anger cameras, 63.33
- Angle of arrival, 4.23–4.26, 4.25*f*, 4.26*f*
- Angle of deflection, 6.40
- Angle scans, for scatterometers, 1.14
- Angled stripe designs, of SOAs, 19.8, 19.8*f*
- Angular apertures, of NPM AOTFs, 6.41
- Angular dispersion, 38.6
- Angular resolution:
 in astronomical x-ray optics, 47.10–47.11, 47.11*f*
 in Wolter x-ray optics, 47.2–47.3, 47.2*f*
- Angular spectrum of plane waves (ASW)
 approach, 6.14–6.15
- Anisoplanatism, 5.19
 and laser beacons, 5.27–5.29, 5.28*f*–5.30*f*
 and subaperture size, 5.42–5.43
- Anisotropic acoustic beam collimation, 6.25
- Anisotropic diffraction, 6.10
- Anisotropy, 17.2
- Annealed proton exchange (APE), 7.30
- Annular pupils, 4.10–4.16, 4.11*f*–4.15*f*, 4.15*t*
- Anodes, as x-ray tube sources, 54.8–54.9,
 54.12, 54.13
- ANSI (American National Standards Institute),
 23.4, 23.6
- Antiguides, for VCSELs, 13.46
- Antireflection (AR) coatings, 19.8, 19.8*f*, 19.20
- Antiresonance, 11.12
- Anti-Stokes scattering, 10.5
- Anti-Stokes waves, 10.9
- Apertures:
 angular, 6.41
 Gaussian, 37.6
 numerical, 9.4, 25.2, 25.18, 42.2
 order-selecting, 40.4, 40.5
 pinhole, 32.3
 and subaperture size, 5.40–5.43, 5.42*f*
- Apodization, 17.6–17.7, 17.7*f*
- Arabidopsis thaliana*, 37.10–37.11, 37.10*f*
- Area-detection x-ray imaging, 61.2, 61.2*f*
- ARROW structures, 11.15
- Arsenic trisulfide (As_2S_3), 12.7
- ASML Alpha Tool, 34.2, 34.3*f*
- Astigmatism, 13.12, 64.4, 64.6
- ASTM (American Society for Testing
 Materials), 1.4
- Astrale aberrations, 45.4
- Astronomical x-ray optics, 47.1–47.11
 angular resolution of, 47.10–47.11, 47.11*f*
 hard, 47.9–47.10
 history of, 47.1–47.2
 Kirkpatrick-Baez optics, 47.7–47.8, 47.8*f*, 47.9*f*
 Wolter, 47.2–47.7, 47.2*f*, 47.4*f*, 47.6*f*, 47.7*f*
- Astronomy, x-ray, 33.1–33.4, 33.3*t*
- ATM (asynchronous transfer mode), fiber optic
 standards and, 23.6
- Atmosphere, standard, 3.6–3.11, 3.7*t*, 3.8*f*–3.11*f*
- Atmospheric Infrared Sounder (AIRS), 3.39
- Atmospheric optics, 3.3–3.45
 and absorption of light, 3.4–3.5, 3.4*f*
 and atmospheric optical transmission,
 3.22–3.26, 3.22*f*–3.27*f*
 and composition of standard atmosphere,
 3.6–3.11, 3.7*t*, 3.8*f*–3.11*f*
 and global climate change, 3.43–3.45, 3.44*f*
 and meteorological optics, 3.40–3.43,
 3.41*f*–3.43*f*
 remote sensing in, 3.36–3.40, 3.37*f*–3.40*f*
 and theory of interaction of light with
 atmosphere, 3.11–3.22
 inelastic optical processes, 3.21–3.22
 Mie scattering, 3.16–3.18, 3.17*f*–3.19*f*
 molecular absorption, 3.12–3.15, 3.13*f*
 molecular emission and thermal spectral
 radiance, 3.18, 3.20, 3.20*f*
 molecular Rayleigh scattering, 3.15–3.16
 surface reflectivity and multiple scattering,
 3.21, 3.21*f*

- Atmospheric optics (*Cont.*):
 turbulence in, 3.26, 3.28–3.36
 beam spreading, 3.32–3.33, 3.33*f*
 beam wander, 3.31–3.32
 imaging and heterodyne detection, 3.34
 parameters for, 3.28–3.31, 3.29*f*, 3.30*f*
 scintillation, 3.34–3.36, 3.35*f*
- Atmospheric tilt, 5.14, 5.14*f*–5.15*f*
- Atmospheric turbulence, imaging through,
 4.1–4.37
 aberration variance and approximate Strehl
 ratio for, 4.27–4.28, 4.28*f*
 and adaptive optics, 4.35–4.36
 angle of arrival fluctuations, 4.23–4.26, 4.25*f*,
 4.26*f*
 and covariance and variance of expansion
 coefficients, 4.20–4.22, 4.21*t*,
 4.22*f*–4.23*f*
 Kolmogorov turbulence and atmospheric
 coherence length, 4.7–4.10, 4.8*f*, 4.9*f*
 long-exposure images, 4.3–4.7
 modal correction of turbulence, 4.28–4.30,
 4.29*t*, 4.30*f*
 and modal expansion of aberration function,
 4.17–4.20, 4.17*f*–4.18*f*, 4.19*t*, 4.20*t*
 and resolution of telescopes, 4.2–4.3
 short-exposure image, 4.31–4.35,
 4.31*f*–4.34*f*, 4.35*t*
 and systems with annular pupils, 4.10–4.16,
 4.11*f*–4.15*f*, 4.15*t*
- Atomic energy levels, 2.2–2.5, 2.4*f*
- Atomic scattering, x-ray optics and, 36.1, 36.2*f*
- Atomic spectra, 2.13
- AT&T (American Telephone & Telegraph), 21.1
- Attenuated total reflectance (ATR) waveguides,
 12.11
- Attenuation:
 defined, 21.13
 fiber, 21.13–21.14
 for fiber optic communication links, 15.7
 Lambert-Beer law of, 63.11
 linear, 31.1, 31.2
 neutron, 63.11–63.12
 in optical fibers, 9.4, 9.5*f*
 in photonic crystal fibers, 11.19–11.22,
 11.20*f*, 11.21*f*
 x-ray, 31.1–31.4, 31.2*f*, 31.3*f*
- Attenuators:
 for fiber-based couplers, 16.4
 for networking, 18.2, 18.9
 variable optical, 21.12, 21.13*f*
- Auger cascades, 59.4
- Auger energies, 36.3*t*, 36.9*t*
- Auger excitation peaks, 29.3
- Auger recombination, 13.27–13.28
- Autocovariance (ACV) function, 44.13
- Avalanche photodiode (APD) receivers, 9.8,
 9.10–9.11
- Avalanche photodiodes, 13.63, 13.71–13.73
- Average unregistered detected point spread
 function (AUDPSF), 44.14, 44.15*f*
- Axial and circumferential slope errors,
 45.7–45.8
- Babinet-Soleil compensators, 7.22
- Backward Brillouin scattering, 11.25
- Ballistic neutron guides, 63.17–63.18
- Balloon payloads, in astronomical x-ray optics,
 47.10
- Band edges, of photonic crystal fibers,
 11.9–11.10
- Band gap energy, 19.2
- Band-filling modulators, 13.62
- Bandpass response, in acousto-optic
 interaction, 6.15–6.16
- Bandwidth:
 defined, 20.1
 fiber, 21.2–21.3, 21.2*f*
- Batman doping, 25.20
- Beacons, laser, 5.27–5.34
 focus anisoplanatism, 5.27–5.29, 5.28*f*–5.30*f*
 mesospheric sodium laser beams, 5.32–5.34,
 5.33*f*
 Rayleigh, 5.30–5.32, 5.31*f*
- Beam spatial coherence, 58.4
- Beam splitters, 18.6, 18.6*f*
- Beam spreading, 3.32–3.33, 3.33*f*
- Beam steering, 6.27–6.29, 54.11
- Beam wander, 3.31–3.32
- Beamlines, for multilayer Laue lenses,
 42.9–42.10, 42.9*f*–42.12*f*
- Beer-Lambert law, 3.11, 3.20, 3.21
- Bend, of liquid crystals, 8.22, 8.23*f*
- Bend loss, of photonic crystal fibers,
 11.21–11.22, 11.21*f*
- Bendable optics, 50.3–50.4, 50.4*f*
- Bending magnet synchrotron radiation sources:
 brightness of, 55.9
 power of, 55.8–55.9, 55.8*f*
 radiation from, 55.3–55.6, 55.5*f*, 55.6*f*
- Bent crystals, 39.5–39.6, 39.5*f*–39.6*f*
- BepiColombo mission, 49.6

- Bessel functions, 6.6
 Bias, of electroabsorption modulators, 13.59
 Bidirectional reflectance distribution function (BRDF or BDRF), 1.4–1.6, 1.5*f*, 3.21
 Bidirectional scatter distribution function (BSDF), 1.6–1.7
 Bidirectional scatter distribution function (BSDF) scatterometers, 1.8, 1.8*f*
 Bidirectional transmission distribution function (BTDF), 1.6
 Bimorph mirrors, 5.37, 5.37*f*, 50.4, 50.4*f*, 50.6*f*
 Binary digits, in OTDM networks, 20.8
 Binary units, for electro-optic modulators, 7.27
 Biphas coding, 20.9, 20.9*f*
 Birefringence, 6.17, 8.19, 11.17
 Birefringent diffraction bandshapes, 6.13, 6.14, 6.14*f*
 Birefringent fibers, 16.5–16.6
 Birefringent phased array deflectors, 6.29
 Birefringent tangential phase matching, 6.25
 Bit error rate (BER), for fiber optic communications, 15.1–15.5, 15.3*f*, 15.4*f*, 15.8, 15.13, 15.15, 15.17
 Bit error ratio (BER), DQPSK and, 21.34
 Bit rate, for optical fibers, 9.12
 Blackman window function, 46.9, 46.10*f*
 Bloch waves, 11.22
 Blockage defects, in polycapillary x-ray optics, 53.4, 53.5, 53.5*f*
 Boltzmann population factor, 3.14, 3.20
 Bormann transmission, 43.3, 43.3*f*
 Bormann triangle, 63.27
 Born approximation, 27.2, 63.4, 63.5
 Boron neutron capture therapy (BNCT), 53.19
 Boron [$^{10}\text{B}(n, \alpha)$] reaction, 63.31
 Boundary quality, of multilayers, 41.5–41.7, 41.7*t*
 Bragg angle, 6.9, 30.5, 39.3
 Bragg cells (wideband), in acousto-optic devices, 6.27, 6.30, 6.31*t*
 Bragg condition, 6.32, 42.3*f*, 42.4, 42.10, 42.11*f*
 Bragg diffraction:
 and acousto-optic modulators, 6.4, 6.6, 6.7, 6.14
 and brightness of x-ray tube sources, 54.16
 far and near, 6.8–6.9, 6.12
 in neutron optics, 63.23
 order of, 20.14
 and phase matching equations, 6.11
 of x-rays, 42.2
 Bragg fibers, 11.4
 Bragg geometry, for crystal monochromators, 39.2–39.4, 39.2*f*, 39.3*f*, 39.6
 Bragg gratings:
 and DBR lasers, 20.14
 fiber, 17.1–17.9
 applications, 17.8–17.9, 17.8*f*
 chirped, 21.22–21.23, 21.23*f*, 21.25–21.26, 21.25*f*
 fabrication, 17.4–17.8, 17.5*f*–17.7*f*
 and fiber lasers, 25.8, 25.16, 25.18, 25.30, 25.31
 long-period gratings vs., 24.9, 24.11
 photosensitivity, 17.2–17.3
 properties of, 17.3–17.4, 17.4*f*
 sensors based on, 24.5–24.8, 24.6*f*–24.7*f*
 volume, 25.29, 25.30
 Bragg limit, 6.8
 Bragg mirrors, 13.28, 13.44
 Bragg planes, of crystal monochromators, 39.1, 39.2, 39.5
 Bragg reflection:
 of crystals, 40.9
 in interferometers, 63.26, 63.27
 and linear polarization, 43.2–43.4, 43.3*f*, 43.4*f*, 43.6, 43.8
 and liquid crystals, 8.10
 in monochromators, 39.1, 63.24
 and multilayers, 41.2
 and x-ray absorption spectroscopy, 30.2, 30.4
 Bragg reflection monochromator, 39.4, 39.5
 Bragg reflector lasers, 13.7, 13.28–13.29, 13.29*f*
 Bragg reflectors, 13.45
 Bragg scattering cross section, 63.18
 Bragg transmission phase retarders, 43.6
 Bragg wavelength, 17.3, 20.15, 24.7
 Bragg-Brentano powder diffractometer, 28.1, 28.2*f*, 28.3, 28.3*f*
 Bragg-Fresnel lenses, 40.9–40.10, 40.9*f*
 Bragg's law, 26.8, 28.3, 30.1, 39.1, 39.4, 43.7, 63.14, 63.16, 63.24
 Bragg-symmetric crystals, 35.3
 Braking radiation (*see* Bremsstrahlung radiation)
 Bremsstrahlung photons, 63.12
 Bremsstrahlung radiation, 31.3
 continuous, 54.4–54.6, 54.5*f*
 from laser-generated plasmas, 56.1, 56.8–56.10
 from pinch plasma sources, 57.1
 and x-ray fluorescence, 29.3, 29.5, 29.6, 29.11

- Brewster angle, in VUV and x-ray region, **41.9**
- Bridge fiber method, of mode matching, **25.17**, **25.17f**
- Brightness:
 of synchrotron radiation, **55.1**, **55.9**
 x-ray tube sources, **54.11–54.15**, **54.13f–54.15f**
- Brillouin frequency shift, **10.8**
- Brillouin scattering:
 backward, **11.25**
 forward, **11.25**, **11.26**
 photonic crystal fibers, **11.25–11.26**, **11.25f**
 stimulated, **10.1**, **10.7–10.9**, **15.8**, **21.20**, **25.6**
- Brillouin zones, **11.9**
- Broadband transmission, **3.23–3.24**, **3.25f–3.26f**
- Broadening:
 Doppler, **3.14**, **5.32**, **56.5**
 homogeneous, **20.1**
 Lorentzian, **3.14**
 pressure, **56.5–56.6**
 spectral, **56.6–56.8**
- Bulk electro-optic modulators, **7.3**, **7.16–7.28**
 amplitude modulation, **7.22–7.24**, **7.23f**, **7.24f**
 frequency modulation, **7.24–7.25**, **7.25f**
 phase modulation, **7.18–7.20**
 polarization modulation (dynamic retardation), **7.20–7.22**, **7.20f**, **7.21f**
 scanners, **7.26–7.28**, **7.26f–7.28f**
- Bulk lasers, **25.5–25.6**
- Bunches, electron, **55.17**
- Buried heterostructure (BH) lasers, **13.5**
- Cable television (CATV), **9.16–9.17**
- Calamitic liquid crystals (LCs), **8.4**, **8.5**, **8.9**, **8.11f**
- Calar Alto telescope, **5.27**
- Cameras:
 Anger, **63.33**
 gamma, **32.2**
- Canada-France-Hawaii Telescope, **5.23**
- Capillary discharge devices, **57.3**, **57.3t**
- Capillary optics, **28.5** (*See also* Monocapillary x-ray optics; Polycapillary optics)
- Carbon dioxide lasers, **12.3t**, **12.9**, **12.13**
- Carbon nanotubes (CNTs), **54.10–54.11**
- Carrier sense multiple access with collision detection (CSMA/CD), **23.7**
- Carrier-suppressed return-to-zero (CSRZ) formats, in WDM networks, **21.30**, **21.31f**, **21.36t**, **21.37f**
- Carrier-to-noise ratio (CNR), for optical fibers, **9.15–9.16**, **9.16f**
- Cassegrain telescopes, **44.4**
- Catadioptric lens systems, **35.1**
- Cathodes, as x-ray tube sources, **54.10–54.11**
- Catoptric lens systems, **35.1**
- Cauchy equations, for liquid crystals, **8.20**, **8.21**
- Čerenkov effects, **11.24**
- Chalcogenides:
 in fiber lasers, **25.27t**, **25.29**
 as infrared fibers, **12.2**, **12.2f**, **12.3t**, **12.6**, **12.7**, **12.7f**, **12.13**
- Chandra X-Ray Observatory, **33.2–33.4**, **33.3t**, **44.4**, **44.10**, **47.1**, **47.4f**, **47.5**, **47.10**, **64.7**
- Channel electron multipliers (CEMs), **60.6–60.7**
- Channel power equalization, for EDFAs, **21.41**, **21.42f**
- Characteristic radiation:
 Bremsstrahlung radiation as, **56.8–56.10**
 from laser-generated plasmas, **56.2–56.10**
 recombination radiation as, **56.10**
 spectral lines as, **56.2–56.8**
 from x-ray tube sources, **54.6–54.8**, **54.7f**
- Charge-coupled devices (CCDs):
 in adaptive optics, **5.21**
 as x-ray detectors, **60.7**, **60.8**, **60.10t**
 x-ray imaging detectors in, **61.7–61.8**, **61.8f**
- Charge-injected devices (CIDs), **60.7**, **60.10t**
- Chebyshev polynomials, **46.6**
- Chemical vapor deposition (CVD), **25.26**
- Chip screening, of SOAs, **19.17**, **19.17f**
- Chiral liquid crystals (LCs), **8.8–8.13**, **8.11f**
- Chirally coupled core (CCC) fibers, **25.2**, **25.19f**, **25.21–25.22**
- Chirp and chirping, **20.1**
 of electro-absorption modulators, **13.58**
 of fiber Bragg gratings, **17.7**
 frequency, **13.1**, **13.17–13.18**, **13.17f**
 of lasers, **9.8**
 and optical fibers, **10.3**, **10.11**
 of solitons, **22.3**
- Chirped fiber Bragg gratings (FBGs), **21.22–21.23**, **21.23f**, **21.25–21.26**, **21.25f**
- Chirped pulse amplification (CPA), **25.2**, **25.32**, **25.33**

- Cholesteric liquid crystal display (Ch-LCD), 8.32
- Chromatic dispersion:
and fiber optic communication links, 15.9, 15.10
in WDM networks, 21.14–21.16, 21.15f, 21.16f
- Chromium ions (Cr^{3+}):
absorption and photoluminescence of, 2.19–2.21, 2.19f, 2.20f
optical spectroscopy of, 2.9–2.10, 2.10f
polarization spectroscopy of, 2.21–2.22, 2.23f
- Circuits, optical tank, 20.21, 20.21f
- Circuit-switched networks, 21.7–21.10, 21.8f–21.9f
- Circular gratings, 40.1
- Circular polarization:
analyzers for, 43.6–43.8, 43.7f
phase plates for, 43.5f
and synchrotron radiation, 55.6–55.7
- Circulators, for networking, 18.3, 18.3f, 18.10
- Circulators, optical: fiber Bragg gratings, 17.8f, 17.9
- Circumferential coordinates, 45.7
- Circumferential slope errors, 45.8
- Cladding:
defined, 25.2
photonic crystal fibers in, 11.7–11.11, 11.8f–11.10f
- Cleaving, of photonic crystal fibers, 11.26
- Climate change, global, 3.43–3.45, 3.44f
- Clock recovery, in OTDM networks, 20.21–20.22, 20.21f
- Coarse wavelength division multiplexing (CWDM) systems, 19.27
- Coatings:
antireflection, 19.8, 19.8f, 19.20
for infrared optical fibers, 12.4, 12.7, 12.9
reflective (multilayers), 41.1–41.10
and calculation of multilayer properties, 41.3–41.4
for diffractive imaging, 41.9–41.10, 41.10f
fabrication methods and performance of, 41.4–41.9, 41.5f, 41.6t, 41.7t, 41.8f
properties of, 41.1–41.3, 41.2f
- Coblentz sphere, 1.10, 1.10f, 1.11
- Code mark inversion (CMI), 20.9
- Code V (program), 35.1
- Coding, in OTDM networks, 20.9–20.10, 20.9f
- Coherence:
beam spatial, 58.4
in coherent x-ray optics, 27.5
in long wavelength limit, 55.17–55.20, 55.18f, 55.19f
mutual coherence function (MCF), 4.4, 4.7, 4.10
temporal, 55.17–55.18, 55.18f
transverse, 55.18–55.20, 55.19f
transverse spatial, 55.16
- Coherence collapse, from laser diodes, 13.22–13.23
- Coherence diameter, Fried's, 5.2, 5.7, 5.9–5.10
- Coherence length:
atmospheric, 4.7–4.10
Fried's, 4.8
and imaging through turbulence, 4.8–4.10, 4.8f, 4.9f
spatial, 27.2
- Coherent area, of atmosphere, 4.16
- Coherent diffraction microscopy, 27.4–27.5, 27.4f, 27.5f
- Coherent Doppler LIDAR, 3.38, 3.39f
- Coherent scattering, 26.7, 63.7, 63.8
- Coherent x-ray optics, 27.1–27.5, 27.3f–27.5f
- Coiling, of LMA fibers, 25.18
- Cold cathode fluorescent lamps (CCFLs), 8.30, 8.30f
- Cold-cathode field emission, 54.10–54.11
- Collimating polycapillary optics, 53.14, 53.14f
- Collimating single crystal diffraction, 53.12, 53.12f
- Collimation and collimators:
anisotropic acoustic beam, 6.25
neutron, 63.15–63.16
in neutron and x-ray optics, 26.11, 26.11f
in polycapillary x-ray optics, 53.8–53.9, 53.8f–53.9f
with refractive x-ray lenses, 37.8
Soller, 28.2, 28.2f, 28.3
in SPECT imaging, 32.3
- Collinear beam acousto-optic tunable filters (CBAOTFs), 6.43, 6.43f, 6.45t
- Collision length, for solitons, 22.9–22.10, 22.15
- Coma, of grazing incidence telescopes, 44.9, 44.10

Communication networks and systems:

- fiber-optic standards for, 23.1–23.8
 - ATM/SONET, 23.6
 - ESCON, 23.1–23.2, 23.2f
 - Ethernet, 23.7
 - DDI, 23.2–23.3, 23.3f
 - Fibre Channel standard, 23.4, 23.5f, 23.5t
 - InfiniBand, 23.8, 23.8t
- optical fibers in, 9.3–9.17
 - analog transmission, 9.15–9.17
 - bit rate, 9.12
 - distance limits, 9.12–9.13
 - fiber for, 9.4–9.7, 9.5f, 9.6f
 - fiber-optic networks, 9.14–9.15
 - optical amplifiers, 9.13–9.14
 - photodetectors, 9.8
 - receiver sensitivity, 9.8–9.11
 - repeater spacing, 9.12–9.13
 - technology, 9.4–9.8
 - transmitting sources, 9.7–9.8
- optical time-division multiplexed, 20.1–20.25
 - analog to digital conversion, 20.8, 20.8f
 - binary digits and line coding in, 20.8–20.10, 20.9f
 - device technology, 20.12–20.24
 - history of, 20.3
 - interleaving in, 20.6–20.7, 20.7f
 - modulation, 20.17–20.20, 20.17f, 20.19f, 20.20f
 - multiplexing and demultiplexing, 20.3–20.12, 20.7f, 20.22, 20.23f
 - optical clock recovery, 20.21–20.22, 20.21f
 - sampling, 20.4–20.6, 20.5f, 20.6f
 - serial vs. parallel, 20.12, 20.13f
 - timing recovery, 20.10–20.12, 20.10f, 20.11f
 - transmitters, 20.12–20.17, 20.14f–20.16f
 - ultrahigh-speed, 20.23–20.24, 20.24f
- solitons in, 22.1–22.17
 - classical solitons, 22.2–22.4, 22.2f–22.4f
 - design of transmission systems, 22.5–22.7
 - dispersion-managed solitons, 22.12–22.15, 22.13f, 22.14f
 - frequency-guiding filters, 22.7–22.9
 - wavelength division multiplexing, 22.9–22.12

Communication networks and systems (*Cont.*):

- wavelength-division multiplexed, 21.1–21.44
 - carrier-suppressed return-to-zero and duobinary, 21.30–21.33, 21.31f, 21.32f
 - chromatic dispersion in, 21.14–21.16, 21.15f, 21.16f
 - circuit and packet switching in, 21.7–21.11, 21.8f–21.11f
 - dispersion and nonlinearities of, 21.16–21.26, 21.17f–21.21f, 21.23f–21.27f
 - DPSK and DQSK, 21.33–21.36, 21.33f–21.35f, 21.36t, 21.37t
 - fiber attenuation and optical power loss, 21.13–21.14
 - fiber bandwidth, 21.2–21.3, 21.2f
 - fiber system impairments, 21.13–21.26
 - history of, 21.1–21.2
 - network reconfigurability, 21.12–21.13, 21.12f, 21.13f
 - optical amplifiers in, 21.37–21.44, 21.37f, 21.38f–21.42f
 - optical modulation formats, 21.27–21.36, 21.28f–21.30f
 - point-to-point links, 21.4
 - in real systems, 21.3–21.4, 21.3f, 21.4f
 - star, ring, and mesh topologies, 21.5–21.7, 21.5f–21.7f
 - wavelength-routed networks, 21.5, 21.5f
 - WDM dispersion managed soliton transmission, 22.15–22.17
- Commutators, 20.1, 20.7f
- Compensators, Babinet-Soleil, 7.22
- Complementary metal oxide semiconductors (CMOS) signal processing, 62.5
- Complex index of refraction, for x-ray optics, 48.1
- Compton radiation sources, 55.2–55.3, 55.3t
- Compton scattering, 59.1
 - and circular polarization, 43.6
 - inverse, 59.1
 - and microfocus x-ray fluorescence, 29.5, 29.8f
 - and polycapillary x-ray optics, 53.3, 53.15, 53.18
 - and refractive x-ray lenses, 37.5, 37.7
 - and x-ray attenuation, 31.2
 - and x-ray optics, 26.7, 36.1
- Computed tomography (CT), 31.5–31.7, 31.6f, 31.7f
- Condensing monocapillary x-ray optics, 52.6

- Cone effect, of laser beacons, 5.27
- Cone-beam computed tomography (CT), 31.7, 31.7f
- Configurational relaxation, in solids, 2.14–2.17, 2.15f–2.18f
- Confinement:
- in double heterostructure laser diodes, 13.4–13.6, 13.4f, 13.6f
 - and photonic crystal fibers, 11.22
 - in semiconductor optical amplifiers, 19.6
- Connector losses, in fiber optic communication, 15.7, 15.8t
- Constellation-X Observatory, 33.4
- Continuous readout x-ray detectors, 61.2
- Continuous-wave (CW) lasers, 25.4, 25.5, 25.5f, 25.7
- diode lasers, 13.49
 - dye lasers, 5.32, 5.33f, 5.34
- Contrast ratio, for fiber optic modulation, 13.57
- Contrast-to-noise ration (CNR), of x-ray detectors, 61.3
- Controlled-drift x-ray detectors, 62.5
- Converging neutron guides, 63.17
- Coordinate-measuring machines (CMMs), 46.2
- Coordinates, circumferential, 45.7
- Copper distributed data interface (CDDI), 23.3
- Core (term), 25.2
- Core drilling, of fiber lasers, 25.26–25.27
- Core-cladding index difference, of photonic crystal fibers, 11.12–11.17, 11.12f–11.17f
- Cornell High Energy Synchrotron Sources (CHESS), 52.3, 52.5
- Coulomb explosion, 2.5, 2.6
- Coulomb fields, of ions, 56.8
- Coulomb interactions, of 3d electrons, 2.9
- Coulomb repulsion, 2.5
- Coupled quantum wells (CQWs), 13.58
- Coupled-wave-theory (CWT), for multilayer Laue lenses, 42.4, 42.12
- Couplers and coupling:
- directional, 18.2, 18.3, 18.3f, 18.9, 18.9f
 - evanescent, 25.11
 - fiber-based, 16.1–16.6, 16.2f, 16.5f
 - nonfused fiber, 25.10
 - vertical, 13.60
 - wavelength-selective, 14.2
- Coupling coefficient, of DBR lasers, 13.28
- Coupling loss, of fiber-optic components, 18.1
- Covariance mapping, 2.5
- Cross-gain modulation (XGM), 19.12, 19.13f, 19.27, 19.29–19.30, 19.29f, 19.32, 19.35–19.36
- Crossover frequency, liquid crystals and, 8.16
- Cross-phase modulation (XPM):
- in optical fibers, 10.3–10.4
 - and SOAs, 19.13, 19.30–19.32, 19.31f, 19.33f, 19.35–19.36
 - and solitons, 22.5, 22.13–22.15
 - in WDM networks, 21.19
- Crucifix images, 49.4, 49.4f
- Cryogenic x-ray detectors, 60.9, 60.9t, 60.10t
- Crystal diffraction:
- and EDXRF, 29.6–29.7, 29.8f–29.9f
 - single, 53.12–53.14, 53.12f–53.13f
 - and WDXRF, 29.2
- Crystal interferometers, 63.26–63.27, 63.27f
- Crystal monochromators:
- and bent crystals, 39.1–39.6, 39.2t, 39.3f, 39.5f–39.6f
 - in neutron optics, 63.23–63.25
- Crystal optics:
- and electro-optic modulators, 7.3–7.4, 7.4f, 7.8t–7.10t
 - and the index ellipsoid, 7.3–7.7, 7.4f–7.6f, 7.8f–7.10f
 - polarizing, 43.1–43.8, 43.3f–43.5f, 43.7f
 - and x-ray diffraction, 28.3–28.4, 28.4f
- Crystalline infrared fibers, 12.2t, 12.3t, 12.7–12.10, 12.8f, 12.10f
- Crystals:
- bent, 39.5–39.6, 39.5f–39.6f
 - Bragg-symmetric, 35.3
 - Darwin width of, 63.24, 63.26
 - doubly curved, 29.6–29.7, 29.8f
 - in electro-optic modulators, 7.33–7.34
 - Fankuchen-cut, 63.25
 - mosaic, 39.2
 - organic, 7.33–7.34
 - SHADOW code for, 35.2
 - [See also Liquid crystals; Photonic crystal fibers (PCFs)]
- Current:
- dark, 13.69, 13.73
 - saturation, 13.69
 - trap-assisted thermal generation, 13.69
- Current confinement, in VCSELs, 13.45, 13.45f
- Curvature of best focal surface, 45.5

- Damage, optical, 13.54
- Damage threshold, of fiber lasers, 25.7
- Dark current, 13.69, 13.73
- Darwin widths, 43.6, 63.24, 63.26
- Data communication systems, 15.1–15.2
- Data transmission formats, 20.9–20.10, 20.9*f*
- dc Kerr effect, 7.11
- De Broglie neutron waves, 63.3, 63.5
- Debye-Waller factor, 41.4
- Decentering errors, for grazing incidence optics, 45.6–45.7
- Decommutator, 20.1, 20.7*f*
- Deep saturation regime, for fiber amplifiers, 14.4
- Deflectors, 6.22–6.31, 6.23*t*
 - anisotropic acoustic beam collimation by, 6.25
 - birefringent phased array, 6.29
 - high-resolution, 6.29, 6.29*t*
 - isotropic AO diffraction by, 6.23–6.24, 6.24*f*
 - phase array beam steering by, 6.27–6.29, 6.28*f*
 - resolution of, 6.25
 - tangential phase matching by, 6.25–6.27, 6.26*f*
- Defocus errors, for grazing incidence optics, 45.6
- Deformable mirrors, 5.4, 5.4*f*, 5.37–5.38, 5.37*f*, 5.38*f*
- Degree of modulation, for electro-optic modulators, 7.35–7.36
- Dektak stylus profiler, 46.2
- Delay-line technique, 60.5
- Demultiplexers and demultiplexing, 20.1
 - for networking, 18.4, 18.10–18.11, 18.10*f*, 18.11*f*
 - in OTDM networks, 20.7–20.8, 20.7*f*, 20.22, 20.23*f*
 - terahertz asymmetric optical, 20.22
- Dense wavelength division multiplexing (DWDM):
 - and AOTFs, 6.43, 6.44
 - and optical fiber amplifiers, 14.1
 - and SOAs, 19.25–19.27, 19.25*f*, 19.26*f*
- Density of states (DOS), for photonic crystal fibers, 11.8, 11.9*f*
- Depletion layer, of semiconductor detectors, 60.5
- Depletion region, of pin photodiodes, 13.64
- Deposition:
 - modified chemical vapor, 25.2, 25.21, 25.26, 25.28
 - outside vapor, 25.2, 25.26
 - physical vapor, 61.6
 - vapor axial, 25.3, 25.26
- Deposition, chemical vapor, 25.26
- Depth of focus (DOF), of multilayer Laue lenses, 42.17, 42.17*f*
- Depth-of-amplitude modulation, 7.22
- Depth-of-phase modulation, 7.19
- Despace errors, for grazing incidence optics, 45.6
- Destructive ports, of M-Z interferometers, 21.34
- Detected point spread function (DPSF), 44.14, 44.15*f*
- Detection and detectors:
 - activated phosphor, 60.7–60.8
 - collision, 23.7
 - direct-conversion flat panel, 61.4*f*, 61.6–61.7, 61.6*t*, 61.7*f*
 - energy dispersive, 62.2–62.4
 - for fiber optic systems, 13.2–13.3, 13.63–13.73
 - avalanche photodiodes, 13.71–13.73
 - MSM detectors, 13.73
 - pin diodes, 13.63–13.71, 13.64*f*, 13.66*f*, 13.66*t*, 13.68*f*
 - Schottky photodiodes, 13.73
 - heterodyne, 3.34
 - light detection and ranging (LIDAR), 3.38–3.39, 3.38*f*, 25.25
 - for medical imaging, 31.4, 31.4*f*, 61.2
 - MSM photoconductive, 13.63, 13.73
 - in neutron optics, 63.31–63.34
 - photodetectors, 9.8
 - semiconductor, 60.5
 - solid state, 63.33–63.34
 - x-ray, 31.4*f*
 - controlled-drift, 62.5
 - cryogenic, 60.8–60.9, 60.9*t*, 60.10*t*
 - film, 60.8, 60.9*t*, 60.10*t*
 - ionization, 60.3–60.7, 60.9*t*, 60.10*t*
 - scintillation, 60.7–60.8, 60.9*t*, 60.10*t*
 - for x-ray imaging, 61.1–61.8
 - CCD detectors, 61.7–61.8, 61.8*f*
 - flat panel detectors, 61.3–61.7, 61.4*f*, 61.6*t*, 61.7*f*
 - geometries for and classifications of, 61.1–61.3, 61.2*f*

- Detective quantum efficiency (DQE), of x-ray detectors, **61.2**, **61.3**, **61.5–61.7**
- Detrending, in x-ray mirror metrology, **46.6–46.7**, **46.7f**
- Detuning, **13.29**, **30.3**
- Dichroism:
 defined, **25.2**
 magnetic circular, **55.7–55.9**, **55.7f**
- Dielectric properties, of liquid crystals, **8.14–8.18**, **8.15f**, **8.18f**
- Differential group delay (DGD), **21.17**
- Differential phase shift keying (DPSK), **19.27**, **21.33–21.34**, **21.36**, **21.36t**, **21.37t**
- Differential quadrature phase-shift-keying (DQPSK), **21.34–21.36**, **21.35f**, **21.36t**, **21.37t**
- Differential quantum efficiency, **13.9**
- Diffraction:
 acousto-optic, **6.4**, **6.9**
 anisotropic, **6.10**
 and birefringent diffraction bandshapes, **6.13**, **6.14**, **6.14f**
- Bragg
 and acousto-optic modulators, **6.4**, **6.6**, **6.7**, **6.14**
 and brightness of x-ray tube sources, **54.16**
 far and near, **6.8–6.9**, **6.12**
 in neutron optics, **63.23**
 order of, **20.14**
 and phase matching equations, **6.11**
 of x-rays, **42.2**
- by crystals, **39.2–39.3**
 and energy-dispersive x-ray fluorescence, **29.6–29.7**, **29.8f–29.9f**
- Fraunhofer, **42.2**, **63.25**
- Fresnel, **27.2**, **27.4**, **40.9**, **63.25**
- by gratings and monochromators, **38.1–38.3**, **38.2f**
- isotropic
 and acousto-optic interactions, **6.9–6.10**, **6.10f**, **6.13**, **6.13f**
 by deflectors, **6.23–6.24**, **6.24f**
 and isotropic diffraction bandshape, **6.13**, **6.13f**
- kinematic theory of, **63.5**
- lysozyme, **53.12–53.14**, **53.12f**, **53.13f**
- near- and far-field, **27.2–27.3**, **27.3f**
- in neutron optics, **63.3**, **63.25**
- order of, **18.6**
- Diffraction (*Cont.*):
 powder, **53.14**, **53.14f**
 single crystal, **53.12–53.14**, **53.12f–53.13f**
 and wavelength-dispersive x-ray fluorescence, **29.2**
 x-ray, **28.5–28.6**, **28.6f**
 and x-ray optics, **26.7**, **26.8**
- Diffraction efficiencies, of zone plates, **40.4–40.8**, **40.5f**, **40.7f**, **40.8f**
- Diffraction geometry, tilted, **42.4**
- Diffraction imaging, **41.9–41.10**, **41.10f**
- Diffraction optics, **26.8**, **26.8f**, **26.9f**
- Diffraction scattering, **64.2–64.3**
- Diffraction meters, **28.1–28.3**, **28.2f**, **28.3f**
- Diffuse scattering, **63.4**
- Digital, analog conversion to, **20.4**, **20.8**, **20.8f**
- Digital displays, in medical imaging, **31.8–31.9**, **31.8f–31.10f**
- Digital modulation, **6.33**
- Digital on-off-keying receivers, **9.9–9.11**
- Digital tomosynthesis, **31.7–31.8**
- Digital Video Broadcasters Measurement
 Group standards, **15.4**
- Diode pumps, **25.12–25.13**
- Diodes:
 laser, **13.3–13.24**
 double heterostructure, **13.3–13.8**, **13.3f**, **13.4f**, **13.6f**, **13.7f**
 noise characteristics of, **13.18–13.24**, **13.19f–13.21f**
 operating characteristics of, **13.8–13.13**, **13.10f**
 transient response of, **13.13–13.18**, **13.14f**, **13.16f**, **13.17f**
- light-emitting, **13.1**, **13.36–13.42**, **13.38f**
 edge-emitting, **13.40**
 operating characteristics of, **13.40–13.42**, **13.40f**, **13.41–13.42**
 and optical fibers, **9.7**
 surface-emitting, **13.38–13.40**, **13.38f**
 and transmissive TFT LCDs, **8.29–8.31**, **8.30f**
- pin, **13.2**, **13.63–13.71**, **13.66t**
 dark current, **13.69**
 geometry of, **13.64–13.65**, **13.64f**
 noise, **13.70–13.71**
 sensitivity, **13.65–13.66**, **13.66f**
 speed, **13.67–13.68**
 and untraveling-carrier (UTC) photodiodes, **13.68–13.69**, **13.68f**
 (See also Photodiodes)

- Dioptric lens systems, 35.1
- Direct modulation, 20.17–20.18, 20.17*f*
- Direct-conversion flat panel detectors, 61.4*f*, 61.6–61.7, 61.6*t*, 61.7*f*
- Directional couplers, for networking, 18.2, 18.3, 18.3*f*, 18.9, 18.9*f*
- Directionality (isolation), of fiber-optic components, 18.1
- Discrete electroabsorption modulators, 13.59
- Dispenser cathodes, 54.10
- Dispersion:
 - angular, 38.6
 - chromatic
 - and fiber optic communication links, 15.9, 15.10
 - in WDM networks, 21.14–21.16, 21.15*f*, 21.16*f*
 - in fiber optic communication links, 15.9–15.11, 15.10*f*
 - by gratings and monochromators, 38.6–38.7
 - group velocity, 11.18, 11.18*f*, 11.19*f*
 - in optical fibers, 9.5–9.7, 9.6*f*
 - theory of, 63.5
 - in WDM networks, 21.20–21.26, 21.21*f*
 - chromatic, 21.14–21.16, 21.15*f*, 21.16*f*
 - electronic solutions for, 21.26, 21.27*f*
 - fixed dispersion compensation, 21.22–21.23, 21.23*f*
 - tunable dispersion compensation, 21.23–21.26, 21.24*f*–21.26*f*
- Dispersion compensating fiber (DCF), 15.10, 21.22
- Dispersion length, of solitons, 22.3
- Dispersion optimizing fiber, 15.10
- Dispersion-managed (DM) solitons, 22.12–22.15, 22.13*f*, 22.14*f*
- Dispersion-shifted fibers, 9.7
- Displacement vectors, bulk modulators and, 7.18
- Displays, for medical imaging, 31.8–31.9
- Distance limits, of optical fibers, 9.12–9.13
- Distorted object approach, in coherent x-ray optics, 27.2–27.4
- Distortion(s):
 - in fiber optic communication links, 15.5–15.6, 15.6*f*
 - in optical fibers, 9.17
- Distributed Bragg reflector (DBR) lasers, 9.8, 13.7, 13.28–13.29, 13.29*f*, 20.14
- Distributed feedback (DFB) lasers:
 - in fiber optic systems, 13.7, 13.30–13.32, 13.30*f*, 13.31*f*
 - optical fibers for, 9.8
 - in OTDM communication networks, 20.14–20.15, 20.15*f*
- Distributed feedback (DFB) threshold, of fiber optic systems, 13.30–13.32
- Distributed Raman amplifiers (DRAs), 21.44
- Divalent rare-earth ions, 2.11
- Divided voltage method, for transreflective LCDs, 8.35, 8.35*f*
- Dopant profiling, for LMA fibers, 25.18
- Dopants, for fiber lasers, 25.22–25.26, 25.23*t*
- Doppler broadening, 3.14, 5.32, 56.5
- Doppler effect, 2.3, 6.9
- Doppler LIDAR systems, 3.38–3.39
- Doppler profiles, of CW lasers, 5.32, 5.34
- Doppler shifts, 10.7
- Doppler-dominated lineshapes, 3.23
- Double heterostructure laser diodes, 13.3–13.8, 13.3*f*, 13.4*f*, 13.6*f*, 13.7*f*
- Double heterostructure waveguides, 19.3*f*, 19.5
- Double-bounce Wolter mirrors, 52.4
- Doubly curved crystals (DCCs), 29.6–29.7, 29.8*f*
- Dragon (monochromator) systems, 38.3
- Drive circuitry, of LEDs, 13.42
- Drive power, of NPM AOTFs, 6.41
- Dual attach FDDI nodes, 23.3
- Dual frequency effect, on LCDs, 8.18
- Dual-cell-gap transreflective LCDs, 8.32–8.33, 8.33*f*
- Duobinary formats, for WDM networks, 21.32–21.33, 21.32*f*, 21.36, 21.36*t*, 21.37*t*
- Duplexers, for networking, 18.4, 18.4*f*, 18.10–18.11
- Dye-doped polymers, in electro-optic modulators, 7.34
- Dynamic Jahn-Teller effect, 2.9
- Dynamic range:
 - of fiber amplifiers, 14.3
 - of fiber optic communication links, 15.5–15.6, 15.6*f*
 - of wideband Bragg cells, 6.30
- Dynamic retardation, of modulators, 7.20–7.22, 7.20*f*, 7.21*f*
- Dynamical theory of diffraction, for crystals, 39.2

- Edge-coupled pin waveguides, **13.68**
- Edge-defined film-fed growth (EFG) technique, **12.9, 12.10f**
- Edge-emitting lasers, **13.3**
- Edge-emitting light-emitting diodes (E-LEDs), **13.36, 13.37, 13.40**
- EEGRAZE code, **44.13**
- Effective group refractive index, **13.13**
- Efficiency:
 - detective quantum, **61.2, 61.3, 61.5–61.7**
 - differential quantum, **13.9**
 - diffraction, **40.4–40.8, 40.5f, 40.7f, 40.8f**
 - external quantum, **13.9**
 - external slope, **13.9**
 - gain, **14.6**
 - internal quantum, **13.9, 13.10**
 - modulation, **7.36**
 - quantum, **13.65, 13.66**
 - quantum detection, **60.7, 61.2–61.3**
 - radiative quantum, **14.7**
- Eigenpolarization, **7.13–7.14, 7.14f**
- Einstein coefficient for spontaneous emission, **2.13**
- Einstein Observatory, **44.4, 44.10, 47.1, 47.5**
- Elastic constants, of liquid crystals, **8.22, 8.23f**
- Elastic scattering, in neutron optics, **63.5**
- Elasto-optic effect, **6.5–6.6**
- Electric dipole selection rules, **56.3**
- Electrical injection:
 - in laser diodes, **13.4, 13.5**
 - in VCSELs, **13.45, 13.45f**
- Electrical time domain multiplexed (ETDM)
 - transmission, **20.3, 20.25**
- Electroabsorption, **13.55–13.56, 13.56f**
 - applying fields in semiconductors, **13.58**
- Electroabsorption modulators (EAMs):
 - in fiber optic systems, **13.55–13.60, 13.56f, 13.57f**
 - in OTDM networks, **20.18, 20.20, 20.20f**
- Electroluminescence, of LEDs, **13.37**
- Electron beam steering, **54.11**
- Electron binding energies, **36.3t–36.6t**
- Electron excitation, WDXRF and, **29.2**
- Electron linacs, **63.12**
- Electron-beam lithography (EBL), **40.8**
- Electrons, **2.9, 54.12, 55.17, 56.2, 58.1**
- Electro-optic effect, **7.6–7.16**
 - and eigenpolarization/phase velocity indices of refraction, **7.13–7.16, 7.14f, 7.16f**
 - Jacobi method, **7.11–7.13**
 - linear, **7.7, 7.8t, 7.10**
 - and lithium niobate modulators, **13.49–13.51, 13.51f**
 - quadratic (Kerr), **7.9t–7.10t, 7.11**
- Electro-optic modulators, **7.1–7.39**
 - applications for, **7.36–7.39, 7.37f–7.38f**
 - bulk modulators, **7.16–7.28**
 - amplitude modulation, **7.22–7.24, 7.23f, 7.24f**
 - frequency modulation, **7.24–7.25, 7.25f**
 - phase modulation, **7.18–7.20**
 - polarization modulation (dynamic retardation), **7.20–7.22, 7.20f, 7.21f**
 - scanners, **7.26–7.28, 7.26f–7.28f**
 - crystal optics and the index ellipsoid, **7.3–7.7, 7.4f–7.6f, 7.8f–7.10f**
 - and electro-optic effect, **7.6–7.16**
 - eigenpolarization/phase velocity indices of refraction, **7.13–7.16, 7.14f, 7.16f**
 - Jacobi method, **7.11–7.13**
 - linear, **7.7, 7.8t, 7.10**
 - quadratic (Kerr), **7.9t–7.10t, 7.11**
 - and Euler angles, **7.39**
 - in fiber optic systems, **13.61**
 - geometries, **7.16–7.18, 7.17f**
 - light propagation in, **7.3**
 - longitudinal, **7.16, 7.17, 7.17f**
 - materials for, **7.33–7.34**
 - in OTDM networks, **20.18–20.19, 20.19f**
 - performance criteria for, **7.34–7.36**
 - polymer modulators, **13.55**
 - transverse modulators, **7.16, 7.17, 7.17f**
 - traveling wave modulators, **7.28–7.30, 7.29f**
 - waveguide modulators, **7.30–7.32, 7.31f–7.33f**
- Electro-optic sampling, **7.36–7.37, 7.37f–7.38f**
- Electrorefraction, **13.2**
- Electrorefractive modulators, **13.61–13.62**
- Elements (chemical), x-ray properties of, **36.1–36.9, 36.3t–36.9t**
- Elliptical reflectors, **64.3, 64.4, 64.4f**
- Emission:
 - amplified spontaneous, **9.13, 10.11, 14.3, 14.6**
 - cold-cathode field, **54.10–54.11**
 - molecular, **3.18, 3.20, 3.20f**
 - particle-induced x-ray, **29.4**
 - self-amplified spontaneous, **58.1**
 - spontaneous, **2.13, 19.3, 20.1**
 - stimulated, **20.2**

- Emission lines, K-shell and L-shell, **36.3t–36.8t**
- Endlessly single-mode photonic crystal fiber (ESM-PCF), **11.12, 11.13, 11.21, 11.21f**
- End-pumped schemes, for fiber lasers, **25.9–25.10, 25.28**
- Energy(-ies):
 atomic, **2.2–2.5, 2.4f**
 Auger, **36.3t, 36.9t**
 band gap, **19.2**
 conservation of, **6.9**
 electron binding, **36.3t–36.6t**
 equipartition, **22.6**
 Fermi level, **30.1**
 filtering of, **53.10**
 photon, **36.7t–36.8t**
 (See also *specific sources, e.g.: Fiber lasers*)
- Energy dispersive detectors (EDS), **62.2–62.4**
- Energy-dispersive x-ray fluorescence (EDXRF), **29.3–29.11**
 with doubly curved crystal diffraction, **29.6–29.7, 29.8f–29.9f**
 monocapillary micro-XRF, **29.4**
 polycapillary micro-XRF, **29.4–29.6, 29.5f, 29.6f**
 ultrahigh resolution, **29.9–29.11, 29.9f–29.11f**
- Entrance slits, of gratings and monochromators, **38.7**
- Environmental control and correction, for adaptive optics, **50.5, 50.6**
- Epitaxial growth, **13.2**
- Epithermal neutrons, **63.18**
- Equatorial divergence, x-ray diffraction and, **28.1**
- Equipartition energy, **22.6**
- Erbium-doped fiber amplifiers (EDFAs):
 energy levels, **14.4**
 fast power transients, **21.39–21.41, 21.39f, 21.40f**
 gain flattening, **14.6–14.7, 21.38–21.39, 21.39f**
 gain formation, **14.4–14.5, 14.5f**
 gain peaking, **21.38, 21.38f**
 noise, **14.6**
 pump wavelength options, **14.5–14.6**
 semiconductor amplifiers vs., **9.13, 9.14, 14.1, 14.2t**
 static gain dynamic and channel power equalization, **21.41, 21.41f–21.42f**
 in WDM networks, **21.2–21.3, 21.2f**
- Erbium-doped fibers, **25.23t, 25.24, 25.32, 25.33**
- Erbium-doped yttrium aluminum garnet (Er:YAG) lasers, **12.3t, 12.6, 12.13**
- Erbium/ytterbium-doped fiber amplifiers (EYDFAs), **14.2, 14.2t, 14.7–14.8**
- ESCON (Enterprise System Connection) standard, **23.1–23.2, 23.2f**
- ESO telescope, **5.35**
- Etch and regrowth fabrication, of electroabsorption modulators, **13.59–13.60**
- Ethernet standard, **23.7**
- Euler angles, **7.13, 7.39**
- Euler's constant, **56.9**
- European Synchrotron Radiation Facility (ESRF), **37.5, 37.8, 50.5, 50.6**
- European X-Ray Free-Electron Laser (XFEL), **58.1**
- Evanescent coupling, **25.11**
- Evanescent wave spectroscopy (EWS), **12.13**
- Evanescent-wave coupled pin waveguides, **13.68**
- Excitation, of electrons, **56.2**
- Excitation spectroscopy, **2.15f, 2.21**
- Excited state absorption (ESA), **14.6**
- Exit slits, of gratings and monochromators, **38.7**
- Expansion coefficients, in atmospheric optics, **4.20–4.22, 4.21t, 4.22f–4.23f**
- Extended Gordon-Haus effect, **22.11**
- Extended x-ray absorption fine structure (EXAFS), **30.2, 30.4**
- External circuits, noise from, **13.70**
- External mirrors, **13.32–13.33, 13.33f**
- External modulation, in OTDM networks, **20.18–20.20, 20.19f, 20.20f**
- External optical feedback, from laser diodes, **13.21–13.24**
- External quantum efficiency, **13.9**
- External slope efficiency, **13.9**
- Extinction ratio, **7.35, 15.13**
- Extreme ultraviolet explorer (EUV), **44.4, 44.5**
- Extreme ultraviolet (EUV) lasers, **58.2–58.4, 58.3f**
- Extreme ultraviolet lithography (EUV-L), **34.1–34.6**
 and EUV-interferometric lithography, **34.4–34.5, 34.5f**
 limitations of, **34.5–34.6, 34.5f, 34.6f**
 and multifoil optics, **48.1**
 and multilayers, **41.5, 41.7, 41.8**
 in semiconductor industry, **34.1–34.2, 34.2t**
 technology for, **34.2–34.5, 34.3f, 34.4f**

- Extreme ultraviolet (EUV) region,
 Schwarzschild objective for, 51.3
 Extreme ultraviolet-interferometric lithography
 (EUV-IL), 34.4–34.5, 34.5f
 Extrinsic Fabry-Perot interferometric (EFPI)
 sensors, 24.2–24.4, 24.2f, 24.3f
 Extrusion, of fiber lasers, 25.27
 Eye degradations, of optical fiber receivers, 9.11
 Eye openings, optical fiber receivers and, 9.11
 Eye safety, 15.1
- Fabry-Perot cavities, 13.65, 19.19, 20.21, 20.21f,
 24.2
 Fabry-Perot filters, 13.33, 22.8
 Fabry-Perot interferometers, 3.36, 24.2–24.5,
 24.2f–24.4f
 Fabry-Perot lasers, 9.11, 13.12–13.13, 13.29,
 15.11, 15.14
 Fabry-Perot semiconductor lasers, 20.13–20.14,
 20.14f
 Fankuchen-cut crystals, 63.25
 Far Bragg diffraction acousto-optic interaction
 and, 6.8
 Far field, 4.10
 Faraday rotators, 18.7, 18.7f, 18.10
 Far-field diffraction, 27.2–27.3, 27.3f
 FASCODE program, 3.23, 3.24f
 Fast power transients, for EDFAs, 21.39–21.41,
 21.45f
 FDDI (fiber distributed data interface) standard,
 13.41, 23.2–23.3, 23.3f
 Feedback:
 in fiber optic systems, 13.30–13.32
 from laser diodes, 13.21–13.24
 from SOAs, 19.8, 19.8f
 Fermi choppers, 63.14
 Fermi level, of energy, 30.1
 Fermi pseudopotential, 63.4, 63.5
 Ferroelectric smectic phase, of liquid crystals,
 8.12, 8.12f
 Fiber (material), for optical fibers, 9.4–9.7
 Fiber amplifiers, 14.1–14.11
 categories and features of, 14.1–14.2, 14.2t
 erbium-doped
 energy levels, 14.4
 fast power transients, 21.39–21.41, 21.39f,
 21.40f
 gain flattening, 14.6–14.7, 21.38–21.39,
 21.39f
 gain formation, 14.4–14.5, 14.5f
 Fiber amplifiers, erbium-doped (*Cont.*):
 gain peaking, 21.38, 21.38f
 noise, 14.6
 pump wavelength options, 14.5–14.6
 semiconductor amplifiers vs., 9.13, 9.14,
 14.1, 14.2t
 static gain dynamic and channel power
 equalization, 21.41, 21.41f–21.42f
 in WDM networks, 21.2–21.3, 21.2f
 erbium/ytterbium-doped, 14.2, 14.2t, 14.7–14.8
 infrared fibers for, 12.3t
 parametric, 14.10–14.11
 praseodymium-doped fiber amplifiers
 (PDFAs), 14.7
 Raman, 14.8–14.9, 14.8f, 14.9, 14.10f
 rare-earth-doped, 14.2–14.4, 14.3f
 semiconductors vs., 9.13–9.14
 ytterbium-doped, 14.7
 Fiber attenuation, optical power loss and,
 21.13–21.14
 Fiber bandwidth, of WDM networks,
 21.2–21.3, 21.2f
 Fiber Bragg gratings (FBGs), 17.1–17.9
 applications, 17.8–17.9, 17.8f
 chirped, 21.22–21.23, 21.23f, 21.25–21.26,
 21.25f
 fabrication, 17.4–17.8, 17.5f–17.7f
 and fiber lasers, 25.8, 25.16, 25.18, 25.30, 25.31
 long-period gratings vs., 24.9, 24.11
 photosensitivity of, 17.2–17.3
 properties of, 17.3–17.4, 17.4f
 sensors based on, 24.5–24.8, 24.6f–24.7f
 Fiber lasers, 25.1–25.33
 architectures, 25.9–25.18, 25.19f
 all-fiber monolithic systems, 25.16–25.18,
 25.16f
 free space, 25.13–25.15, 25.14f, 25.15f
 pumping techniques, 25.9–25.13, 25.11f,
 25.12f
 bulk lasers vs., 25.5–25.6
 dopants for, 25.22–25.26, 25.23t
 fabrication of, 25.26–25.29, 25.27t
 growth of, 25.5, 25.5f
 history of, 25.3–25.4, 25.4f
 infrared fibers for, 12.3t
 limitations of, 25.6–25.7
 LMA designs for, 25.18–25.22, 25.19f
 operation of, 25.7–25.8
 spectral and temporal modalities of,
 25.29–25.33

- Fiber length, for rare-earth-doped amplifiers, 14.2–14.3, 14.3f
 - Fiber optic amplifiers, 14.1–14.11
 - categories and features of, 14.1–14.2, 14.2t
 - erbium-doped
 - energy levels, 14.4
 - fast power transients, 21.39–21.41, 21.39f, 21.40f
 - gain flattening, 14.6–14.7, 21.38–21.39, 21.39f
 - gain formation, 14.4–14.5, 14.5f
 - gain peaking, 21.38, 21.38f
 - noise, 14.6
 - pump wavelength options, 14.5–14.6
 - semiconductor amplifiers vs., 9.13, 9.14, 14.1, 14.2t
 - static gain dynamic and channel power equalization, 21.41, 21.41f–21.42f
 - in WDM networks, 21.2–21.3, 21.2f
 - erbium/ytterbium-doped, 14.7–14.8
 - parametric, 14.10–14.11
 - praseodymium-doped, 14.7
 - Raman fiber, 14.8–14.9, 14.8f, 14.10f
 - rare-earth-doped, 14.2–14.4, 14.3f
 - ytterbium-doped, 14.7
 - Fiber optic chemical sensors, 12.3t
 - Fiber optic communication links, 15.1–15.17
 - distortions and dynamics range of, 15.5–15.6, 15.6f
 - figures of merit for, 15.2–15.6, 15.3f, 15.4f
 - link budget analysis for, 15.6–15.17
 - extinction ratio, 15.13
 - installation loss, 15.6–15.7, 15.8t
 - optical power penalties, 15.8–15.17, 15.10f
 - Fiber optic communication standards, 23.1–23.8
 - ATM/SONET, 23.6
 - ESCON, 23.1–23.2, 23.2f
 - Ethernet, 23.7
 - FDDI, 23.2–23.3, 23.3f
 - Fibre Channel standard, 23.4, 23.5f, 23.5t
 - InfiniBand, 23.8, 23.8t
 - Fiber optic networking, micro-optics-based
 - components for, 18.1–18.12
 - attenuators, 18.2, 18.9
 - beam splitters, 18.6, 18.6f
 - circulators, 18.3, 18.3f, 18.10
 - directional couplers, 18.2, 18.3, 18.3f, 18.9, 18.9f
 - Faraday rotators, 18.7, 18.7f
 - filters, 18.6
 - Fiber optic networking, micro-optics-based components for (*Cont.*):
 - gratings, 18.5–18.6, 18.6f
 - GRIN-rod lenses, 18.7, 18.8, 18.8f
 - isolators, 18.3, 18.10, 18.10f
 - mechanical switches, 18.4, 18.5, 18.5f, 18.11, 18.11f, 18.12f
 - MEMS mirrors and switches, 18.8, 18.8f, 18.11, 18.12f
 - multiplexers/demultiplexers/duplexers, 18.4, 18.4f, 18.10–18.11
 - network functions, 18.2–18.5
 - polarizers, 18.7, 18.7f
 - power splitters, 18.2–18.3, 18.2f, 18.9, 18.9f
 - prisms, 18.5, 18.5f
 - Fiber optic networks and systems, 9.14–9.15
 - detectors in, 13.2–13.3, 13.63–13.73
 - avalanche photodiodes, 13.71–13.73
 - MSM detectors, 13.73
 - pin diodes, 13.63–13.71, 13.64f, 13.66f, 13.66t, 13.68f
 - Schottky photodiodes, 13.73
 - modulators in, 13.2, 13.48–13.63
 - electroabsorption, 13.55–13.60, 13.56f, 13.57f
 - electro-optic, 13.61
 - electrorefractive, 13.61–13.62
 - lithium niobate, 13.48–13.55, 13.49f, 13.51f, 13.54f
 - semiconductor interferometric, 13.63
 - sources for, 13.1–13.48
 - distributed Bragg reflector lasers, 13.28–13.29, 13.29f
 - distributed feedback lasers, 13.30–13.32, 13.30f, 13.31f
 - laser diodes, 13.3–13.24, 13.3f, 13.4f, 13.6f, 13.7f, 13.10f, 13.14f, 13.16f, 13.17f, 13.19f–13.21f
 - light-emitting diodes, 13.36–13.42, 13.38f, 13.40f
 - quantum well lasers, 13.24–13.28, 13.25f–13.27f
 - strained layer quantum well lasers, 13.26–13.28, 13.26f, 13.27f
 - tunable lasers, 13.32–13.36, 13.33f–13.36f
 - vertical cavity surface-emitting lasers, 13.42–13.48, 13.43f, 13.45f
- [See also related topics, e.g.: Optical time-division multiplexed (OTDM)]
communication networks

- Fiber optic sensors, **24.1–24.13**
 extrinsic Fabry-Perot interferometric,
 24.2–24.4, 24.2f, 24.3f
 fiber Bragg grating, **24.5–24.8, 24.6f–24.7f**
 intrinsic Fabry-Perot interferometric sensors,
 24.4–24.5, 24.4f
 long-period grating sensors, **24.8–24.13,**
 24.9f–24.12f, 24.11t, 24.13t
- Fiber pigtail connection, **13.8**
- Fiber pulling, for fiber lasers, **25.26**
- Fiber Raman lasers, **10.7**
- Fiber-based couplers, **16.1–16.6, 16.2f, 16.5f**
- Fibre Channel Arbitrated Loop (FC-AL), **23.4**
- Fibre Channel standard, **23.2, 23.4, 23.5f, 23.5t**
- Field of view (FOV):
 and multifoil optics, **48.1**
 for scatterometers, **1.6, 1.10, 1.12**
- Field-effect transistor (FET) amplifiers, **13.70**
- Field-weighted-average resolution, **44.10**
- Fifth-order oblique spherical aberration, **45.4**
- Figures of merit:
 acousto-optic, **6.16–6.17**
 fiber optic communication links, **15.2–15.6,**
 15.3f, 15.4f
- Filling factor, electro-optic effect and, **13.50**
- Film x-ray detectors, **60.8, 60.9t, 60.10t**
- Filtered backprojection (FBP), **32.2**
- Filters and filtering:
 Bragg, **63.14**
 Fabry-Perot, **13.33, 22.8**
 in fiber optic networking, **18.6**
 frequency-guiding, **22.7–22.9**
 guided-mode resonance, **25.2, 25.30**
 Mach-Zehnder, **14.6, 21.39**
 for networking, **18.6**
 neutron, **63.18–63.19, 63.18t, 63.19f**
 of x-ray tube source spectra, **54.9**
 [See also Acousto-optic tunable filters
 (AOTFs)]
- Finite-difference time-domain (FDTD)
 analysis, for PCFs, **11.7**
- Fitting error, **5.41**
- Fixed dispersion compensation, **21.22–21.23,**
 21.23f
- Fizeau techniques, for surface figure metrology,
 46.3
- Flame brushing, of fiber Bragg gratings, **17.2**
- FLASH free-electron laser facility, **58.1**
- Flat panel detectors, for x-ray imaging,
 61.3–61.7, 61.4f, 61.6t, 61.7f
- Fluorescence:
 laser-induced, **3.21**
 x-ray, **28.1, 54.8, 62.5, 62.6f**
 and polycapillary x-ray optics,
 53.10–53.11, 53.11f
 and x-ray diffraction, **28.5–28.6, 28.6f**
- Fluorescence line narrowing (FLN), **2.13–2.14,**
 2.14f
- Fluorescent lamps, cold cathode, **8.30, 8.30f**
- Fluorides, for fiber lasers, **25.28–25.29**
- Fluoroaluminate glass, **12.4, 12.4t**
- Fluorozirconate glass (ZBLAN), **12.5, 12.5f**
 and fiber lasers, **25.3, 25.24, 25.27t, 25.28**
 fluoroaluminate glass vs., **12.4, 12.4t**
- Flux, of electron bunches, **55.17**
- Focal surfaces, of grazing incidence optics,
 45.5
- Focus anisoplanatism, **5.27–5.29, 5.28f–5.30f,**
 5.42–5.43
- Focused beams, **54.16**
- Focused single crystal diffraction, **53.12–53.14,**
 53.12f–53.13f
- Focusing:
 of gratings and monochromators, **38.3–38.6,**
 38.3f, 38.4t–38.5t
 in grazing-incidence neutron optics,
 64.3–64.7, 64.4f–64.7f
 Kerr, **7.39**
 in neutron and x-ray optics, **26.9–26.11, 63.22**
 with refractive x-ray lenses, **37.8–37.11,**
 37.9f, 37.10f
 (See also Hard x-rays, nanofocusing of)
- Focusing polycapillary x-ray optics, **53.9–53.10,**
 53.10t, 53.14
- Forward Brillouin scattering, **11.25, 11.26**
- Forward error correction (FEC) coding, **19.27,**
 21.26
- Fourier crosstalk matrix, **32.3**
- Fourier differentiation, **46.11–46.12**
- Fourier intervals, **27.4**
- Fourier transform spectroscopy, **2.5, 2.6f**
- Fourier transforming infrared spectrometer
 (FTIR), **1.14**
- Fourier transforms, **6.15, 16.4, 46.8, 55.9, 55.10**
- Four-wave mixing (FWM):
 in optical fibers, **10.2, 10.9–10.11, 10.11f**
 and SOAs, **19.13, 19.14f, 19.27, 19.33–19.35,**
 19.33f, 19.34f
 and solitons, **22.11, 22.15**
 in WDM networks, **21.19–21.20, 21.20f**

- Frames, in OTDM, 20.7
- Frank elastic constants, 8.22
- Franz-Keldysh effect, 13.59
- Fraunhofer diffraction, 42.2, 63.25
- Fraunhofer regime, in coherent x-ray optics, 27.2
- Free carriers, in fiber optic systems, 13.4
- Free electron lasers (FELs), 29.4, 41.9–41.10, 41.10f, 48.1, 58.1–58.2
- Free space fiber lasers, 25.9, 25.10, 25.13–25.15, 25.14f, 25.15f
- Freedericksz threshold voltage, of LC cells, 8.27
- Frequency:
 - and acousto-optic interaction, 6.12–6.14, 6.13f, 6.14f, 6.16
 - crossover, 8.16
 - of electro-optic modulators, 7.35
 - Greenwood, 5.19, 5.22, 5.42
 - of liquid crystals, 8.17–8.18, 8.18f
 - Nyquist, 27.4
 - Stokes, 10.8, 21.42
 - Tyler, 5.17
- Frequency chirping, 13.1, 13.17–13.18, 13.17f
- Frequency modulation, 7.24–7.25, 7.25f
- Frequency shifters, in electro-optic phase shifts, 13.51
- Frequency-division multiplexing (FDM), 9.16
- Frequency-guiding filters, 22.7–22.9
- Frequency-modulated (FM) systems, 9.16
- Fresnel diffraction, 27.2, 27.4, 40.9, 63.25
- Fresnel equations, 41.3, 64.2
- Fresnel integrals, 6.15
- Fresnel phase zones, 27.3
- Fresnel reflection, 13.6, 13.53, 17.3, 25.8
- Fresnel rhomb, 43.5–43.6, 43.5f
- Fresnel waves, 27.2
- Fresnel zone plates, 40.2, 42.3, 42.3f, 55.16
- Fresnel zones, MLLs and, 42.16f
- Fresnel's law, 63.20
- Fresnel-Soret zone plates, 40.2
- Fried's coherence diameter, 5.2, 5.7, 5.9–5.10
- Fried's coherence length, 4.8
- Fringes, in sensor signals, 24.3
- Fringes of equal chromatic order (FECO) technique, 46.2
- Full width at half-maximum (FWHM), of fiber Bragg gratings, 17.6–17.7, 17.7f, 24.8
- Full-aperture techniques, in surface figure metrology, 46.3
- Fusion splicing, 25.18
- Gabor zone plates, 40.4, 40.5, 40.7
- Gain:
 - of avalanche photodiodes, 13.71–13.72
 - of EDFAs, 14.4–14.5, 14.5f
 - polarization-dependent, 14.9, 19.18–19.20, 21.18
 - Raman, 10.5, 10.6, 21.42f
 - of SOAs, 19.4–19.6, 19.4f–19.6f
- Gain clamping, 19.14–19.15, 19.14f
- Gain dynamics, of SOAs, 19.12–19.13, 19.13f, 19.14f
- Gain efficiency, of EDFAs, 14.6
- Gain flattening, 14.6–14.7, 21.38–21.39, 21.39f
- Gain peaking, 21.38, 21.38f
- Gain per unit length, of lasers, 13.7, 13.8
- Gain ripple, 19.8, 19.8f, 19.19, 19.19f
- Gain saturation, 13.15, 13.17
- Gain-guided index antiguided fibers (GG IAG), 25.2, 25.19f, 25.22
- Gain-guided laser diodes, 13.5
- Gallium phosphide (GaP), 6.16, 6.17t, 6.29t, 6.30, 6.31t, 6.34t
- Gamma cameras, 32.2
- Gamma rays, in SPECT imaging, 32.1–32.2
- Gas detectors, 63.32–63.33
- Gas-puff sources, 57.3, 57.3t
- Gaunt factors, 56.9–56.10
- Gaussian apertures, 37.6
- Gaussian approximation of sensitivity, for optical fiber receivers, 9.10
- Gaussian error function, 15.3
- Gaussian image point, 38.5–38.6
- Gaussian integral, for noise, 15.2, 15.4
- Gaussian line profiles, 3.14, 56.5–56.7
- Gaussian spectra, 15.13
- Gaussian statistics, for wave propagation, 5.9
- Gaussian transmission, 37.5
- Gauss-Seidel iterative method, 3.21
- Geiger counters, 60.5, 60.9t
- Geiger region, 60.5
- Gemini North telescope, 5.20, 5.21f
- GENLN2 (code), 3.23
- Geometrical point spread function (GPSF), 44.13
- Germanate:
 - in fiber lasers, 25.27t, 25.28
 - in optical fibers, 12.3–12.4, 12.6, 12.6f
- Glancing angle, of crystal monochromators, 39.1, 39.2

- Glass:
- for fiber lasers, 25.27t, 25.28–25.29
 - fluorozirconate (ZBLAN), 12.5, 12.5f
 - and fiber lasers, 25.3, 25.24, 25.27t, 25.28
 - fluoraluminate glass vs., 12.3t, 12.4, 12.4t
 - fluoroaluminate, 12.4, 12.4t
 - heavy-metal fluoride, 12.1–12.5, 12.2f, 12.3t, 12.4t, 12.5f
 - heavy-metal oxide, 12.2t, 12.3–12.7, 12.3t, 12.4t, 12.5f–12.7f
 - for hollow waveguides, 12.2f, 12.11–12.13, 12.12f
 - negative core-cladding index difference, 11.14, 11.15
 - Raman bands of, 11.24
 - Rayleigh scattering in, 11.21
 - Zerodur, 47.5
- Glass micro-pore optics, 49.1–49.6, 49.2f–49.6f
- Global climate change, 3.43–3.45, 3.44f
- Goebel mirrors, 26.10
- GOES-13 satellite, 44.16–44.17, 44.17f
- Gooch-Tarry first minimum condition, 8.26
- Goos-Hanschen shifts, 13.55
- Gordon-Haus effect, 22.7–22.8, 22.11
- Graded index (GRIN) fibers, 15.17
- Graded index separate confinement
 - heterostructure (GRINSCH), 13.4f, 13.5
- Graded index-rod (GRIN-rod) lenses, 18.7, 18.8, 18.8f, 18.10, 18.11f
- Gradient tilt (G-tilt), 4.3
 - and adaptive optics, 5.14–5.16
 - and angle of arrival, 4.23, 4.25, 4.26
- Gradients, of wavefronts, 5.23
- Graphical user interface (GUI), for SHADOW code, 35.3
- Grasshopper* monochromator, 38.3
- Grating equation, for VUV and soft x-ray region, 38.2–38.3
- Gratings:
- Bragg, 20.14, 25.29, 25.30
 - circular, 40.1
 - fiber Bragg, 17.1–17.9
 - applications, 17.8–17.9, 17.8f
 - chirped, 21.22–21.23, 21.23f, 21.25–21.26, 21.25f
 - fabrication, 17.4–17.8, 17.5f–17.7f
 - and fiber lasers, 25.8, 25.16, 25.18, 25.30, 25.31
 - long-period gratings vs., 24.9, 24.11
 - photosensitivity, 17.2–17.3
- Gratings, fiber Bragg (*Cont.*):
- properties of, 17.3–17.4, 17.4f
 - sensors based on, 24.5–24.8, 24.6f–24.7f
- Fresnel diffraction of, 40.9
- Hill, 17.5
- long-period, 21.39, 21.39f, 24.10–24.11, 24.11t, 24.13t
- for networking, 18.5–18.6, 18.6f
- sampled, 13.34, 13.34f
- short-period, 24.6
- superstructure, 13.34, 13.35f
- in VUV and soft x-ray region, 38.1–38.8
 - diffraction properties, 38.1–38.3, 38.2f
 - dispersion properties, 38.6–38.7
 - efficiency, 38.8
 - focusing properties, 38.3–38.6, 38.3f, 38.4t–38.5t
 - resolution properties, 38.7
- Grazing incidence optics:
- aberrations of, 44.6–44.12, 44.7f, 44.9f–44.11f, 45.1–45.8, 45.2f
 - image formation with, 44.3–44.18
 - and multifoil optics, 48.1–48.2
 - and pumping by EUV lasers, 58.3, 58.4f
 - telescopes with, 44.6–44.12, 44.7f, 44.9f–44.11f
 - and x-ray mirrors, 44.3–44.6, 44.4f–44.6f
- Grazing incidence x-ray optics, 44.12–44.18, 44.12f–44.18f
- Grazing-angle reflection, 63.21
- Grazing-incidence neutron optics, 64.1–64.7
 - diffractive scattering and mirror surface roughness, 64.2–64.3
 - and imaging focusing optics, 64.3–64.7, 64.4f–64.7f
 - total external reflection, 64.1–64.2
- Green flashes, 3.43
- Green's function solution, 6.14
- Greenwood frequency, 5.19, 5.22, 5.42
- Group velocity dispersion (GVD), 11.18, 11.18f, 11.19f, 15.11
- Grüneisen constant, 6.17
- GTWave technology, 25.11
- Guidance, in photonic crystal fibers, 11.11–11.26
 - attenuation mechanisms, 11.19–11.22, 11.20f, 11.21f
 - birefringence, 11.17
 - Brillouin scattering, 11.25–11.26, 11.25f

- Guidance, in photonic crystal fibers (*Cont.*):
 - group velocity dispersion, 11.18, 11.18*f*, 11.19*f*
 - Kerr nonlinearities, 11.22–11.24, 11.24*f*
 - negative core-cladding index difference, 11.14–11.17, 11.14*f*–11.17*f*
 - positive core-cladding index difference, 11.12–11.14, 11.12*f*, 11.13*f*
 - Raman scattering, 11.24
 - resonance and antiresonance, 11.12
- Guided-mode resonance filters (GMRFs), 25.2, 25.30
- Guides, neutron, 63.15–63.18
- Half-period zones, of zone plates, 40.1
- Half-wave voltage, of electro-optic modulators, 7.19, 7.22
- Hard x-ray beamlines, SHADOW code for, 35.5, 35.5*f*
- Hard x-ray optics, astronomical, 47.9–47.10
- Hard x-ray telescopes, 50.2
- Hard x-rays, nanofocusing of, 42.1–42.17
 - history of, 42.2–42.4, 42.2*f*, 42.3*f*
 - instrumental beamline arrangement and measurements for, 42.9–42.10, 42.9*f*–42.12*f*
 - limitations of, 42.15–42.17, 42.16*f*–42.17*f*
 - with magnetron-sputtered MLLs, 42.5–42.7, 42.6*f*–42.8*f*
 - on MLLs with curved interfaces, 42.14, 42.15*f*
 - Takagi-Taupin calculations for, 42.12–42.14
 - volume diffraction calculations for, 42.4–42.5, 42.5*f*
 - with wedged MLLs, 42.12–42.13, 42.13*f*, 42.14*f*
- Hardware implementation, for adaptive optics, 5.21–5.38
 - higher-order wavefront sensing techniques, 5.36–5.37
 - laser beacons, 5.27–5.34, 5.28*f*–5.31*f*, 5.33*f*
 - real-time processors, 5.34–5.35, 5.34*f*, 5.35*f*
 - Shack-Hartmann technique, 5.23–5.27, 5.23*f*, 5.25*f*, 5.26*f*
 - tracking, 5.21–5.23, 5.22*f*
 - wavefront correctors, 5.37–5.38, 5.38*f*
- Heavy water, scattering in, 63.10
- Heavy-hole (HH) bands, of strained layer quantum well lasers, 13.27
- Heavy-metal fluoride glass (HMFG) fibers, 12.1–12.5, 12.2*f*, 12.3*t*, 12.4*t*, 12.5*f*
- Heavy-metal oxide glass fibers, 12.2, 12.3–12.7, 12.3*t*, 12.4*t*, 12.5*f*–12.7*f*
- HEFT balloon payloads, 47.10
- Height profilometry, 46.3
- Helium atoms, 2.3
- Helium [$^3\text{He}(n, p)$] reaction, 63.31
- Helmholtz equation, 5.8
- Hermetic enclosure, for laser diodes, 13.7, 13.7*f*
- Hermite-Gaussian functions, 11.7
- HERO balloon payload, 47.10
- Heterodyne detection, 3.34
- Heterostructures, of fiber optic devices, 13.2
- High aspect ratio microlithography (HARM), 61.3
- High harmonic production, of x-ray lasers, 58.2
- High Resolution Doppler Imager (HRDI), 3.36, 3.37*f*
- Higher-order mode (HOM) fibers, 25.2, 25.19*f*, 25.22
- High-power spectrally controlled fiber lasers, 25.29
- High-power ultrashort pulse technologies, for fiber lasers, 25.32–25.33
- High-power USP oscillators, 25.32–25.33
- High-reflectivity mirrors, 41.7–41.8, 41.8*f*
- High-resolution (HR) acousto-optic deflectors, 6.29, 6.29*t*
- Hill gratings, 17.5
- HITRAN database, 3.14, 3.22–3.23, 3.22*f*, 3.23*f*
- HITRAN-PC program, 3.26
- Hobby-Eberly telescopes, 5.2
- Holeburning:
 - optical, 2.13, 2.14*f*
 - spatial, 20.2
- Hollow glass waveguides (HGWs), 12.2*f*, 12.11–12.13, 12.12*f*
- Hollow waveguides, 12.2*f*, 12.2*t*, 12.3*t*, 12.10–12.13, 12.12*f*
- Hollow-core photonic crystal fibers:
 - attenuation in, 11.20–11.22, 11.20*f*
 - birefringence of, 11.17
 - and group velocity dispersion, 11.18, 11.19*f*
 - Kerr effects for, 11.23
 - and negative core-cladding index difference, 11.14–11.15, 11.14*f*
- Holmium-doped fibers, 25.23*t*, 25.25–25.26
- Homogeneous broadening, 20.1

- Hot bands, in molecular spectroscopy, 2.5
 Hot filament sources, of x-ray tubes, 54.10
 HRCam system, 5.23
 Huang-Rhys factor, 2.15
 Huesler alloy, 63.28–63.29
 Hufnagel model, of atmospheric turbulence, 3.29
 Hufnagel-Valley model, of atmospheric turbulence, 3.30, 5.7, 5.8
 Huygens-Fresnel approximation, 3.31–3.33
 Hybrid mode locking, 20.17
 Hybrid network topologies, for WDM networks, 21.7
 Hydrogen (H^-) ions, negative, 2.3
 Hydrogen loading, 17.2
 Hyperboloid-hyperboloid (HH) grazing incidence x-ray telescopes, 44.10–44.12, 44.11f
- IBM, 23.1, 23.2
 Ice, in standard atmosphere, 3.6, 3.42, 3.42f
 Ideal receivers, 9.9
 Ideality factor, of semiconductor diodes, 13.69
 IEEE (Institute of Electrical and Electronics Engineers) standards, 23.2, 23.7
 Image (scophony) AO modulators, 6.34–6.35
 Image receptors, in medical imaging, 31.4, 31.4f
 Image wander, 4.3 (*See also* Angle of arrival)
 Images:
 long-exposure, 4.3–4.7
 from Nomarski microscope, 46.2
 short-exposure, 4.3, 4.31–4.35, 4.31f–4.34f, 4.35t
 x-ray, 31.1–31.4
 Imaging:
 and atmospheric turbulence, 3.34
 diffractive, 41.9–41.10, 41.10f
 with grazing incidence optics, 44.3–44.18
 with grazing-incidence neutron optics, 64.3–64.7, 64.4f–64.7f
 medical, 31.1–31.10
 applications of, 31.9, 31.10
 digital displays, 31.8–31.9, 31.8f–31.10f
 digital tomosynthesis, 31.7–31.8
 and inverse Compton x-ray sources, 59.3–59.4
 and polycapillary x-ray optics, 53.14–53.16, 53.15f–53.16f
 radiography, 31.1–31.4, 31.2f–31.4f
 tomography, 31.1, 31.5–31.7, 31.5f–31.7f
 x-ray detectors for, 61.2
- Imaging (*Cont.*):
 molecular, 32.1
 monochromatic, 53.16–53.17, 53.17f
 multi-energy, 54.9–54.10
 nuclear, 53.17, 53.18, 53.18f
 with refractive x-ray lenses, 37.6–37.7, 37.6f, 37.7f
 scatter rejection in, 53.14–53.16, 53.15f–53.16f
 SPECT, 32.1–32.3
 and spectroscopy, 5.19–5.21, 5.19f, 5.20f
 thermal, 12.3t
 through atmospheric turbulence, 4.1–4.37
 aberration variance and approximate
 Strehl ratio for, 4.27–4.28, 4.28f
 and adaptive optics, 4.35–4.36
 angle of arrival fluctuations, 4.23–4.26, 4.25f, 4.26f
 and covariance and variance of
 expansion coefficients, 4.20–4.22, 4.21t, 4.22f–4.23f
 Kolmogorov turbulence and atmospheric
 coherence length, 4.7–4.10, 4.8f, 4.9f
 long-exposure images, 4.3–4.7
 modal correction of turbulence, 4.28–4.30, 4.29t, 4.30f
 and modal expansion of aberration
 function, 4.17–4.20, 4.17f–4.18f, 4.19t, 4.20t
 and resolution of telescopes, 4.2–4.3
 short-exposure image, 4.31–4.35, 4.31f–4.34f, 4.35t
 and systems with annular pupils, 4.10–4.16, 4.11f–4.15f, 4.15t
- Imaging detectors (x-ray), 61.1–61.8
 CCD detectors, 61.7–61.8, 61.8f
 flat panel detectors, 61.3–61.7, 61.4f, 61.6t, 61.7f
 geometries for and classifications of, 61.1–61.3, 61.2f
- Imaging plates, in neutron optics, 63.34
 Impact ionization coefficient, 13.72
 Incidence angle, constant, 38.2
 Incident power measurement, in scatterometers, 1.14–1.15
 Incoherent scattering, 26.7, 31.2, 63.7, 63.8
 Index ellipsoid, of electro-optic modulators, 7.4–7.7, 7.5f–7.6f

- Index of refraction:
 - complex, **48.1**
 - and fiber Bragg gratings, **17.2**
 - and Kolmogorov turbulence, **4.7**
 - of liquid crystals, **8.18–8.19**
 - in neutron optics, **63.19–63.20**
 - phase velocity, **7.15–7.16, 7.16f**
 - of photonic crystal fibers, **11.9–11.10**
 - structure function of, **5.6–5.7**
- Index-guided laser stripes, **13.6**
- Indirect modulation, in OTDM networks, **20.17–20.18, 20.17f**
- Indirect-conversion flat panel detectors, **61.4–61.6, 61.4f**
- Inelastic optical processes, **3.21–3.22**
- Inelastic scattering, **21.20, 63.3**
- In-fiber devices, for photonic crystal fibers, **11.27, 11.28**
- InfiniBand standard, **23.8, 23.8t**
- InFOCUS/SUMIT balloon payloads, **47.10**
- Infrared (IR) optical fibers, **12.1–12.13**
 - applications, **12.13**
 - categories and properties of, **12.1–12.3, 12.2f, 12.2t, 12.3t**
 - crystalline, **12.2t, 12.3t, 12.7–12.10, 12.8f, 12.10f**
 - heavy-metal oxide glass in, **12.2t–12.4t, 12.3–12.7, 12.5f–12.7f**
 - in hollow waveguides, **12.2t, 12.3t, 12.10–12.13, 12.12f**
- Injection:
 - electrical, **13.4, 13.5, 13.45, 13.45f**
 - in LEDs, **13.37**
- Injection seeding, optical clock recovery and, **20.21–20.22**
- In-line semiconductor optical amplifiers (SOAs), **19.24**
- Inner scale of turbulence, **4.7**
- Inorganic crystals, **7.33**
- In-plane switching (IPS) cells, **8.16, 8.26, 8.28f**
- Input saturation power, for fiber amplifiers, **14.3**
- Insertion devices:
 - for synchrotron radiation sources, **55.9–55.16, 55.10f, 55.12f, 55.13f**
- Insertion loss, **7.36, 13.53, 18.1**
- Installation loss, for fiber optic communication links, **15.6–15.7, 15.8t**
- Instrument signature, of scatterometers, **1.6, 1.11–1.13, 1.11f, 1.13t**
- Instrument transfer function effects, in x-ray mirror metrology, **46.9–46.11**
- Instrumental line spread function, **38.7**
- Integrated Mach-Zehnder interferometers, **20.18–20.19, 20.19f**
- Integrated planar lightwave circuits (iPLCs), **21.12, 21.13f**
- Integrated-optic modulators, **7.3, 7.30–7.32, 7.31f–7.33f**
- Integrating x-ray detectors, **61.2**
- Intelligent Physical Protocol Enhanced Physical Project, **23.4**
- Intensity, of spectral lines, **56.2–56.4**
- Intensity modulation and modulators, **6.34, 13.57–13.58**
- Interaction zone, for Compton scattering, **59.1**
- Interband processes, SOAs and, **19.12, 19.13**
- Interconnected switchable networks, **23.4**
- Interference:
 - multiple-Bragg-beam, **43.1**
 - in neutron optics, **63.25–63.27, 63.27f**
 - Nomarski differential, **46.4**
 - Pendellösung, **63.26**
 - in SOAs, **19.23**
 - in x-ray optics, **26.8, 26.8f, 26.9f**
- Interferometers:
 - Bragg reflection in, **63.26, 63.27**
 - crystal, **63.26–63.27, 63.27f**
 - Fabry-Perot, **3.36, 24.2–24.5, 24.2f–24.4f**
 - Laue transmission in, **63.26**
 - Linnik, **46.2**
 - Mach-Zehnder, **21.34**
 - and Bragg grating sensors, **24.8**
 - and DPSK, **21.33–21.34, 21.33f**
 - and electro-optic modulators, **7.22, 7.24, 7.32, 7.38**
 - and fiber Bragg gratings, **17.8**
 - integrated, **20.18–20.19, 20.19f**
 - in OTDM networks, **20.22, 20.23f**
 - SOAs in, **19.31–19.32, 19.31f, 19.33f, 19.36**
 - and supercontinuum generation, **11.23**
 - Michelson, **17.8–17.9, 17.8f**
 - perfect crystal, **63.26–63.27, 63.27f**
 - phase-measuring, **46.2**
 - Sagnac, **20.22**
 - ultrafast nonlinear, **19.32**
 - unbalanced nonlinear, **20.22**
- Interferometric lithography (IL), **34.4**
- Interferometric Mach-Zehnder modulators, **13.51–13.52, 13.54–13.55**

- Interferometric method, of FBG fabrication, 24.6, 24.6f
- Interferometric modulators, 13.51–13.52
- Intergovernmental Panel on Climate Change, 3.44
- Interleaving, in OTDM networks, 20.6–20.7, 20.7f
- Intermodal dispersion, in optical fibers, 9.5–9.6
- Intermodulation distortions (IMDs), 15.5–15.6
- Intermodulation (IM) products, of acousto-optic devices, 6.30
- Internal quantum efficiency, 13.9, 13.10
- Internal writing technique, for FBGs, 17.4–17.5
- International Space Station, 49.4
- International Telecommunications Union standards, 15.15
- Inter-switch links (ISLs), 23.4
- Intersymbol interference (ISI), 19.23
- Intraband processes, of SOAs, 19.12
- Intramodal dispersion, 9.5–9.6, 9.6f
- Intrinsic Fabry-Perot interferometric (IFPI) sensors, 24.4–24.5, 24.4f
- Inverse Compton scattering, 59.1
- Inverse Compton x-ray sources, 59.1–59.4, 59.3t
- Inverse Raman effect, 10.5, 10.6, 14.9
- Ionization, of electrons, 56.2
- Ionization chambers, 60.3–60.4, 60.9t
- Ionization x-ray detectors, 60.3–60.7, 60.9t, 60.10t
- Ionizing radiation, fiber optic communication links and, 15.17
- ISO standards, 23.2
- Isolation (directionality), of fiber-optic components, 18.1
- Isolators, for networking, 18.3, 18.10, 18.10f
- Isoplastic angle, 5.19
- Isotropic diffraction:
and acousto-optic interactions, 6.9–6.10, 6.10f, 6.13, 6.13f
by deflectors, 6.23–6.24, 6.24f
- Iterative phasing technique, for coherent diffraction microscopy, 27.4
- Jacobi method, of electro-optic effect, 7.11–7.13
- Jahn-Teller effect, 2.9
- Jitter:
in fiber optic communication links, 15.15–15.16
and solitons, 22.6–22.8, 22.11, 22.16
- Jitter transfer function (JTF), 15.16
- Johansson bent/ground focusing monochromator, 39.5
- Johnson noise, of pin diodes, 13.70
- Jülich SANS instrument, 64.4
- Kagomé lattice, 11.5f, 11.11, 11.16
- Karhunen-Loève functions, 4.36
- Keck Observatory, 5.27
- Keck telescopes, 4.36, 5.2, 5.27
- Kerr cells, 7.34
- Kerr effect, 7.11, 14.11, 19.34, 20.1, 20.22, 22.3
- Kerr electro-optic effect, 7.9t–7.10t, 7.11
- Kerr focusing, 7.39
- Kerr interactions, 10.2
- Kerr nonlinearities, 7.38–7.39, 11.22–11.24, 11.24f
- Kerr-lens mode-locking (KLM), 7.11, 7.39
- Kinematic theory of diffraction, in neutron optics, 63.5
- Kinematical theory of diffraction, for crystals, 39.2, 39.3
- Kirkpatrick-Baez (KB) mirrors, 44.4, 44.4f, 63.21, 64.5–64.6, 64.5f
- Kirkpatrick-Baez (KB) optics, 47.7–47.8, 47.8f, 47.9f, 50.5
- Kirkpatrick-Baez (KB) systems, 48.3–48.4, 48.3f
- Kolmogorov model of turbulence, 5.5–5.6, 5.11
- Kolmogorov spatial power spectral density, 5.6
- Kolmogorov spectrum, 3.28, 3.31
- Kolmogorov turbulence, 4.3, 4.7–4.10, 4.8f, 4.9f, 4.27, 4.30, 4.36
- Kossel patterns, 53.11
- Kramer approximation, for Bremsstrahlung radiation, 54.5
- Kramers-Kronig relations, 13.61, 17.3, 20.22
- Kronecker delta, 4.19
- Kronig-Penney model, 2.12
- KRS-5 fiber, 12.7–12.8
- K-shell emission lines, of elements, 36.3t–36.8t
- Kumahov capillary lenses, 53.9
- Lab-based radiation sources, 50.2–50.7, 50.4f, 50.6f, 50.8f
- Lamb dip spectroscopy, 2.5, 2.7f

- Lamb shifts, 2.2, 2.3
- Lambert-Beer law of attenuation, 63.11
- Lamor precession, 63.29–63.30
- Lamps, cold cathode fluorescent, 8.30, 8.30f
- Langmuir-Blodgett techniques, 7.34
- Large Binocular Telescope (LBT), 5.5
- Large flat-mode fibers, in fiber lasers, 25.19f, 25.20
- Large-mode-area (LMA) fibers, 25.2, 25.4, 25.5
 - in all-fiber monolithic systems, 25.16
 - chirally coupled core fibers, 25.21–25.22
 - designs, 25.19f
 - equations for, 25.8–25.9
 - and photonic crystal fibers, 25.20–25.21
 - techniques using, 25.18–25.20
- Laser beacons (laser guide star sensing), 5.21, 5.23, 5.27–5.34
 - focus anisoplanatism of, 5.27–5.29, 5.28f–5.30f
 - and mesospheric sodium laser beams, 5.32–5.34, 5.33f
 - Rayleigh, 5.30–5.32, 5.31f
- Laser beam scanning, by high-resolution deflectors, 6.29
- Laser diodes, 13.3–13.24
 - double heterostructure, 13.3–13.8, 13.3f, 13.4f, 13.6f, 13.7f
 - noise characteristics of, 13.18–13.24, 13.19f–13.21f
 - operating characteristics of, 13.8–13.13, 13.10f
 - transient response of, 13.13–13.18, 13.14f, 13.16f, 13.17f
- Laser guide star (LGS) sensing, 5.21, 5.23, 5.27 (*See also* Laser beacons)
- Laser mode locking, 7.38–7.39
- Laser-generated plasmas, 56.1–56.10
 - and Bremsstrahlung radiation, 56.8–56.10
 - and recombination radiation, 56.10
 - sources of, 56.1
 - and spectral line emission, 56.2–56.8
- Laser-heated pedestal growth (LHPG)
 - technique, 12.9–12.10, 12.10f
- Laser-induced fluorescence, 3.21
- Laser-induced-breakdown spectroscopy (LIBS), 3.39
- Lasers:
 - Bragg reflector, 13.7
 - bulk, 25.5–25.6
 - buried heterostructure, 13.5
- Lasers (*Cont.*):
 - carbon dioxide, 12.3t, 12.9, 12.13
 - chirp of, 9.8
 - continuous-wave, 25.4, 25.5, 25.5f, 25.7
 - diode lasers, 13.49
 - dye lasers, 5.32, 5.33f, 5.34
 - distributed Bragg reflector, 9.8, 13.7, 13.28–13.29, 13.29f, 20.14
 - distributed feedback
 - in fiber optic systems, 13.30–13.32, 13.30f, 13.31f
 - and optical fibers, 9.8
 - in OTDM networks, 20.14–20.15, 20.15f
 - quarter-wavelength-shifted grating, 13.31–13.32, 13.31f
 - edge-emitting, 13.3
 - and electroabsorption modulators, 13.60
 - erbium-doped yttrium aluminum garnet (Er:YAG), 12.3t, 12.6, 12.13
 - European X-ray Free-Electron Laser, 58.1
 - external cavity, 13.22, 13.33
 - extreme ultraviolet, 58.2–58.4, 58.3f
 - Fabry-Perot, 13.12–13.13, 13.29
 - multilongitudinal mode, 9.7–9.8
 - semiconductor, 20.13–20.14, 20.14f
 - fiber, 25.1–25.33
 - all-fiber monolithic systems, 25.16–25.18, 25.16f
 - architectures, 25.9–25.18, 25.12f, 25.19f
 - bulk lasers vs., 25.5–25.6
 - dopants, 25.22–25.26, 25.23t
 - equations for, 25.8–25.9
 - fabrication of, 25.26–25.29, 25.27t
 - free space, 25.13–25.15, 25.14f, 25.15f
 - growth of, 25.5, 25.5f
 - history of, 25.3–25.4, 25.4f
 - limitations, 25.6–25.7
 - LMA fiber designs, 25.18–25.22, 25.19f
 - operation of, 25.7–25.8
 - pumping techniques, 25.9–25.13, 25.11f, 25.12f
 - spectral and temporal modalities, 25.29–25.33
 - fiber Raman, 10.7
 - free electron, 41.9–41.10, 41.10f, 48.1, 58.1–58.2
 - free space fiber, 25.9, 25.10, 25.13–25.15, 25.14f, 25.15f
 - mesospheric sodium, 5.32–5.34, 5.33f
 - mode-locked, 20.15–20.17, 20.16f

Lasers (*Cont.*):

- multiple quantum well, 13.24–13.25
- and optical fibers, 9.7–9.8
- in OTDM communication networks, 20.13–20.17, 20.14f, 20.16f
- phase noise (finite line width) of, 9.8
- planar buried heterostructure, 13.6
- plasma-based EUV, 58.2–58.4, 58.3f
- pulsed-dye, 5.32
- quantum well, 13.24–13.28, 13.25f–13.27f
- relative intensity noise of, 9.11
- ridge waveguide, 13.6
- for scatterometers, 1.8
- semiconductor, 13.1, 20.13–20.14, 20.14f
- single-longitudinal-mode, 9.8
- sum-frequency, 5.32
- tunable, 13.32–13.36, 13.33f–13.36f
- vertical cavity surface-emitting, 9.7n, 13.42–13.48, 13.43f, 13.45f, 19.14
- and wavelength-division multiplexing, 9.8
- x-ray, 58.1–58.4, 58.3f
- Lattice-matched compositions, SOAs and, 19.11
- Lattice-matched epitaxial layers, of fiber optic devices, 13.2
- Laue crystals, 35.3
- Laue equation, 39.1
- Laue geometry, for crystal monochromators, 39.2f, 39.4–39.6
- Laue lenses, multilayer [*see* Multilayer Laue lenses (MLLs)]
- Laue phase retarders, 43.6
- Laue transmission, 26.8, 26.9f, 63.24, 63.26
- Laue-diffracting crystals, 43.2
- Layer adding method, multiple scattering and, 3.21
- LBLRTM (code), 3.23
- Leakage channel fibers, 25.21
- Legendre polynomials, 46.6
- Legendre-Fourier (L-F) polynomials, 45.6
- Lenses:
 - Airy distribution of, 40.3
 - Bragg-Fresnel, 40.9–40.10, 40.9f
 - catadioptric systems of, 35.1
 - catoptric systems of, 35.1
 - dioptric systems of, 35.1
 - GRIN-rod, 18.7, 18.8, 18.8f, 18.10, 18.11f
 - Kerr, 7.11, 7.39
 - Kumuhov capillary, 53.9

Lenses (*Cont.*):

- multilayer Laue, 42.1–42.17
 - with curved interfaces, 42.14, 42.15f
 - history of, 42.2–42.4, 42.2f, 42.3f
 - instrumental beamline arrangement and measurements, 42.9–42.10, 42.9f–42.12f
 - limitations of, 42.15–42.17, 42.16f–42.17f
 - magnetron-sputtered, 42.5–42.7, 42.6f–42.8f
 - Takagi-Taupin calculations, 42.12–42.14
 - volume diffraction calculations, 42.4–42.5, 42.5f
 - wedged, 42.12–42.13, 42.13f, 42.14f
 - and x-ray/neutron optics, 26.10
- in neutron optics, 63.22–63.23, 63.23f
- refractive x-ray, 37.3–37.11
 - applications of, 37.11
 - history of, 37.3
 - nanofocusing, 37.8–37.11, 37.9f, 37.10f
 - parabolic, 37.4–37.8, 37.4f, 37.6f, 37.7f
 - zone plates as, 40.3–40.4
- Leslie viscosity coefficients, 8.24
- Lifetime, photon, 20.1
- Light:
 - absorption of, 3.4–3.5, 3.4f
 - out-of-plane profile of, 13.11
 - propagation of, 7.3
 - retroreflection of guided, 13.6–13.7
 - spatial characteristics of, 13.11–13.12, 13.46
 - spectral characteristics of, 13.12–13.13
 - theory of interaction of atmosphere and, 3.11–3.22
 - inelastic optical processes, 3.21–3.22
 - Mie scattering, 3.16–3.18, 3.17f–3.19f
 - molecular absorption, 3.12–3.15, 3.13f
 - molecular emission and thermal spectral radiance, 3.18, 3.20, 3.20f
 - molecular Rayleigh scattering, 3.15–3.16
 - surface reflectivity and multiple scattering, 3.21, 3.21f
- Light detection and ranging (LIDAR) systems, 3.38–3.39, 3.38f, 25.25
- Light out vs. current in (L-I curve), 13.9–13.11, 13.10f, 13.46–13.47
- Light-emitting diodes (LEDs), 13.1, 13.36–13.42, 13.38f
 - edge-emitting, 13.40
 - operating characteristics of, 13.40–13.42, 13.40f
 - and optical fibers, 9.7

- Light-emitting diodes (LEDs) (*Cont.*):
 - surface-emitting, 13.38–13.40, 13.38*f*
 - and transmissive TFT LCDs, 8.29–8.31, 8.30*f*
- Light-hole (LH) bands, 13.27
- Linac Coherent Light Source (LCLS), 58.1
- Line edge roughness (LER), in extreme ultraviolet lithography, 34.6
- Line width, of lasers, 9.8
- Linear attenuation coefficient, 31.1, 31.2
- Linear dispersion, 38.7
- Linear electro-optic (Pockels) effect:
 - and electro-optic modulators, 7.6–7.11, 7.8*t*, 7.11
 - and liquids, 7.34
 - in OTDM networks, 20.1, 20.18, 20.19
- Linear optical amplifiers (LOAs), 19.14, 19.27
- Linear polarization analyzers, 43.4, 43.4*f*
- Linear polarizers, 43.2–43.3, 43.3*f*
- Linear regime, for rare-earth-doped fiber amplifiers, 14.3
- Linear tomography, 31.5, 31.5*f*
- Linearity, of scatterometers, 1.15
- Line-by-line transmission programs, 3.23, 3.24*f*
- Linewidth enhancement factor, 13.17, 13.20
- Link budget analysis, for fiber optic communication:
 - installation loss, 15.6–15.7, 15.8*t*
 - optical power penalties, 15.8–15.17, 15.10*f*, 15.11*f*
- Linnik interferometers, 46.2
- Liouville's theorem, 54.11, 54.15, 54.16, 63.15, 63.22
- Liquid crystal (LC) cells, 8.25–8.28, 8.26*f*, 8.27*f*, 8.28*f*
- Liquid crystal displays (LCDs), 8.29–8.35, 61.3
 - reflective, 8.31–8.32, 8.31*f*
 - transmissive TFT, 8.29–8.31, 8.30*f*
 - transreflective, 8.32–8.35, 8.33*f*, 8.34*f*
- Liquid crystals (LCs):
 - composition of, 8.2–8.4, 8.3*f*
 - dielectric properties of, 8.14–8.18, 8.15*f*, 8.18*f*
 - elastic properties of, 8.22–8.23, 8.23*f*
 - limitations of, 8.37
 - optical properties, 8.18–8.22, 8.19*f*, 8.20*f*, 8.22*f*
 - phase transitions of, 8.13, 8.13*f*, 8.14*f*
 - phases of, 8.8–8.13, 8.11*f*–8.12*f*
 - physical properties of, 8.13–8.23
 - in polymer/liquid crystal composites, 8.36–8.37, 8.36*f*–8.37*f*
- Liquid crystals (LCs) (*Cont.*):
 - types of, 8.4–8.8, 8.4*f*–8.6*f*, 8.7*t*, 8.9*t*
 - viscosities of, 8.23–8.25, 8.24*f*
- Liquid-crystal-on-silicon (LCoS) displays, 8.32
- Liquid-metal anodes, 54.13
- Liquids, in electro-optic modulators, 7.34
- Lithium niobate (LiNbO_3), 6.30, 6.31, 6.31*t*
- Lithium niobate (LiNbO_3) modulators, 13.2, 13.48–13.55, 13.49*f*
 - electro-optic effect, 13.49–13.51, 13.51*f*
 - electro-optic polymer, 13.55
 - high-speed operation of, 13.52–13.53
 - insertion loss in, 13.53
 - as Mach-Zehnder modulators, 13.51–13.52, 13.54–13.55, 13.54*f*
 - phase modulation by, 13.51
 - photorefractivity and optical damage of, 13.54
 - polarization independence of, 13.53
 - Y-branch interferometric, 13.51–13.52
- Lithium [$^6\text{Li}(n, \alpha)$] reaction, 63.31
- Lithium-drifted silicon x-ray detectors, 60.6
- Lithography:
 - electron-beam, 40.8
 - extreme ultraviolet, 34.1–34.6, 34.2*t*, 34.3*f*–34.6*f*
 - extreme ultraviolet-interferometric, 34.4–34.5, 34.5*f*
 - high aspect ratio microlithography, 61.3
 - interferometric, 34.4
 - optical, 34.1
- L-I-V measurements, for SOAs, 19.17, 19.17*f*
- Lobster-eye (LE) optics, 48.2–48.4, 48.2*f*, 48.3*f*, 49.3–49.4, 49.3*f*–49.4*f*
- Lobster-ISS system, 50.7
- Local area networks (LANs), 9.14, 21.7
 - power splitters and couplers in, 18.2, 18.3*f*
 - standards for, 23.2, 23.6, 23.7
- Local gain per unit length, of lasers, 13.8
- Local oscillators, 9.13
- Local shift variance, in grazing incidence x-ray optics, 44.14
- Log-amplitude structure function, 4.5
- Long trace profiler (LTP), 46.4, 46.5, 46.5*f*
- Long-exposure images, 4.3–4.7
- Long-exposure MCF, 4.10
- Longitudinal electro-optic modulators, 7.16, 7.17, 7.17*f*
- Longitudinal spatial modulation (LSM), 6.12, 6.17

- Long-period gratings (LPGs), 21.39, 21.39f, 24.8–24.13, 24.9f–24.12f, 24.11t, 24.13t
- Long-wave infrared (LWIR) AOTFs, 6.42
- Lorentzian broadening, 3.14
- Lorentzian lineshapes, of spectra, 2.13, 56.4–56.7
- Loss:
- Akhieser, 6.17
 - bend, 11.21–11.22, 11.21f
 - connector, 15.7, 15.8t
 - coupling, 18.1
 - insertion, 7.36, 13.53, 18.1
 - installation, 15.6–15.7, 15.8t
 - optical power, 21.13–21.14
 - polarization-dependent, 19.18, 21.18
 - radiation induced, 15.17
 - splice, 15.7, 15.8t
 - transmission, 15.7
- Loss, of fiber-optic components, 18.1
- Low cost (LC) FDDI, 23.3
- Low leakage guidance, 11.16, 11.17
- Low-frequency fluctuations (LFF), of lasers, 13.23
- LOWTRAN program, 3.23–3.24, 3.25f–3.26f
- L-shell emission lines, of elements, 36.3t–36.8t
- Lump amplification, 21.44
- Lyman series, 56.2
- Lyotropic liquid crystals (LCs), 8.3, 8.5f
- Lysozyme diffraction, 53.12–53.14, 53.12f, 53.13f
- Mach-Zehnder devices, as fiber-based couplers, 16.4–16.5, 16.5f
- Mach-Zehnder filters, 14.6, 21.39
- Mach-Zehnder interferometers (MZIs):
- and Bragg grating sensors, 24.8
 - and DPSK, 21.33–21.34, 21.33f
 - and electro-optic modulators, 7.22, 7.24, 7.32, 7.38
 - and fiber Bragg gratings, 17.8
 - integrated, 20.18–20.19, 20.19f
 - in OTDM networks, 20.22, 20.23f
 - SOAs in, 19.31–19.32, 19.31f, 19.33f, 19.36
 - and supercontinuum generation, 11.23
- Mach-Zehnder modulators:
- interferometric, 13.51–13.52, 13.54–13.55, 13.54f, 13.63
 - in WDM networks, 21.30, 21.31f, 21.32f
- Magnesium oxide (MgO), 2.20, 2.20f, 2.22
- Magnetic circular dichroism (MCD), 55.7–55.9, 55.7f
- Magnetron-sputtered multilayer Laue lenses (MLLs), 42.5–42.7, 42.6f–42.8f
- Magnification, by gratings and monochromators, 38.7
- Maier-Saupe mean-field coupling constant, 8.25
- Mammography, 54.8, 59.3–59.4
- Manchester (biphase) coding, 20.9, 20.9f
- Markov random process approximation, for beam wander, 3.31, 3.32
- Masks, for extreme ultraviolet lithography, 34.2, 34.6
- Master oscillator power amplifier (MOPA) systems, 25.2
- free space designs, 25.13–25.15, 25.15f
 - monolithic designs, 25.16–25.17, 25.16f
 - nanosecond designs, 25.31–25.32
 - narrow linewidths, 25.30
 - ultrashort systems, 25.33
- MathCad program, 5.39
- Mathematica program, 5.39
- Maximum frequency deviation, of electro-optic modulators, 7.35
- Maximum likelihood sequence estimations (MLSEs), 21.26
- Maximum refractive index, for PCFs, 11.9–11.10
- Maximum tolerable input jitter (MTIJ), 15.16
- Maximum-likelihood expectation maximization (MLEM), 32.2
- Maxwell-averaged Gaunt factors, 56.10
- Maxwell-Boltzmann distribution, for EDFA spectra, 14.4
- Maxwell's equations:
- and Bremsstrahlung radiation, 56.8
 - for fiber-based couplers, 16.2, 16.3
 - and photonic crystal fibers, 11.3, 11.6, 11.20
 - and wave propagation, 5.8
- Mean-field theory, of liquid crystals, 8.22–8.23
- Mechanical switches, for networking, 18.4, 18.5, 18.5f, 18.11, 18.11f, 18.12f
- Media access control (MAC), of FDDI, 23.3
- Media interface connectors (MICs), 23.3
- Medical imaging, 31.1–31.10
- applications of, 31.9, 31.10
 - digital displays, 31.8–31.9, 31.8f–31.10f
 - digital tomosynthesis, 31.7–31.8
 - and inverse Compton x-ray sources, 59.3–59.4
 - and polycapillary x-ray optics, 53.14–53.16, 53.15f–53.16f
 - radiography, 31.1–31.4, 31.2f–31.4f
 - tomography, 31.1, 31.5–31.7, 31.5f–31.7f
 - x-ray detectors for, 61.2

- Medicine, nuclear, 32.1–32.4
- Mellin transforms, 5.16
- Mesh topologies, of WDM networks, 21.6
- Mesospheric sodium lasers, 5.32–5.34, 5.33f
- Metallization, for SOAs, 19.16, 19.16f
- Metal-semiconductor-metal (MSM) photoconductive detectors, 13.63, 13.73
- Meteorological optics, 3.40–3.43, 3.41f–3.43f
- Metrology:
 - and magnetron-sputtered MLLs, 42.6–42.7, 42.7f, 42.8f
 - scatterometers in, 1.16
 - surface figure, 46.3–46.6, 46.5f
 - surface finish, 46.2
 - x-ray mirror, 46.1–46.12
 - history of, 46.1–46.2
 - profile analysis considerations, 46.6–46.12, 46.7f, 46.10f
 - surface figure metrology, 46.3–46.6, 46.5f
 - surface finish metrology, 46.2
- Metropolitan area networks (MANs), 9.14, 21.7
- Michelson interferometers, 17.8–17.9, 17.8f
- Microbunches, of electrons, 58.1
- Microcalorimeter detectors, 29.9–29.11, 29.11f, 60.9, 60.9t
- Microchannel plate (MCP) detectors, 63.34
- Microchannel plates (MCPs), 49.2, 49.3f, 49.5f, 50.7, 60.7
- Micro-electromechanical systems (MEMS)
 - mirrors and switches, 18.8, 18.8f, 18.11, 18.12f
- Microfocus x-ray fluorescence (MXRF):
 - with doubly curved crystal diffraction, 29.6–29.7, 29.8f–29.9f
 - monocapillary, 29.4
 - polycapillary, 29.4–29.6, 29.5f, 29.6f
 - ultrahigh resolution, 29.9–29.11, 29.9f–29.11f
- Microfocusing, with refractive x-ray lenses, 37.7–37.8
- Micro-optics-based components, for networking, 18.1–18.12
 - attenuators, 18.2, 18.9
 - beam splitters, 18.6, 18.6f
 - circulators, 18.3, 18.3f, 18.10
 - directional couplers, 18.2, 18.3, 18.3f, 18.9, 18.9f
 - Faraday rotators, 18.7, 18.7f
 - filters, 18.6
 - gratings, 18.5–18.6, 18.6f
 - Micro-optics-based components, for networking (Cont.):
 - GRIN-rod lenses, 18.7, 18.8, 18.8f
 - isolators, 18.3, 18.10, 18.10f
 - mechanical switches, 18.4, 18.5, 18.5f, 18.11, 18.11f, 18.12f
 - MEMS mirrors and switches, 18.8, 18.8f, 18.11, 18.12f
 - multiplexers/demultiplexers/duplexers, 18.4, 18.4f, 18.10–18.11
 - network functions, 18.2–18.5
 - polarizers, 18.7, 18.7f
 - power splitters, 18.2–18.3, 18.2f, 18.9, 18.9f
 - prisms, 18.5, 18.5f
- Micro-pore optics, 49.1–49.6
- Microscopes and microscopy:
 - coherent diffraction, 27.4–27.5, 27.4f, 27.5f
 - Nomarski microscope, 46.2
 - scanning electron
 - and magnetron-sputtered MLLs, 42.6–42.7, 42.7f, 42.8f, 42.13, 42.13f
 - and x-ray spectral detection, 62.1–62.3, 62.2f
 - x-ray, 37.6
- Microsource devices, 28.7
- Microstrip detectors, 63.32–63.33
- Microstructured optical arrays (MOAs),
 - adaptive, 50.7–50.8, 50.8f
- Mid-wave infrared (MWIR) AOTFs, 6.42
- Mie scattering, 3.12, 3.16–3.18, 3.17f–3.19f
- Mie theory, 11.7
- Miesowicz viscosity coefficients, 8.24
- Mirages, 3.42–3.43, 3.42f, 3.43f
- Mirror reflectivity, for VCSELs, 13.44
- Mirror surface roughness:
 - and grazing-incidence neutron optics, 64.2–64.3
 - and Wolter x-ray optics, 47.3–47.5, 47.4f
- Mirrors:
 - bimorph, 50.4, 50.4f, 50.6f
 - Bragg, 13.28, 13.44
 - deformable, 5.4, 5.4f, 5.37–5.38, 5.37f, 5.38f
 - double-bounce Wolter, 52.4
 - external, 13.32–13.33, 13.33f
 - Goebel, 26.10
 - high-reflectivity, 41.7–41.8, 41.8f
 - Kirkpatrick-Baez, 44.4, 44.4f, 63.21, 64.5–64.6, 64.5f
 - in micro-electromechanical systems, 18.8, 18.8f, 18.12f

Mirrors (*Cont.*):

- for neutron optics, 63.20–63.21, 63.21f
 - nonlinear optical loop, 20.22
 - point-spread function of, 64.2
 - polarizing, 63.28
 - semiconductor laser amplifier loop optical, 20.22
 - semiconductor saturable absorber, 25.3, 25.32
 - and SHADOW code, 35.4, 35.5
 - Wolter configurations, 52.4, 64.6, 64.6f
 - (*See also* X-ray mirrors)
- Modal approach, to wavefront error correction, 4.35
- Modal dispersion, 15.9
- Modal filtering, 11.12–11.13, 11.13f
- Modal gain per unit length, 13.7, 13.8
- Modal noise, 15.16–15.17
- Mode field adaptors (MFAs), 25.2, 25.17–25.18
- Mode field diameter (MFD), 25.2
- Mode hopping, 13.13
- Mode matching, in mode field adaptors, 25.17, 25.17f
- Mode partition noise, 20.1
- for fiber optic communication links, 15.11–15.13, 15.11f
 - for laser diodes, 13.19–13.20, 13.20f
- Mode transformers, photonic crystal fibers
- and, 11.26–11.27, 11.27f
- Mode-locked fiber lasers, 25.32–25.33
- Mode-locked lasers, 20.15–20.17, 20.16f
- Moderators, for neutron optics, 63.12, 63.14–63.15
- Modified chemical vapor deposition (MCVD), 25.2, 25.21, 25.26, 25.28
- MODTRAN program, 3.24
- Modulated grating (MG) reflectors, 13.36
- Modulation:
- amplitude, 7.22–7.24, 7.23f, 7.24f
 - cross-gain, 19.12, 19.13f, 19.27, 19.29–19.30, 19.29f, 19.32, 19.35–19.36
 - cross-phase
 - in optical fibers, 10.3–10.4
 - and SOAs, 19.13, 19.30–19.32, 19.31f, 19.33f, 19.35–19.36
 - and solitons, 22.5, 22.13–22.15
 - in WDM networks, 21.19
 - degree of, for electro-optic modulators, 7.35–7.36
 - depth-of-amplitude, 7.22

Modulation (*Cont.*):

- depth-of-phase, 7.19
 - digital, 6.33
 - direct, 20.17–20.18, 20.17f
 - frequency, 7.24–7.25, 7.25f
 - longitudinal spatial, 6.12
 - in OTDM communication networks
 - direct and indirect, 20.17–20.18, 20.17f
 - external, 20.18–20.20, 20.19f, 20.20f
 - percent, 7.35
 - phase, 7.18–7.20
 - bulk electro-optic modulators, 7.18–7.20
 - by lithium niobate modulators, 13.51
 - polarization, 7.20–7.22, 7.20f, 7.21f
 - pulse code, 20.8
 - self-phase
 - in optical fibers, 10.3–10.4
 - and solitons, 22.3–22.4, 22.3f, 22.4f
 - in WDM networks, 21.18–21.19, 21.19f
 - spatial light, 6.4, 6.9
 - transverse spatial, 6.11–6.12, 6.23, 6.30, 6.31
 - in WDM networks, 21.27–21.36
 - basic concepts, 21.27–21.29, 21.28f–21.30f
 - carrier-suppressed return-to-zero and duobinary, 21.30–21.33, 21.31f, 21.32f
 - comparisons of, 21.36, 21.36t, 21.37t
 - DPSK and DQSK, 21.33–21.36, 21.33f–21.35f, 21.36t, 21.37t
- Modulation bandwidth, electro-optic modulators, 7.34
- Modulation efficiency, of electro-optic modulators, 7.36
- Modulation error ratio (MER), 15.4–15.5
- Modulation formats, for WDM networks, 21.27–21.36
- basic concepts, 21.27–21.29, 21.28f–21.30f
 - carrier-suppressed return-to-zero and duobinary, 21.30–21.33, 21.31f, 21.32f
 - comparisons of, 21.36, 21.36t, 21.37t
 - DPSK and DQSK, 21.33–21.36, 21.33f–21.35f, 21.36t, 21.37t
- Modulation instability, 10.3
- Modulation response, of laser diodes, 13.16–13.17, 13.16f
- Modulation transfer functions (MTFs):
- for acousto-optic modulators, 6.32
 - for polycapillary x-ray optics, 53.15, 53.16f
 - for Schwarzschild objectives, 51.1–51.2, 51.2f
 - and SPECT imaging, 32.3
 - for x-ray detectors, 61.3

Modulators:

- acousto-optic, 6.23*t*, 6.31–6.35, 6.32*f*, 6.34*t*
- acousto-optic frequency shifters, 6.35
- and Bragg diffraction, 6.4, 6.6, 6.7, 6.14
- image (scophony), 6.34–6.35
- multi-mode interference, 13.2, 13.35
- principle of operation, 6.32, 6.32*f*
- band-filling, 13.62
- electroabsorption
 - in fiber optic systems, 13.55–13.60, 13.56*f*, 13.60
 - in OTDM networks, 20.18, 20.20, 20.20*f*
- electro-optic, 7.1–7.39, 13.61, 20.18–20.19, 20.19*f*
- applications for, 7.36–7.39
- bulk modulators, 7.16–7.28, 7.21*f*, 7.24*f*–7.28*f*
- crystal optics and the index ellipsoid, 7.3–7.7, 7.4*f*–7.6*f*, 7.8*f*–7.10*f*
- and electro-optic effect, 7.6–7.16, 7.8*t*–7.10*t*, 7.14*f*, 7.16*f*
- electro-optic sampling, 7.36–7.37, 7.37*f*–7.38*f*
- and Euler angles, 7.39
- in fiber optic systems, 13.61
- geometries, 7.16–7.18, 7.17*f*
- laser mode locking, 7.38–7.39
- light propagation in, 7.3
- materials, 7.33–7.34
- in OTDM networks, 20.18–20.19
- performance criteria, 7.34–7.36
- sensors, 7.38
- traveling wave modulators, 7.28–7.30, 7.29*f*
- waveguide or integrated-optic modulators, 7.30–7.32, 7.31*f*–7.33*f*
- electrorefractive, 13.61–13.62
- integrated-optic, 7.3, 7.30–7.32, 7.31*f*–7.33*f*
- interferometric Mach-Zehnder, 13.51–13.52, 13.54–13.55, 13.63
- lithium niobate, 13.2, 13.48–13.55, 13.49*f*
- Mach-Zehnder
 - interferometric, 13.51–13.52, 13.54–13.55, 13.54*f*, 13.63
 - in WDM networks, 21.30, 21.31*f*, 21.32*f*
- Nipi, 13.62
- semiconductor interferometric, 13.63
- separate confinement heterostructure for, 13.4*f*, 13.5
- waveguide, 7.30–7.32, 7.31*f*–7.33*f*, 13.56, 13.57*f*

- Molecular absorption, 3.12–3.15, 3.13*f*
- Molecular absorption line database, 3.22–3.23, 3.22*f*, 3.23*f*
- Molecular emission, 3.18, 3.20, 3.20*f*
- Molecular gases, in standard atmosphere, 3.6, 3.7*t*, 3.8*f*–3.9*f*
- Molecular imaging, 32.1
- Molecular spectroscopy, 2.5–2.6, 2.6*f*, 2.7*f*
- Molecular targets, for SPECT imaging, 32.1
- Monin-Obukhov similarity theory, 3.31
- Monocapillary x-ray optics, 28.5, 52.1–52.6, 52.2*f*–52.5*f*, 52.2*t*
- Monochromatic imaging, in polycapillary x-ray optics, 53.16–53.17, 53.17*f*
- Monochromators:
 - Bragg reflection, 39.4, 39.5
 - Bragg reflections in, 39.1, 63.24
 - crystal
 - and bent crystals, 39.1–39.6, 39.2*t*, 39.3*f*, 39.5*f*–39.6*f*
 - in neutron optics, 63.23–63.25
 - Dragon* systems of, 38.3
 - Grasshopper*, 38.3
 - Johansson bent/ground focusing, 39.5
 - and SHADOW code, 35.4, 35.4*f*
 - spherical-grating, 38.3
 - synchrotron radiation, 39.6
 - toroidal-grating, 38.3
 - in VUV and soft x-ray region, 38.1–38.8
 - diffraction properties, 38.1–38.3, 38.2*f*
 - dispersion properties, 38.6–38.7
 - efficiency of, 38.8
 - focusing properties, 38.3–38.6, 38.3*f*, 38.4*t*–38.5*t*
 - resolution properties, 38.7
 - x-ray, 30.1–30.4, 50.6–50.7
 - and x-ray diffraction, 28.3, 28.4
- Monolithic fiber laser resonators, 25.16
- Monolithic tunable lasers, 13.33–13.36, 13.34*f*–13.36*f*
- Mosaic crystals, 39.2
- Mt. Pinatubo, 3.10, 3.18, 3.39
- Mt. Wilson telescope, 5.27
- Mueller matrices, 1.14, 19.18–19.19
- Multichannel Bragg cells (MCBC), 6.30–6.31
- Multicore fibers, 25.2, 25.22
- Multidomain vertical alignment (MVA) cells, 8.25, 8.27–8.28
- Multi-energy imaging, 54.9–54.10
- Multi-fiber push on (MPO) connectors, 23.8

- Multifoil Kirkpatrick-Baez optics, 48.3–48.4, 48.3f
- Multifoil lobster-eye optics, 48.2–48.4, 48.2f, 48.3f
- Multifoil optics (MFO), 48.1–48.4, 48.2f, 48.3f
- Multilayer Laue lenses (MLLs):
 with curved interfaces, 42.14, 42.15f
 and hard x-rays, 42.1–42.17
 history of, 42.2–42.4, 42.2f, 42.3f
 instrumental beamline arrangement
 and measurements for, 42.9–42.10, 42.9f–42.12f
 limitations of, 42.15–42.17, 42.16f–42.17f
 with magnetron-sputtered MLLs, 42.5–42.7, 42.6f–42.8f
 on MLLs with curved interfaces, 42.14, 42.15f
 Takagi-Taupin calculations for, 42.12–42.14
 volume diffraction calculations for, 42.4–42.5, 42.5f
 with wedged MLLs, 42.12–42.13, 42.13f, 42.14f
 history of, 42.2–42.4, 42.2f, 42.3f
 instrumental beamline arrangement and measurements of, 42.9–42.10, 42.9f–42.12f
 limitations of, 42.15–42.17, 42.16f–42.17f
 magnetron-sputtered, 42.5–42.7, 42.6f–42.8f
 Takagi-Taupin calculations for, 42.12–42.14
 volume diffraction calculations for, 42.4–42.5, 42.5f
 wedged, 42.12–42.13, 42.13f, 42.14f
 and x-ray/neutron optics, 26.10
- Multilayers (reflective coatings), 41.1–41.10
 and calculation of multilayer properties, 41.3–41.4
 for diffractive imaging, 41.9–41.10, 41.10f
 fabrication and performance of, 41.4–41.9, 41.5f, 41.6t, 41.7t, 41.8f
 periodic, 42.5–42.6, 42.6f
 properties of, 41.1–41.3, 41.2f
 and x-ray diffraction, 28.5, 28.5f
- Multilongitudinal mode Fabry-Perot laser, 9.7–9.8
- Multimode fibers, for E-LEDs, 13.40
- Multimode interference (MMI) modulators, 13.2, 13.35
- Multimode interferometric Mach-Zehnder modulators, 13.54–13.55, 13.54f
- Multipath interference noise, 15.13–15.14
- Multiple Mirror Telescope (MMT), 5.5
- Multiple quantum well (MQW) lasers, 13.24–13.25
- Multiple quantum wells (MQW), 20.20
- Multiple scattering, 3.21, 3.21f
- Multiple-beam Bragg diffraction (MBD), 43.6–43.8, 43.7f
- Multiple-Bragg-beam interference, 43.1
- Multiplexers and multiplexing:
 for networking, 18.4, 18.10–18.11, 18.11f
 optical add/drop, 21.2, 21.8, 21.8f, 21.9f
 in OTDM networks, 20.1, 20.3–20.12, 20.5f–20.11f, 20.13f
 parallel, 20.12
 serial, 20.12
 time-division, 9.12, 20.3, 21.3
 (See also Wavelength division multiplexing)
- Multiplication, of avalanche photodiodes, 13.71–13.72
- Multiquantum wells (MQWs), 2.11, 19.7, 19.11f, 19.21
- Multiwire proportional counters (MWPCs), 63.31–63.32
- Mutual coherence function (MCF), 4.4, 4.7, 4.10
- Nanofocusing, of hard x-rays (see Hard x-rays, nanofocusing of)
- Nanofocusing lenses (NFLs), 37.8–37.11, 37.9f, 37.10f
- Nano-optic-measuring (NOM) machine, 46.5
- Nanosecond fiber systems, 25.30–25.32
- Narrow linewidth fiber lasers, 25.29–25.30
- Natural guide star (NGS) sensing, 5.21
- Natural line width, of spectral lines, 56.4–56.5
- Near Bragg diffraction, 6.8–6.9, 6.12
- Near field, 4.10
- Near-field diffraction, 27.2–27.3, 27.3f
- Negative core-cladding index difference, of photonic crystal fibers, 11.14–11.17, 11.14f–11.17f
- Negative hydrogen (H^-) ions, 2.3
- Negative orders of radiation, 40.1
- Nematic phase, of liquid crystals, 8.8, 8.11f, 8.11
- Neodymium-doped fibers, 25.23–25.24, 25.23t
- Networking, micro-optics-based components
 for, 18.1–18.12
 attenuators, 18.2, 18.9
 beam splitters, 18.6, 18.6f

- Networking, micro-optics-based components
for (*Cont.*):
circulators, 18.3, 18.3f, 18.10
directional couplers, 18.2, 18.3, 18.3f, 18.9, 18.9f
Faraday rotators, 18.7, 18.7f
filters, 18.6
gratings, 18.5–18.6, 18.6f
GRIN-rod lenses, 18.7, 18.8, 18.8f
isolators, 18.3, 18.10, 18.10f
mechanical switches, 18.4, 18.5, 18.5f, 18.11,
18.11f, 18.12f
MEMS mirrors and switches, 18.8, 18.8f,
18.11, 18.12f
multiplexers/demultiplexers/duplexers, 18.4,
18.4f, 18.10–18.11
network functions, 18.2–18.5
polarizers, 18.7, 18.7f
power splitters, 18.2–18.3, 18.2f, 18.9, 18.9f
prisms, 18.5, 18.5f
(*See also related topics, e.g.: Communication
networks and systems*)
- Neutron attenuation, 63.11–63.12
Neutron collimation, 63.15–63.16
Neutron filters, 63.18–63.19, 63.18t, 63.19f
Neutron gravity spectrometer, 63.21f
Neutron guides, 63.15–63.18, 63.17f
Neutron optics, 63.3–63.34
detection in, 63.31–63.34
devices for, 63.15–63.19, 63.17f, 63.18t, 63.19f
diffraction and interference in, 63.23–63.27,
63.27f
grazing-incidence, 64.1–64.7
diffractive scattering and mirror surface
roughness, 64.2–64.3
imaging focusing optics, 64.3–64.7,
64.4f–64.7f
materials of optical elements, 64.7
total external reflection, 64.1–64.2
and neutron physics, 63.3–63.5
and neutron sources, 63.12–63.15, 63.13f
polarization techniques for, 63.27–63.30,
63.30f
refraction and reflection in, 63.19–63.23,
63.21f, 63.23f
scattering lengths and cross sections,
63.5–63.12, 63.6t, 63.10t
neutron attenuation, 63.11–63.12
scattering length density, 63.9–63.11, 63.11f
and x-ray optics, 26.5–26.11, 26.8f, 26.9f,
26.11f, 36.2f
- Neutron physics, 63.3–63.5
- Neutron polarization, 63.27–63.29, 63.30f
Neutron scintillators, 63.33
Neutron zone plates, 63.25
- Neutrons:
epithermal, 63.18
MCP detectors for, 63.34
scattering cross sections of, 63.6–63.9, 63.6t,
63.10t
scattering length densities of, 63.9–63.11, 63.11f
scattering lengths of, 63.5–63.9, 63.6t, 63.10t
thermal, 63.3
total integrated scatter of, 64.2–64.3
(*See also Neutron optics*)
- NeXT spacecraft, 47.10
Nipi modulators, 13.62
- Noise:
ASE, 19.3, 19.4f, 19.9, 19.18, 19.24, 19.35
avalanche photodiodes, 13.72–13.73
of EDFAs, 14.6
in fiber optic communication links,
15.11–15.14, 15.16–15.17
of laser diodes, 13.18–13.24, 13.20f
modal, 15.16–15.17
mode partition, 13.19–13.20, 15.11–15.13, 15.11f
multipath interference, 15.13–15.14
phase (linewidth), 13.20–13.21, 13.21f
of pin diodes, 13.70–13.71
relative intensity, 13.18–13.19, 13.19f, 15.14
of SOAs, 19.3, 19.4f, 19.9, 19.18, 19.20,
19.24, 19.35
[*See also Signal-to-noise ratio (SNR)*]
- Noise equivalent BDSF (NEBDSF), for
scatterometers, 1.6, 1.8, 1.12–1.13
Noise equivalent power (NEP), 1.13, 13.71
Noise figure, of EDFAs, 14.6
Nomarski differential interference, 46.4
Nomarski microscope, images from, 46.2
- Noncritical phase-matching acousto-optic
tunable filters (NPM AOTFs), 6.37, 6.38f,
6.39–6.42
angle of deflection, 6.40
angular aperture, 6.41
for long-infrared, 6.42
for mid-infrared, 6.42
optical throughput, 6.41
performance of, 6.42–6.44, 6.43t
resolution, 6.40
sidelobe suppression, 6.41–6.42
transmission and drive power, 6.41
tuning relation, 6.39–6.40
for ultraviolet, 6.42

- Nonfused fiber couplers, **25.10**
- Nonlinear acoustic (NA) interaction, in
 - acousto-optic devices, **6.30**
- Nonlinear distortion, **9.17**
- Nonlinear effects:
 - of fiber lasers, **25.6**
 - four-wave mixing, **10.2, 10.9–10.11, 10.11f**
 - in optical fibers, **10.1–10.12**
 - self- and cross-phase modulation, **10.3–10.4**
 - stimulated Brillouin scattering, **10.1, 10.7–10.9**
 - stimulated Raman scattering, **10.1, 10.4–10.7, 10.5f**
- Nonlinear length, of solitons, **22.3**
- Nonlinear optical loop mirrors (NOLM), **20.22**
- Nonlinear optics, WDM networks and,
 - 21.18–21.20, 21.19f, 21.20f**
- Nonlinear Schrödinger equation (NSE), **10.4, 22.2**
- Non-return-to-zero differential-phase-shift-keying (NRZ-DPSK) format, **21.28**
- Non-return-to-zero (NRZ) format:
 - in OTDM networks, **20.8, 20.9f, 20.12**
 - in WDM networks, **21.16, 21.29f, 21.32f**
- Non-return-to-zero on-off keying (NRZ-OOK),
 - 21.34, 21.36t, 21.37t**
- Nonzero dispersion-shifted fiber (NZDSF),
 - 21.21, 21.21f, 21.34f**
- Nuclear imaging, **53.17, 53.18, 53.18f**
- Nuclear medicine, **32.1–32.4**
- Numerical aperture (NA), **9.4, 25.2, 25.18, 42.2**
- NuSTAR spacecraft, **47.6, 47.7, 47.10**
- Nyquist frequency, **27.4**
- Nyquist frequency power, **46.8–46.9, 46.11**
- Nyquist noise, **13.70**

- Objectives, Schwarzschild, **26.10, 51.1–51.3, 51.2f–51.4f**
- Observatories:
 - Chandra, **33.2–33.4, 33.3t, 44.4, 44.10, 47.1, 47.4f, 47.5, 47.10, 64.7**
 - Constellation-X, **33.4**
 - Einstein, **44.4, 44.10, 47.1, 47.5**
 - ROSAT, **47.5**
 - SOHO, **41.3**
 - Suzaku, **33.3–33.4, 33.3t**
 - TRACE, **41.3**
 - W. M. Keck, **5.27**
 - XMM-Newton, **33.3, 47.2, 47.4f, 47.6, 47.6f**
 - XMM-Newton observatory, **33.3t**
 - x-ray, **33.1–33.4, 33.3t**

- On-axis aberrations, **45.6–45.8**
- On-axis optics, **64.3**
- On-axis tangential phase matching, **6.25–6.26**
- On-blaze condition, **38.2**
- $1 \times N$ power splitters, **16.1, 16.4**
- 1D profilometry, **46.6**
- On-off keying (OOK), in WDM networks,
 - 21.29, 21.30, 21.34, 21.36t, 21.37t**
- Open fiber control (OFC), for Fibre Channel standard, **23.4**
- Optical absorption, measurements of, **2.2–2.13, 2.4f, 2.6f–2.8f, 2.10f, 2.12f**
- Optical add/drop multiplexers (OADMs), **21.2, 21.8, 21.8f, 21.9f, 21.12**
- Optical amplifiers:
 - communications applications for, **9.14**
 - semiconductor vs. fiber, **9.13–9.14**
 - in WDM networks, **21.37–21.44, 21.37f**
 - EDFA, **21.38–21.41, 21.38f–21.42f**
 - Raman, **21.42–21.44, 21.42f–21.44f**
 - (*See also* Optical fiber amplifiers)
- Optical burst switching (OBS), **21.11**
- Optical circulators, **17.8f, 17.9**
- Optical clock recovery, **20.21–20.22, 20.21f**
- Optical crossconnects (OXCs), **21.5f, 21.6, 21.8, 21.8f, 21.10, 21.10f**
- Optical damage, **13.54, 25.6**
- Optical electric field, bulk modulators and, **7.18**
- Optical fiber amplifiers, **14.1–14.11**
 - categories and features of, **14.1–14.2, 14.2t**
 - erbium-doped
 - energy levels, **14.4**
 - fast power transients, **21.39–21.41, 21.39f, 21.40f**
 - gain flattening, **14.6–14.7, 21.38–21.39, 21.39f**
 - gain formation, **14.4–14.5, 14.5f**
 - gain peaking, **21.38, 21.38f**
 - noise, **14.6**
 - pump wavelength options, **14.5–14.6**
 - semiconductor amplifiers vs., **9.13, 9.14, 14.1, 14.2t**
 - static gain dynamic and channel power equalization, **21.41, 21.41f–21.42f**
 - in WDM networks, **21.2–21.3, 21.2f**
- erbium/ytterbium-doped, **14.7–14.8**
- parametric, **14.10–14.11**
- praseodymium-doped, **14.7**
- Raman fiber, **14.8–14.9, 14.8f, 14.10f**
- rare-earth-doped, **14.2–14.4, 14.3f**
- ytterbium-doped, **14.7**

- Optical fiber sensors, **24.1–24.13**
 - extrinsic Fabry-Perot interferometric, **24.2–24.4, 24.2f, 24.3f**
 - fiber Bragg grating, **24.5–24.8, 24.6f–24.7f**
 - intrinsic Fabry-Perot interferometric sensors, **24.4–24.5, 24.4f**
 - long-period grating sensors, **24.8–24.13, 24.9f–24.12f, 24.11t, 24.13t**
- Optical fibers:
 - in communication systems, **9.3–9.17**
 - analog transmission, **9.15–9.17**
 - bit rate, **9.12**
 - distance limits, **9.12–9.13**
 - fiber for, **9.4–9.7, 9.5f, 9.6f**
 - fiber-optic networks, **9.14–9.15**
 - optical amplifiers, **9.13–9.14**
 - photodetectors, **9.8**
 - receiver sensitivity, **9.8–9.11**
 - repeater spacing, **9.12–9.13**
 - technology, **9.4–9.8**
 - transmitting sources, **9.7–9.8**
 - infrared, **12.1–12.13**
 - applications, **12.13**
 - categories and properties of, **12.1–12.3, 12.2f, 12.2t, 12.3t**
 - crystalline, **12.2t, 12.3t, 12.7–12.10, 12.8f, 12.10f**
 - heavy-metal oxide glass in, **12.2t–12.4t, 12.3–12.7, 12.5f–12.7f**
 - hollow waveguides, **12.2t, 12.3t, 12.10–12.13, 12.12f**
 - nonlinear effects in, **10.1–10.12**
 - four-wave mixing, **10.2, 10.9–10.11, 10.11f**
 - self- and cross-phase modulation, **10.3–10.4**
 - stimulated Brillouin scattering, **10.1, 10.7–10.9**
 - stimulated Raman scattering, **10.1, 10.4–10.7, 10.5f**

[See also related topics, e.g.: Photonic crystal fibers (PCFs)]

- Optical holeburning (OHB), **2.13, 2.14f**
- Optical insertion loss, of electro-optic modulators, **7.36**
- Optical Kerr effect, **7.11**
- Optical lithography (OL), **34.1**
- Optical mode conditioners, **23.7**
- Optical power dependence, of electroabsorption modulators, **13.59**
- Optical power penalties, for fiber optic communication links, **15.8–15.17, 15.10f, 15.11f**
- Optical signal-to-noise ratio (OSNR):
 - of SOAs, **19.24–19.27**
 - for WDM networks, **21.20, 21.28, 21.34**
- Optical spectrum analyzers (OSAs), **19.18**
- Optical strength, of turbulence (C_n^2), **5.6–5.8, 5.7f, 5.8f**
- Optical tank circuits, **20.21, 20.21f**
- Optical throughput, of NPM AOTFs, **6.41**
- Optical time-division multiplexed (OTDM) communication networks:
 - and all-optical switching for demultiplexing, **20.22, 20.23f**
 - device technology, **20.12–20.24**
 - direct and indirect modulation in, **20.17–20.18, 20.17f**
 - external modulation in, **20.18–20.20, 20.19f, 20.20f**
 - history of, **20.3**
 - multiplexing in, **20.1, 20.3–20.12, 20.5f–20.11f, 20.13f**
 - and optical clock recovery, **20.21–20.22, 20.21f**
 - serial vs. parallel, **20.12, 20.13f**
 - transmitters in, **20.12–20.17, 20.14f–20.16f**
 - ultrahigh-speed OTDM, **20.23–20.24, 20.24f**
 - and WDM, **21.2**
- Optical transfer function (OTF):
 - and adaptive optics, **5.19–5.20, 5.20f**
 - and atmospheric turbulence, **4.3, 4.6–4.7**
 - of systems with annular pupils, **4.10–4.13, 4.12f, 4.13f**
- Optical transmission, atmospheric, **3.22–3.26, 3.22f–3.27f**
- Optical-electrical field overlap parameter, of electro-optic effect, **13.50**
- Optically detected magnetic resonance (ODMR), **2.23–2.24, 2.24f**
- Optically rotated tangential phase matching, **6.26f, 6.27**
- Order selecting aperture (OSA), of zone plates, **40.4, 40.5**
- Ordered dye-doped polymers, **7.34**
- Ordered subsets expectation maximum (OSEM), in SPECT imaging, **32.2**
- Organic crystals, **7.33–7.34**
- Oscillator strength, **56.3–56.4**

- Oscillators:
 and fiber lasers, 25.13, 25.14f,
 25.30–25.33
 high-power USP, 25.32–25.33
 local, 9.13
 parametric, 11.23, 11.24
 Q-switched, 25.30–25.31
 voltage-controlled, 20.11, 20.11f
 [*See also* Master oscillator power amplifier
 (MOPA) systems]
Outer scale of turbulence, 4.7
Out-of-plane profile, of emitted light, 13.11
Outside vapor deposition (OVD), 25.2,
25.26
Overlapping integral, of acousto-optic
interaction, 6.15
- Pacific Northwest National Laboratory
program, 3.26
Packaging, of SOAs, 19.17, 19.17f
Packet-switched networks, 21.7, 21.10–21.11,
21.10f–21.11f
Parabolic reflectors, for neutron beams, 64.3,
64.4, 64.4f
Parallel beams, and x-ray tube sources, 54.16
Parallel multiplexing, 20.12
Parallel-hole collimators, 32.3
Parametric amplifiers and oscillators, 11.23,
11.24, 14.2, 14.10–14.11
Parseval's theorem, 46.8
Partial coherence length, 4.10
Particle-induced x-ray emission (PIXE), 29.4
Particulate matter, in standard atmosphere,
3.6–3.7, 3.9–3.11, 3.10f, 3.11f
Passbands, 20.1
Passive mode locking, 20.17
Path function, for gratings and monochromators,
38.3–38.5, 38.3f, 38.4t–38.5t
Patterned vertical alignment (PVA) cells,
8.27–8.28, 8.28f
Pauli exclusion principle, 56.2
Pearson IV function, 56.7–56.8
Pearson VII function, 56.7, 56.8
Pendellösung interference, 63.26
Perfect crystal interferometers, 63.26–63.27,
63.27f
Periodogram estimator, for 1D profiles, 46.8
Periodic multilayers, of MLLs, 42.5–42.6,
42.6f
PETRA III (synchrotron source), 37.10
- Phase aberration function:
 correction of, 4.28–4.30
 modal expansion of, 4.17–4.20, 4.17f–4.18f,
 4.19t, 4.20t
Phase array beam steering, 6.27–6.29, 6.28f
Phase fluctuations, adaptive optics and,
5.18–5.19
Phase mask method, of FBG fabrication,
17.5–17.6, 17.6f, 17.8, 24.7, 24.7f
Phase matching:
 for acousto-optic devices, 6.9–6.12, 6.10f
 birefringent tangential, 6.25
 on-axis tangential, 6.25–6.26
 optically rotated tangential, 6.26f, 6.27
 tangential, 6.12, 6.13, 6.17, 6.25–6.27,
 6.26f
Phase modulation, 7.18–7.20, 13.51 (*See also*
 Cross-phase modulation; Self-phase
 modulation)
Phase modulation index, 7.19
Phase noise, 9.8, 13.1, 13.20–13.21, 13.21f
Phase plates, for circular polarization,
43.5–43.6, 43.5f
Phase retarders, 41.9, 43.6
Phase structure function, 4.5, 5.9, 5.10
Phase transitions, of liquid crystals, 8.13–8.14,
8.13f, 8.14f
Phase velocity indices of refraction, 7.15–7.16,
7.16f
Phase zone plates, 40.5–40.7, 40.7f
Phase-locked loops (PLLs), 20.11, 20.11f
Phase-measuring interferometers
(PMIs), 46.2
Phase-space acceptance, by gratings and
monochromators, 38.7
Phosphates, for fiber lasers, 25.27t, 25.28
Phosphor x-ray detectors, 60.7–60.8, 60.10t
Photoabsorption, 36.1
Photodetectors, 9.8
Photodiodes:
 avalanche, 13.63, 13.71–13.73
 pin, 13.64–13.66
 resonant, 13.65
 Schottky, 13.63, 13.73
 unitraveling-carrier, 13.68–13.69, 13.68f
Photographic film, neutron detection with,
63.33, 63.34
Photomultiplier tubes (PMT), 31.5
Photonic bandgaps (PBGs), 11.2–11.3, 11.8,
11.8f, 11.10f, 11.11, 11.11f, 11.14

- Photonic crystal fibers (PCFs), 11.1–11.28, 25.2
 all solid-core, 25.21
 Bragg fibers, 11.4
 in cladding, 11.7–11.11, 11.8f–11.11f
 cleaving and splicing of, 11.26
 design and fabrication of, 11.4–11.6, 11.5f
 endlessly single-mode, 11.12, 11.13, 11.21, 11.21f
 in fiber lasers, 25.19f, 25.20–25.21, 25.27
 and guidance
 attenuation mechanisms, 11.19–11.22, 11.20f, 11.21f
 birefringence, 11.17
 core-cladding index difference, 11.12–11.17, 11.12f–11.17f
 group velocity dispersion, 11.18, 11.18f, 11.19f
 Kerr nonlinearities, 11.22–11.24, 11.24f
 resonance and antiresonance, 11.12
 scattering, 11.24–11.26, 11.25f
 history of, 11.2–11.4, 11.3f
 in-fiber devices for, 11.27, 11.28
 mode transformers, 11.26–11.27, 11.27f
 modeling and analysis of, 11.6–11.7
- Photonic integrated circuits (PICs), 19.1, 19.36
- Photons:
 energies of, 36.7t–36.8t
 lifetimes of, 20.1
- Photorefractivity, of lithium niobate modulators, 13.54
- Photoresist, for extreme ultraviolet lithography, 34.2, 34.6
- Photosensitivity, of fiber Bragg gratings, 17.2–17.3
- Photostimulable phosphors, 63.34
- Physical layer (PHY) implementation, of FDDI connectors, 23.2–23.3
- Physical vapor deposition (PVD), 61.6
- Picosecond (unit), 20.1
- Pigtail connection, fiber, 13.8
- Pin diodes, 13.2, 13.63–13.71, 13.66t
 dark current, 13.69
 geometry of, 13.64–13.65, 13.64f
 noise, 13.70–13.71
 sensitivity, 13.65–13.66, 13.66f
 speed, 13.67–13.68
 untraveling-carrier (UTC) photodiodes, 13.68–13.69, 13.68f
- Pin holes, in neutron and x-ray optics, 26.7
- Pin junctions, in fiber optic systems, 13.59
- Pin photodiodes, 13.64–13.66
- Pin waveguides, 13.68
- Pinch plasma, 57.1–57.5, 57.2f, 57.3t, 57.4f
- Pinhole apertures, for SPECT imaging, 32.3
- Pitch, of liquid crystals, 8.10
- Planar buried heterostructure (PBH) lasers, 13.6
- Planck radiation law, 3.18, 3.20
- Plane wave analysis, for acousto-optic interaction, 6.6–6.9
- Plasma:
 and atomic spectroscopy, 2.4–2.5
 for extreme ultraviolet lithography, 34.5
 laser-generated, 56.1–56.10
 Bremsstrahlung, 56.8–56.10
 and characteristic radiation, 56.2–56.10
 recombination radiation, 56.10
 spectral line broadening, 56.2–56.8
 pinch, 57.1–57.5, 57.2f, 57.3t, 57.4f
- Plasma focus, for z-pinch radiation, 57.3, 57.3t
- Plasma-based EUV lasers, 58.2–58.4, 58.3f
- P-n junctions, 20.1
- Pockels cells, 7.33
- Pockels (linear electro-optic) effect:
 and electro-optic modulators, 7.6–7.11, 7.8t, 7.11
 and liquids, 7.34
 in OTDM networks, 20.1, 20.18, 20.19
- Pockels' theory, elasto-optic effect and, 6.5, 6.7
- Poincaré sphere, 19.18–19.19
- Point-by-point technique, for fiber Bragg gratings, 17.8
- Point-spread function (PSF), 4.1, 4.34–4.36, 4.34f, 44.13, 44.14, 44.15f, 64.2
- Point-to-point links, in WDM networks, 21.3f, 21.4
- Poisson distribution, 9.10
- Poisson-effect cross coupling, 24.3
- Poisson's equation, 5.36
- Polarization:
 circular, 43.5–43.8, 43.7f
 phase plates for, 43.5–43.6, 43.5f
 and synchrotron radiation, 55.6–55.9, 55.7f
 eigen-, 7.13–7.16, 7.14f, 7.16f
 of insertion devices, 55.15–55.16
 of laser diodes, 13.13
 linear, 43.2–43.4, 43.3f, 43.4f, 43.6, 43.8
 neutron, 63.27–63.29, 63.30f
 Stokes-Poincaré parameters for, 43.2
 transverse electric, 19.7
 transverse magnetic, 19.7
 and VCSELs, 13.48
 of x-rays, 43.1–43.2

- Polarization analyzers, 43.4, 43.4f, 43.6–43.8, 43.7f
- Polarization dependence, 13.58–13.59, 19.7, 19.7f, 19.32
- Polarization independence, 6.44, 13.53
- Polarization modulation (dynamic retardation), 7.20–7.22, 7.20f, 7.21f
- Polarization scrambling, 19.18
- Polarization spectroscopy, 2.21, 2.22, 2.23f
- Polarization-dependent gain (PDG), 14.9, 19.18–19.20, 21.18
- Polarization-dependent loss (PDL), 19.18, 21.18
- Polarization-maintaining (PM) fibers, 25.3, 25.12
- Polarization-mode dispersion (PMD), 21.16–21.18, 21.17f–21.18f
- Polarizers:
 - fiber-based couplers as, 16.5–16.6
 - linear, 43.2–43.3
 - multilayers, 41.9
 - for networking, 18.7, 18.7f, 18.10
- Polarizing crystal optics, 43.1–43.8
 - circular polarization analyzers, 43.6–43.8, 43.7f
 - linear polarization analyzers, 43.4, 43.4f
 - linear polarizers, 43.2–43.3, 43.3f
 - phase plates for circular polarization, 43.5–43.6, 43.5f
 - and polarization of x-rays, 43.1–43.2
- Polarizing mirrors, 63.28
- Polycapillary optics:
 - and brightness of x-ray tube sources, 54.16
 - collimating, 53.14, 53.14f
 - and neutron optics, 63.21–63.22
 - x-ray diffraction, 28.5
- Polycapillary x-ray optics, 53.1–53.19
 - alignment and measurement in, 53.5–53.8, 53.6f, 53.7f
 - applications of, 53.10–53.19, 53.11f–53.18f
 - collimation, 53.8–53.9, 53.8f–53.9f
 - focusing, 53.9–53.10, 53.10t
 - history of, 53.1–53.3, 53.2f
 - radiation resistance in, 53.5
 - simulations and defect analysis in, 53.3–53.5, 53.3f–53.5f
- Polycrystalline (PC) fibers, 12.2, 12.3t, 12.8–12.9, 12.8f
- Polymer stabilized cholesteric texture (PSCT), of liquid crystals, 8.37, 8.37f
- Polymer sustained alignment (PSA) technique, for LC cells, 8.28
- Polymer-dispersed liquid crystals (PDLCs), 8.36, 8.36f
- Polymer/liquid crystal composites, 8.36–8.37, 8.36f–8.37f
- Polymer-stabilized liquid crystals (PSLCs), 8.36, 8.36f, 8.37
- Polynomials:
 - Chebyshev, 46.6
 - Legendre, 46.6
 - Legendre-Fourier, 45.6
 - Zernike, 4.17–4.20, 4.20t, 5.10, 46.6
- Population inversions, amplification and, 19.2
- Pore optics, 49.1–49.7, 49.2f–49.6f
- Positive core-cladding index difference, for PFCs, 11.12–11.14, 11.12f, 11.13f
- Positive orders of radiation, 40.1
- Positive-intrinsic-negative (PIN) receivers, 9.8–9.10
- Positron emission tomography (PET), 32.1
- Postprocessing, of SOAs, 19.16, 19.17
- Powder diffraction, in polycapillary x-ray optics, 53.14, 53.14f
- Power:
 - ASE, 19.20
 - channel, for EDFAs, 21.41, 21.42f
 - and dispersion-managed solitons, 22.12
 - for extreme ultraviolet lithography, 34.5
 - incident, of scatterometers, 1.14–1.15
 - input saturation, 14.3
 - insertion device, 55.15
 - noise equivalent, 1.13, 13.71
 - of NPM AOTFs, 6.41
 - Nyquist frequency, 46.8–46.9, 46.11
 - saturation output, 14.3, 14.4
 - Stokes, 14.9
 - of synchrotron radiation, 55.8–55.9, 55.8f
- Power amplifiers, 14.4
- Power dependence, of electroabsorption modulators, 13.59
- Power loss, 21.13–21.14
- Power penalties, of fiber optic devices, 9.11, 15.8–15.17
- Power per unit bandwidth, 7.34–7.35
- Power spectral density (PSD) function, 41.6, 41.7, 46.7–46.9
- Power splitters, 16.1, 16.4, 18.2–18.3, 18.2f, 18.3f, 18.9, 18.9f
- Poynting vectors, 7.3, 7.5, 42.2

- Praseodymium-doped fiber amplifiers (PDFAs), 14.2, 14.2*t*, 14.7
- Pre-amp semiconductor optical amplifiers (SOAs), 19.23
- Prefocusing, with refractive x-ray lenses, 37.8
- Preform manufacture, of fiber lasers, 25.26–25.27
- Pressure broadening, of spectral lines, 56.5–56.6
- Prewhitening techniques, in x-ray mirror metrology, 46.9
- Prisms, 18.5, 18.5*f*, 63.22
- Processors, real-time, 5.34–5.35, 5.34*f*, 5.35*f*
- Profile analysis, in x-ray mirror metrology, 46.6–46.12, 46.7*f*, 46.10*f*
- Profile errors, in polycapillary x-ray optics, 53.3, 53.3*f*
- Profilometry:
 - height, 46.3
 - 1D, 2D, and 3D, 46.6
 - slope, 46.3–46.6, 46.5*f*
- Proportional counters, x-ray detectors and, 60.4–60.5, 60.9*t*, 60.10*t*
- Pulse amplification, chirped, 25.2, 25.32, 25.33
- Pulse code modulation, 20.8
- Pulsed-dye lasers, 5.32
- Pump cladding, 25.3
- Pump wavelength, EDFAs and, 14.5–14.6
- Pumping techniques, for fiber lasers, 25.9–25.13, 25.11*f*, 25.12*f*
- Pumps, for rare-earth-doped amplifiers, 14.2–14.3, 14.3*f*
- Pupils, annular, 4.10–4.16, 4.11*f*–4.15*f*, 4.15*t*
- Q factor, for OTDM networks, 20.11
- Q-switched oscillators, 25.30–25.31
- Quadratic (Kerr) electro-optic effect, 7.6, 7.9*t*–7.10*t*, 7.11
- Quadratic Stark effect, 56.6
- Quantum detection efficiency (QDE), 60.7, 61.2–61.3
- Quantum dots, 19.12, 19.21
- Quantum efficiency, of pin photodiodes, 13.65, 13.66
- Quantum limit, of digital on-off keying receivers, 9.9
- Quantum well (QW) lasers, 13.24–13.28, 13.25*f*, 13.26*f*
- Quantum wells (QWs), 20.1
 - coupled, 13.58
 - and EA modulators, 20.20
 - and electrorefractive modulators, 13.62
 - in fiber optic systems, 13.4
 - and SOAs, 19.7, 19.11, 19.12
 - and VCSELs, 13.43
- Quantum-confined Stark effect (QCSE), 13.2, 13.56, 13.56*t*, 20.1
- Quarter-wave phase plates, 43.6
- Quarter-wavelength-shifted gratings, 13.31–13.32, 13.31*f*
- Quaternary structure, of fiber optic devices, 13.2
- Radiance, thermal spectral, 3.18, 3.20, 3.20*f*
- Radiation:
 - Bremsstrahlung, 31.3
 - continuous, 54.4–54.6, 54.5*f*
 - from laser-generated plasmas, 56.1, 56.8–56.10
 - from pinch plasma sources, 57.1
 - and x-ray fluorescence, 29.3, 29.5, 29.6, 29.11
 - characteristic
 - Bremsstrahlung radiation as, 56.8–56.10
 - from laser-generated plasmas, 56.2–56.10
 - recombination radiation as, 56.10
 - spectral lines as, 56.2–56.8
 - from x-ray tube sources, 54.6–54.8, 54.7*f*
 - Compton sources of, 55.2–55.3, 55.3*t*
 - ionizing, 15.17
 - lab-based sources of, 50.2–50.7, 50.4*f*, 50.6*f*, 50.8*f*
 - from laser-generated plasmas, 56.2–56.10
 - negative orders of, 40.1
 - Planck radiation law, 3.18, 3.20
 - positive orders of, 40.1
 - recombination, 56.10
 - synchrotron sources of (*see* Synchrotron radiation sources)
 - Unruh, 58.2
 - X-pinch sources of, 57.3, 57.3*t*, 57.4
 - zero order of, 40.1
- Radiation induced loss, 15.17
- Radiation resistance, of polycapillary x-ray optics, 53.5
- Radiative quantum efficiency, of PDFAs, 14.7
- Radiography, 31.1–31.4, 31.2*f*–31.4*f*, 62.1
- Radiometry, infrared fibers for, 12.3*t*

- Rainbows, 3.41, 3.41*f*
- Raman amplifiers, 19.27, 21.42–21.44, 21.42*f*–21.44*f*
- Raman bands, of glass, 11.24
- Raman effect, inverse, 10.5, 10.6, 14.9
- Raman fiber amplifiers, 14.1, 14.2, 14.2*t*, 14.8–14.9, 14.8*f*, 14.10*f*
- Raman gain, 10.5, 10.6, 21.42*f*
- Raman resonance, 14.8
- Raman scattering:
 and atmospheric optics, 3.12, 3.21
 and configurational relaxation of solids, 2.15–2.17, 2.18*f*
 stimulated, 25.6
 and fiber optic communication links, 15.8
 and optical fiber amplifiers, 14.2, 14.8
 in optical fibers, 10.1, 10.4–10.7, 10.5*f*
 and photonic crystal fiber guidance, 11.23, 11.24, 11.26
 and WDM networks, 21.20
- Raman threshold, 10.7
- Raman-Nath diffraction regime, 6.4, 6.6
- Raman-Nath equations, 6.6–6.7
- Random Device Slope Scanner, 46.3
- Rare-earth ions, 2.7–2.8, 2.11
- Rare-earth-doped fiber lasers, 25.22–25.26, 25.23*t*
- Rare-earth-doped optical fiber amplifiers, 14.1, 14.2–14.4, 14.3*f*
- Ray tracing, for x-ray optics, 35.1–35.6, 35.4*f*, 35.5*f*
- Rayleigh backscattering, 10.8, 14.3
- Rayleigh beacons, 5.29–5.31, 5.31*f*
- Rayleigh criterion, 7.26, 34.1, 40.3, 42.2, 51.3
- Rayleigh range, inverse Compton scattering and, 59.1, 59.2
- Rayleigh resolution limit, 11.13
- Rayleigh scattering:
 in glass, 11.21
 and green flashes, 3.43
 for HMFG fibers, 12.4
 and laser beacons, 5.27, 5.30, 5.32
 and MXRE, 29.5, 29.8*f*
 and optical fibers, 9.4, 9.12
 and Raman fiber amplifiers, 14.9
 in theory of interaction of light and atmosphere, 3.12, 3.15–3.16
 and x-ray optics, 26.7
- Rayleigh scattering extinction coefficient, 3.16
- Reabsorption, of spectral lines, 56.8
- Real-time processors, for adaptive optics, 5.34–5.35, 5.34*f*, 5.35*f*
- Receivers:
 avalanche photodiode, 9.8, 9.10–9.11
 digital on-off-keying, 9.9–9.11
 ideal, 9.9
 in OTDM networks, 20.7–20.8
 positive-intrinsic-negative, 9.8–9.10
 in scatterometers, 1.9–1.10, 1.9*f*
 sensitivity of fiber optic, 9.8–9.11
- Receptors, image, 31.4, 31.4*f*
- Reciprocal linear dispersion, 38.7
- Recombination radiation, 56.10
- Reconfigurability, of WDM networks, 21.12–21.13, 21.12*f*, 21.13*f*
- Reconfigurable optical add/drop multiplexers (ROADMs), 21.12
- Red-shifts, in peak gain wavelength, 25.24
- Reference method (scatterometer calibration), 1.15
- Reflectance, in interaction of light and atmosphere, 3.21, 3.21*f*
- Reflection(s):
 Bragg
 of crystals, 40.9
 in interferometers, 63.26, 63.27
 and linear polarization, 43.2–43.4, 43.3*f*, 43.4*f*, 43.6, 43.8
 and liquid crystals, 8.10
 in monochromators, 63.24
 and multilayers, 41.2
 simultaneous, 43.7 [See also Multiple-beam Bragg diffraction (MBD)]
 and x-ray absorption spectroscopy, 30.2, 30.4
 Fresnel, 13.6, 13.53, 17.3, 25.8
 grazing-angle, 63.21
 in neutron and x-ray optics, 26.8–26.9, 63.20–63.21, 63.21*f*
 reference, for Fabry-Perot sensors, 24.2
 retro-, 13.6–13.7
 sensing, for Fabry-Perot sensors, 24.2
 total external, 64.1–64.2
 total internal, 11.2, 11.3, 11.14, 11.15
- Reflective LCDs, 8.31–8.32, 8.31*f*
- Reflective semiconductor optical amplifiers (SOAs), 19.28
- Reflectivity, of Wolter x-ray optics, 47.5

- Reflectors:
 Bragg, 13.45
 modulated grating, 13.36
 for neutron beams, 64.3, 64.4, 64.4f
- Refraction:
 electro-, 13.2
 in neutron optics, 63.19–63.23, 63.21f, 63.23f
 in x-ray optics, 26.6–26.8, 26.8f
- Refractive index (*see* Index of refraction)
- Refractive x-ray lenses, 37.3–37.11
 applications of, 37.11
 history of, 37.3
 nanofocusing, 37.8–37.11, 37.9f, 37.10f
 parabolic, 37.4–37.8, 37.4f, 37.6f, 37.7f
- Registered detected point spread function (RDPSF), 44.14, 44.15f
- Relative intensity noise (RIN):
 in fiber optic communication links, 15.2, 15.14
 of laser diodes, 13.18–13.19, 13.19f, 13.23
 and optical fibers, 9.11, 9.16
- Relaxation, configurational, 2.14–2.17, 2.15f–2.18f
- Relaxation oscillations, of laser diodes, 13.14–13.16
- Remote sensing, in atmospheric optics, 3.36–3.40, 3.37f–3.40f
- Repeater spacing, for optical fibers, 9.12–9.13
- Resolution:
 angular, 47.2–47.3, 47.2f, 47.10–47.11, 47.11f
 of deflectors, 6.25, 6.29, 6.30
 field-weighted-average, 44.10
 of gratings and monochromators, 38.7
 of NPM AOTFs, 6.40
 spatial, 26.10–26.11
 of telescopes, 4.2–4.3
 in x-ray imaging, 62.3–62.5
- Resolution enhancement techniques (RET),
 for extreme ultraviolet lithography, 34.1
- Resonance, of PCFs, 11.12
- Resonant cavity light-emitting diodes (RC-LEDs), 13.39, 13.40
- Resonant circuits, electro-optic modulators
 and, 7.34
- Resonant photodiodes, 13.65
- Resonators, 25.13, 25.16
- Responsivity, of pin diodes, 13.65–13.66, 13.66f
- Restimulable phosphor detectors, 60.8, 60.10t
- Retarders, phase, 41.9, 43.6
- Retroreflection, of guided light, 13.6–13.7
- Return-to-zero differential quadrature
 phase-shift-keying (RZ-DQPSK) format, 21.35, 21.35f, 21.36t, 21.37t
- Return-to-zero differential-phase-shift-keying (RZ-DPSK) format, 21.28, 21.34f, 21.36t, 21.37t
- Return-to-zero (RZ) format, 20.8, 20.9f, 20.10f, 21.16, 21.29f, 21.31f
- Return-to-zero on-off keying (RZ-OOK),
 21.34, 21.36t
- Rician density function, 3.36
- Ridge waveguide (RWG) laser, 13.6
- Rigid-body motions, 45.6–45.7
- Ring topologies, for WDM networks,
 21.5–21.6, 21.5f–21.7f
- Robustness, of solitons, 22.4–22.5
- ROSAT observatory, 47.5
- ROSAT (Rontgensatellit) telescope, 44.4, 44.6f, 44.10
- Rose model, for x-ray attenuation, 31.2
- Rotating anodes, as x-ray tube sources, 54.12
- Rotationally parabolic profiles, for refractive
 x-ray lenses, 37.4–37.8, 37.4f, 37.6f, 37.7f
- Rotators, Faraday, 18.7, 18.7f, 18.10
- Roto-optic effect, 6.6
- Roughness, polycapillary x-ray optics, 53.4
- Rowland circle, 39.5, 39.6
- Rowland spherical grating, 38.6
- Russell Saunders coupling, 2.11
- Rydberg energy, 54.7
- Rytov transformation, 5.9
- SAGE II satellite system, 3.39, 3.40f
- Sagittal-focusing geometry, for monochromators,
 39.6
- Sagnac interferometers, 20.22
- Sagnac loops, 17.8
- Sample mounts, for scatterometers, 1.9
- Sampled gratings (SG), 13.34, 13.34f
- Sampling, in OTDM networks, 20.1, 20.4–20.6, 20.5f, 20.6f
- Sapphire, 12.3t, 12.9–12.10, 12.10f
- Saturated output power, of rare-earth-doped
 amplifiers, 14.4
- Saturation, of SOAs, 19.9–19.10, 19.10f
- Saturation current, of semiconductor diodes,
 13.69
- Saturation output power, of rare-earth-doped
 amplifiers, 14.3

- Saturation regime, for rare-earth-doped amplifiers, 14.3
- Scaling law, or photonic crystal fibers, 11.7
- Scan rate, of high-resolution deflectors, 6.29
- Scanners, 7.26–7.28, 7.26f–7.28f
- Scanning electron microscopy (SEM):
and magnetron-sputtered MLLs, 42.6–42.7, 42.7f, 42.8f, 42.13, 42.13f
and x-ray fluorescence, 29.2, 29.3
and x-ray spectral detection, 62.1–62.3, 62.2f
- Scatter function (cosine-corrected BRDF), 1.6
- Scatter rejection, polycapillary x-ray optics and, 53.14–53.16, 53.15f–53.16f
- Scattering:
anti-Stokes, 10.5
Brillouin
backward, 11.25
forward, 11.25, 11.26
photonic crystal fibers, 11.25–11.26, 11.25f
stimulated, 10.1, 10.7–10.9, 15.8, 21.20, 25.6
Compton, 59.1
and circular polarization, 43.6
inverse, 59.1
and MXRF, 29.5, 29.8f
and polycapillary x-ray optics, 53.3, 53.15, 53.18
and refractive x-ray lenses, 37.5, 37.7
and x-ray attenuation, 31.2
and x-ray optics, 26.7, 36.1
elastic, 63.5
in heavy water, 63.10
incoherent, 26.7, 31.2, 63.7, 63.8
inelastic, 63.3
and photonic crystal fibers, 11.24–11.26, 11.25f
Raman, 11.24
and atmospheric optics, 3.12, 3.21
and configurational relaxation of solids, 2.15–2.17, 2.18f
Rayleigh, 11.21
backscattering, 10.8, 14.3
in glass, 11.21
and green flashes, 3.43
for HMFG fibers, 12.4
and laser beacons, 5.27, 5.30, 5.32
and MXRF, 29.5, 29.8f
and optical fibers, 9.4, 9.12
and Raman fiber amplifiers, 14.9
Scattering, Rayleigh (*Cont.*):
in theory of interaction of light and atmosphere, 3.12, 3.15–3.16
and x-ray optics, 26.7
by silica-air photonic crystal fibers, 11.25
small angle neutron, 64.1
stimulated Raman, 25.6
and fiber optic communication links, 15.8
and optical fiber amplifiers, 14.2, 14.8
in optical fibers, 10.1, 10.4–10.7, 10.5f
and photonic crystal fiber guidance, 11.23, 11.24, 11.26
and WDM networks, 21.20
Thompson, 26.7
Thomson backscattering, 59.1
in water, 63.10, 63.10t, 63.11f
Scattering cross sections, for neutrons, 63.6–63.9, 63.6t, 63.10t
Scattering length densities, of neutrons, 63.9–63.11, 63.11f
Scattering lengths, of neutrons, 63.5–63.9, 63.6t, 63.10t
Scatterometers, 1.3–1.16
BSDF, 1.8, 1.8f
calibration of, 1.14–1.15
configurations and components for, 1.7–1.11, 1.8f–1.10f
error analysis of, 1.15
incident power measurement of, 1.14–1.15
instrument signature and quality, 1.11–1.13, 1.11f, 1.13t
measurement issues with, 1.13–1.14
in metrology, 1.16
specifications of, 1.5–1.7, 1.5f
Schawlow-Townes linewidth, 13.21
Schottky barriers, 13.58, 13.69
Schottky photodiodes, 13.63, 13.73
Schroder-Van Laar equation, 8.14
Schrödinger equation, 63.4–63.5
Schwarzschild objectives, 26.10, 51.1–51.3, 51.2f–51.4f
Schwarzschild optics, 44.6, 48.2
Schwinger formula, 55.15
Scintigraphy, 32.1, 53.17, 53.18, 53.18f
Scintillation, 3.26, 3.34–3.36, 3.35f
Scintillation x-ray detectors, 60.7–60.8, 60.9t, 60.10t
Scintillator-based flat panel detectors, 61.4, 61.4f, 61.7–61.8, 61.8f
Scintillators, neutron, 63.33

- Secondary electrons, 54.12
- Segmented deformable mirrors, 5.37, 5.37*f*
- Selective area epitaxy, 13.60
- Self-amplified spontaneous emission (SASE)
 - mode, of lasers, 58.1
- Self-healing ring networks, for WDM, 21.6, 21.6*f*
- Self-phase modulation (SPM):
 - in optical fibers, 10.3–10.4
 - and solitons, 22.3–22.4, 22.3*f*, 22.4*f*, 22.13
 - in WDM networks, 21.18–21.19, 21.19*f*
- Self-saturation effect, of EDFAs, 14.6
- Self-steepening, 10.4
- Semiconductor (x-ray) detectors, 29.3, 60.5–60.6, 60.9*t*, 60.10*t*
- Semiconductor Equipment and Materials International (SEMI), 1.4
- Semiconductor interferometric modulators, 13.2, 13.63
- Semiconductor laser amplifier loop optical mirrors (SLALOM), 20.22
- Semiconductor laser amplifiers (SLAs), 9.13–9.14
- Semiconductor lasers, 13.1
- Semiconductor optical amplifiers (SOAs), 19.1–19.36
 - amplification in, 19.1*f*–19.2*f*, 19.2, 19.22–19.27, 19.22*f*–19.26*f*
 - ASE noise, 19.3, 19.4*f*
 - confinement factor, 19.6
 - device characterization, 19.17–19.21, 19.17*f*, 19.19*f*–19.21*f*
 - fabrication, 19.15–19.17, 19.15*f*–19.17*f*
 - gain, 19.4–19.6, 19.4*f*–19.6*f*
 - gain clamping, 19.14–19.15, 19.14*f*
 - gain dynamics, 19.12–19.13, 19.13*f*, 19.14*f*
 - gain ripple and feedback reduction, 19.8, 19.8*f*
 - history of, 19.1
 - material systems, 19.11–19.12, 19.11*f*–19.12*f*
 - noise figure, 19.9
 - nonlinear applications, 19.29–19.36, 19.29*f*, 19.31*f*, 19.33*f*, 19.34*f*
 - and photonic integrated circuits, 19.36
 - polarization dependence, 19.7, 19.7*f*
 - saturation, 19.9–19.10, 19.10*f*
 - switching and modulation, 19.28, 19.28*f*
- Semiconductor saturable absorber mirror (SESAM), 25.3, 25.32
- Semiconductors:
 - and electroabsorption modulators, 13.58
 - in electro-optic modulators, 7.34
 - and extreme ultraviolet lithography, 34.1–34.2, 34.2*t*
 - fiber amplifiers vs., 9.13–9.14
 - signal processing in complementary metal oxide, 62.5
- Sensing reflection, for Fabry-Perot sensors, 24.2
- Sensitivity (responsivity), of pin diodes, 13.65–13.66, 13.66*f*
- Sensors:
 - electro-optic modulators, 7.38
 - extrinsic Fabry-Perot interferometric (EFPI), 24.2–24.4, 24.2*f*, 24.3*f*
 - fiber optic chemical, 12.3*t*
 - intrinsic Fabry-Perot interferometric (IFPI), 24.4–24.5, 24.4*f*
 - optical fiber, 24.1–24.13
 - comparison of, 24.13
 - extrinsic Fabry-Perot interferometric, 24.2–24.4, 24.2*f*, 24.3*f*
 - fiber Bragg grating, 24.5–24.8, 24.6*f*–24.7*f*
 - intrinsic Fabry-Perot interferometric, 24.4–24.5, 24.4*f*
 - long-period grating sensors, 24.8–24.13, 24.9*f*–24.12*f*, 24.11*t*, 24.13*t*
 - Shack-Hartmann, 5.21, 5.36, 5.40–5.43
 - Shack-Hartmann wavefront, 50.5, 50.6*f*
 - transition-edge, 60.9
- Separate confinement heterostructure (SCH), in fiber optic modulators, 13.4*f*, 13.5
- Serial byte connection (SBCON) standard, 23.1
- Serial multiplexing, 20.12
- Servo lag, subaperture size and, 5.41–5.42
- Shack-Hartmann sensors, 5.21, 5.36, 5.40–5.43
- Shack-Hartmann technique, for adaptive optics, 5.23–5.27, 5.23*f*, 5.25*f*, 5.26*f*
- Shack-Hartmann wavefront sensors, 50.5, 50.6*f*
- SHADOW code, 35.2–35.5, 35.4*f*, 35.5*f*
- SHADOWVUI interface, 35.3, 35.5, 35.5*f*
- Shane telescope, 5.27, 5.32
- Shielding, in neutron optics, 63.13–63.14
- Short-exposure images, 4.3, 4.31–4.35, 4.31*f*–4.34*f*, 4.35*t*
- Short-period gratings (SPGs), 24.6
- Shot noise, 13.70, 13.73
- Sidelobe suppression, in AOTFs, 6.41–40.2, 6.44–6.45

- Signal processing, by wideband AO Bragg cells, 6.30
- Signal-dependent noises, 9.11
- Signal-to-noise ratio (SNR):
 for electro-optic modulators, 7.16
 for fiber optic communication links, 15.2, 15.4, 15.5
 for laser diodes, 13.19
 in neutron optics, 63.13
 optical
 for SOAs, 19.24–19.27
 for WDM networks, 21.20, 21.28, 21.34
 for optical fibers, 9.9
 for solitons, 22.1, 22.6–22.8
- Silica fibers, in lasers, 25.27*t*, 25.28
- Silica-air photonic crystal fibers, 11.14, 11.14*f*, 11.25
- Silicon pore optics, 49.6–49.7, 49.7*f*
- Silver halide fibers, 12.3*t*, 12.8–12.9, 12.8*f*
- Simbol-X spacecraft, 47.10
- Simultaneous Bragg reflection (multiple-beam Bragg diffraction), 43.6–43.8, 43.7*f*
- Single attach FDDI nodes, 23.3
- Single capillaries (*see* Monocapillary x-ray optics)
- Single crystal diffraction, 53.12–53.14, 53.12*f*–53.13*f*
- Single crystal (SC) fibers, as infrared optical fibers, 12.3*t*, 12.9–12.10, 12.10*f*
- Single mirror resonators, 25.13
- Single photon emission computed tomography (SPECT), 32.1–32.4
- Single-bounce monocapillary x-ray optics, 52.5–52.6, 52.5*f*
- Single-cell-gap transreflective LCDs, 8.33–8.35, 8.34*f*
- Single-channel systems, of SOAs, 19.22–19.24, 19.22*f*–19.24*f*
- Single-longitudinal-mode (SLM) lasers, 9.8
- Single-mode excitation technique (SMET), 25.3, 25.20
- Single-mode fibers (SMFs):
 dispersion in, 21.16, 21.16*f*, 21.21, 21.21*f*
 for E-LEDs, 13.40
 and photonic crystal fibers, 11.2, 11.11, 11.13, 11.23–11.25
- Singular-value-decomposition (SVD) algorithms, 5.25, 5.26
- Slab-coupled optical waveguide amplifiers (SCOWAs), 19.21
- Slater integrals, 2.8
- Slope errors, 38.7, 45.7–45.8
- Slope measurement analysis, in x-ray mirror metrology, 46.11–46.12
- Slope profilometry, 46.3–46.6
- Slot-detection x-ray imaging, 61.2, 61.2*f*
- Slowly varying envelope approximation (SVEA), 10.2, 10.6
- Small angle neutron scattering (SANS), 64.1
- Smectic phase, of liquid crystals, 8.8–8.12, 8.11*f*, 8.12*f*
- Snell's law, 19.8, 26.8
- Soft x-ray region, gratings and monochromators in, 38.1–38.8
 diffraction properties, 38.1–38.3, 38.2*f*
 dispersion properties, 38.6–38.7
 efficiency, 38.8
 focusing properties, 38.3–38.6, 38.3*f*, 38.4*t*–38.5*t*
 resolution properties, 38.7
- Soft x-ray telescopes, 50.2
- Soft x-rays, circularly polarized, 55.6–55.7
- Software implementation, for adaptive optics, 5.21–5.38
 higher-order wavefront sensing techniques, 5.36–5.37
 laser beacons, 5.27–5.34, 5.28*f*–5.31*f*, 5.33*f*
 real-time processors, 5.34–5.35, 5.34*f*, 5.35*f*
 Shack-Hartmann technique, 5.23–5.27, 5.23*f*, 5.25*f*, 5.26*f*
 tracking, 5.21–5.23, 5.22*f*
 wavefront correctors, 5.37–5.38, 5.38*f*
- SOHO observatory, 41.3
- Solar x-ray Imager (SXI), 44.9, 44.12–44.13, 44.16–44.17, 44.16*f*, 44.17*f*
- Solid core photonic crystal fibers:
 attenuation in, 11.20, 11.21
 Brillouin scattering in, 11.25
 group velocity dispersion, 11.18, 11.18*f*, 11.19*f*
 Kerr effects for, 11.23
- Solid state detectors, in neutron optics, 63.33–63.34
- Solids:
 configurational relaxation in, 2.14–2.17, 2.15*f*–2.18*f*
 optical absorption measurements of, 2.7–2.13, 2.8*f*, 2.10*f*, 2.12*f*
 zero-phonon lines in, 2.13–2.14, 2.14*f*
- Solid-state ion chambers, 60.5–60.6

- Solitons:
 classical, 22.2–22.4, 22.2f–22.4f
 in communication systems, 22.1–22.17
 dispersion-managed, 22.12–22.15, 22.13f, 22.14f
 effects of, 22.1–22.2
 errors, 22.6
 and frequency-guiding filters, 22.7–22.9
 in optical amplifiers, 9.14
 and optical fibers, 10.1
 properties of, 22.4–22.5
 self-frequency shift cancellation, 11.24
 transmission systems for, 22.5–22.7
 and wavelength division multiplexing, 22.2, 22.9–22.12, 22.15–22.17
- Soller collimators, 28.2, 28.2f
- Soller slits, 26.7
- SOR telescope, 5.27, 5.32, 5.34, 5.38
- Source debris, in extreme ultraviolet lithography, 34.5
- Source depth measurements, for x-ray tube sources, 54.14, 54.14f
- Source spot measurements, for x-ray tube sources, 54.14, 54.14f, 54.15, 54.15f
- Source-model technique, for photonic crystal fibers, 11.7
- Space-charge region, of pin photodiodes, 13.64
- Spatial coherence, 4.10
- Spatial coherence length, 27.2
- Spatial frequency spectrum, 6.16
- Spatial hole burning, 13.13, 20.2
- Spatial light modulation (SLM), 6.4, 6.9
- Spatial resolution, 26.10–26.11
- Spatial scale, adaptive optics and, 5.10–5.13, 5.11t, 5.12f, 5.13f
- Spectra (code), 55.14
- Spectral and temporal modalities, of fiber lasers, 25.29–25.33
- Spectral hole burning, 13.13
- Spectral inversion, four-wave mixing and, 10.11
- Spectral lines:
 broadening of
 Doppler, 3.14, 5.32, 56.5
 homogeneous, 20.1
 Lorentzian, 3.14
 pressure, 56.5–56.6
 spectral, 56.6–56.8
 from laser-generated plasmas, 56.2–56.8
- Spectral radiance, thermal, 3.18, 3.20, 3.20f
- Spectrometers, 1.14, 63.21f
- Spectroscopy:
 and absorption/photoluminescence of Cr^{3+} , 2.19–2.21, 2.19f, 2.20f
 in adaptive optics, 5.19–5.21, 5.19f, 5.20f
 atomic, 2.4–2.5
 evanescent wave, 12.13
 excitation, 2.15f, 2.21
 Fourier transform, 2.5, 2.6f
 and homogeneous lineshapes of spectra, 2.13–2.17, 2.14f–2.18f
 Lamb dip, 2.5, 2.7f
 Laser-induced-breakdown, 3.39
 measurements from, 2.1–2.24
 molecular, 2.5–2.6, 2.6f, 2.7f
 optical, 2.9–2.10, 2.10f
 optical absorption measurements
 of atomic energy levels, 2.2–2.5, 2.4f
 of solids, 2.7–2.13, 2.8f, 2.10f, 2.12f
 polarization, 2.21–2.22, 2.23f
 time-domain, 7.37, 7.37f–7.38f
 time-of-flight, 2.5
 x-ray absorption, 30.1–30.5, 30.2f–30.5f
 x-ray absorption near edge, 30.2, 30.2f, 30.4f
 Zeeman, 2.23–2.24, 2.24f
- Spectrum(-a):
 atomic, 2.13
 homogeneous lineshapes of, 2.13–2.17, 2.14f–2.18f
 Kolmogorov, 3.28, 3.31
 spatial frequency, 6.16
 von Kármán, 3.28, 5.6
 of x-ray tube sources, 54.4–54.10, 54.5f, 54.8f
 (See also Spectral lines)
- Speed:
 of avalanche photodiodes, 13.72
 of pin diodes, 13.67–13.68
- Spherical aberrations, 45.4
- Spherical-grating monochromators (SGMs), 38.3
- Spin flippers, 63.29
- Splay, of liquid crystals, 8.22, 8.23f
- Splice losses, for fiber optic communication links, 15.7, 15.8t
- Splicing, of PCFs, 11.26
- Split-off band, of strained layer quantum well lasers, 13.27
- Spontaneous emission, 2.13, 19.3, 20.1
- Spring-8 Compact SASE Source (SCSS), 58.1
- Spurious-free dynamic range (SFDR), 15.6
- SRW (code), 55.14

- Stack and draw method, of fiber laser fabrication, 25.27
- Stacked actuator continuous facesheet deformable mirrors, 5.37f, 5.38, 5.38f
- Standard atmosphere, composition of, 3.6–3.11, 3.7t, 3.8f–3.11f
- Standard dispersion-shifted fibers (DSFs), 24.10–24.11, 24.11t, 24.13t
- Star topologies, for WDM networks, 21.5–21.6, 21.5f, 21.7f
- Starfire Optical Range, 5.23
- Stark effect, 2.21, 20.20, 25.24, 56.6
- Stark levels, of EDFAs, 14.4, 14.5, 14.5f
- Static gain dynamic, for EDFAs, 21.41, 21.41f
- Static Jahn-Teller effect, 2.9
- Stationary anodes, of x-ray tube sources, 54.12
- Step-and-scan stage, of extreme ultraviolet lithography, 34.2
- Stern-Gerlach experiment, 63.4
- Stimulated Brillouin scattering (SBS), 10.1, 10.7–10.9, 15.8, 21.20, 25.6
- Stimulated emission, 20.2
- Stimulated Raman scattering (SRS), 25.6
 - and fiber optic communication links, 15.8
 - and optical fiber amplifiers, 14.2, 14.8
 - in optical fibers, 10.1, 10.4–10.7, 10.5f
 - and photonic crystal fiber guidance, 11.23, 11.24, 11.26
 - and WDM networks, 21.20
- Stokes frequency, 10.8, 21.42
- Stokes intensity, 10.6, 10.7
- Stokes power, 14.9
- Stokes shifts, 2.15–2.16
- Stokes waves, 10.4–10.9, 14.2
- Stokes-Poincaré parameters, for polarization, 43.2, 43.8
- Strain, in fiber optic devices, 13.2
- Strained layer quantum well lasers, 13.26–13.28, 13.26f, 13.27f
- Strained-layer superlattices (SLs), 2.11–2.13, 2.12f
- Strehl ratio:
 - and adaptive optics, 5.14, 5.14f–5.15f, 5.17, 5.17f, 5.20–5.22, 5.22f, 5.29, 5.29f, 5.35, 5.35f, 5.39, 5.40, 5.43, 5.43f–5.46f, 5.46
 - and imaging through atmospheric turbulence, 4.13, 4.14f, 4.15t, 4.27–4.28, 4.28f, 4.32–4.33, 4.34f, 4.36
- Strip loading, of RWG lasers, 13.6
- Subaperture size, for adaptive optics, 5.40–5.43, 5.42f
- Sum-frequency lasers, 5.32
- Superconducting quantum (SQUID) detector, 60.9
- Superconducting tunneling junctions (STJ) detectors, 60.9, 60.9t
- Supercontinuum (SC) generation, 11.23, 11.24f
- Supermirrors, 41.7, 41.8, 64.7
- Superstructure gratings (SSGs), 13.34, 13.35f
- Surface axial coordinates, in grazing incidence optics, 45.7
- Surface figure metrology, 46.3–46.6, 46.5f
- Surface finish metrology, 46.2
- Surface-emitting light-emitting diodes (S-LEDs), 13.36–13.40, 13.38f
- Suzaku observatory, 33.3–33.4, 33.3t
- SUZUKU spacecraft, 47.6, 47.7f
- Swift Burst Alert Telescope (BAT), 33.1
- Switches and switching:
 - for fiber-based couplers, 16.4
 - for networking, 18.4, 18.5, 18.5f, 18.11, 18.11f, 18.12f
 - and SOAs, 19.28, 19.28f
- Synchronous digital hierarchy (SDH), 9.17
- Synchronous optical networks (SONETs), 9.17, 21.6, 23.6
- Synchrotron beamlines, SHADOW code and, 35.1–35.2, 35.4
- Synchrotron radiation (SR) monochromators, 39.6
- Synchrotron radiation sources, 55.1–55.20
 - adaptive x-ray optics for, 50.2–50.7, 50.4f, 50.6f, 50.8f
 - coherence of, 55.17–55.20, 55.18f–55.19f
 - Compton sources vs., 55.2–55.3, 55.3t
 - history of, 55.1–55.2
 - insertion devices, 55.9–55.16, 55.10f, 55.12f, 55.13f
 - as linear polarizers, 43.2, 43.3
 - and refractive x-ray lenses, 37.5, 37.11
 - theory of synchrotron radiation emission, 55.2–55.9, 55.2f, 55.3f, 55.5f–55.8f
- Takagi-Taupin calculations, for MLLs, 42.12–42.14
- Talystep stylus profiler, 46.2
- Tanabe-Sugano diagrams, 2.9–2.11, 2.10f, 2.19
- Tangential phase matching (TPM), 6.12, 6.13, 6.17, 6.25–6.27, 6.26f

- Tapered fiber bundles (TFBs), 25.3, 25.10–25.11, 25.11*f*, 25.16
- Tapered fiber coupler process, 16.1–16.3, 16.2*f*
- Tapered fiber method, for adaptor fabrication, 25.17, 25.17*f*
- Teflon coatings, for infrared optical fibers, 12.4, 12.7, 12.9
- Telecommunication systems, data communication vs., 15.1–15.2
- Telescopes:
- and adaptive optics, 5.2–5.5, 5.3*f*, 5.4*f*, 5.43–5.46, 5.43*f*–5.46*f*
 - Burst Alert, 33.1
 - Calar Alto, 5.27
 - Canada-France-Hawaii, 5.23
 - Cassegrain, 44.4
 - ESO, 5.35
 - Gemini North, 5.20, 5.21*f*
 - with grazing incidence optics, 44.6–44.12, 44.7*f*, 44.9*f*–44.11*f*, 45.1
 - hard vs. soft x-ray, 50.2
 - Hobby-Eberly, 5.2
 - hyperboloid-hyperboloid (HH) grazing incidence x-ray, 44.10–44.12, 44.11*f*
 - Keck, 4.36, 5.2, 5.27
 - Large Binocular, 5.5
 - Mt. Wilson, 5.27
 - Multiple Mirror, 5.5
 - resolution of, 4.2–4.3
 - ROSAT, 44.4, 44.6*f*
 - Shane, 5.27
 - SOR, 5.27
 - Swift Burst Alert, 33.1
 - 10-m, 5.45–5.46, 5.45*f*–5.46*f*
 - 3.5-m, 5.43–5.45, 5.43*f*–5.45*f*
 - wide-field lobster-eye, 48.4
 - Wolter, 44.4, 44.5*f*, 44.6–44.10, 44.7*f*, 44.10*f*, 45.1–45.5, 45.2*f*
 - Wolter-Schwarzschild, 44.7, 44.11, 45.1
- Tellurites, for fiber lasers, 25.27*t*, 25.28
- Tellurium dioxide (TeO_2), 6.17, 6.21*t*, 6.25, 6.29*t*, 6.34*t*, 6.39, 6.42
- Temperature:
- and laser diodes, 13.11
 - and liquid crystals, 8.17, 8.21–8.22, 8.22*f*
 - and long-period grating sensors, 24.13, 24.13*t*
- Temporal coherence, of synchrotron radiation sources, 55.17–55.18, 55.18*f*
- 10-m telescope systems, 5.45–5.46, 5.45*f*–5.46*f*
- Terabit (unit), 20.2
- Terahertz asymmetric optical demultiplexers (TOAD), 20.22
- Ternary layers, in fiber optic devices, 13.2
- Thermal control and correction, for adaptive optics, 50.5, 50.6
- Thermal imaging, 12.3*t*
- Thermal neutrons, 63.3
- Thermal (Johnson) noise, 13.70
- Thermal runaway, of lasers, 13.11
- Thermal spectral radiance, 3.18, 3.20, 3.20*f*
- Thermally expanded cores (TECs), 25.3, 25.17, 25.17*f*
- Thermoelectric coolers (TECs), 19.4*f*, 19.5
- Thin film transistors (TFTs), 61.3, 61.4, 61.6
- Thin lenses, zone plates as, 40.3–40.4
- Thompson scattering, 26.7
- Thomson backscattering, 59.1
- Thomson scattering cross-section, 59.2
- 3D profilometry, 46.6
- 3.5-m telescopes, 5.43–5.45, 5.43*f*–5.45*f*
- Thulium-doped fibers, 25.23*t*, 25.25, 25.32
- Tight binding approximations, for photonic crystal fibers, 11.8
- Tilt:
- atmospheric, 5.14, 5.14*f*–5.15*f*
 - gradient (G-tilt), 4.3
 - and adaptive optics, 5.14–5.16
 - and angle of arrival, 4.23, 4.25, 4.26
 - wavefront (Z-tilt), 4.3, 4.23–4.25, 5.14, 5.15
 - Zernike, 4.23–4.26
- Tilt errors, for grazing incidence optics, 45.7
- Tilt-corrected phase variance, 4.31–4.32, 4.31*f*
- Tilted diffraction geometry, 42.4
- Time shifts, of solitons, 22.15, 22.16
- Time-division multiplexing (TDM), 9.12, 20.3, 21.3
- Time-domain spectroscopy (TDS) systems, 7.37, 7.37*f*–7.38*f*
- Time-of-flight (TOF) spectroscopy, 2.5
- Timing recovery, in OTDM networks, 20.10–20.12, 20.10*f*, 20.11*f*
- Tokens, of FDDI connectors, 23.3
- Tomography, 31.1, 31.5–31.7, 31.5*f*–31.7*f*
- Tomosynthesis, digital, 31.7–31.8
- Toroidal reflectors, for neutron beams, 64.4
- Toroidal-grating monochromators (TGMs), 38.3
- Total external reflection, 64.1–64.2

- Total integrated scatter (TIS):
 of neutrons, **64.2–64.3**
 and scatterometers, **1.4, 1.6–1.7, 1.10–1.11**
- Total internal reflection (TIR), of photonic crystal fibers, **11.2, 11.3, 11.14, 11.15**
- TRACE observatory, **41.3**
- Tracking, for adaptive optics, **5.15–5.18, 5.17f, 5.21–5.23, 5.22f**
- Transient response:
 of light-emitting diodes (LEDs), **13.41–13.42**
- Transient response, of laser diodes, **13.13–13.18, 13.14f, 13.16f, 13.17f**
- Transition-edge sensor (TES) microcalorimeter detectors, **29.9–29.11, 29.9f, 29.11f**
- Transition-edge sensors (TES), **60.9**
- Transition-metal ions, spectra of, **2.8–2.9, 2.8f**
- Transmission:
 of amplitude modulators, **7.22**
 analog, **9.15–9.17, 9.16f**
 atmospheric optical, **3.22–3.26, 3.22f–3.27f**
 Bormann, **43.3, 43.3f**
 broadband, **3.23–3.24, 3.25f–3.26f**
 electrical time domain multiplexed, **20.3, 20.25**
 formats for data, **20.9–20.10, 20.9f**
 Gaussian, **37.5**
 Laue, **26.8, 26.9f, 63.24, 63.26**
 line-by-line, **3.23, 3.24f**
 of NPM AOTFs, **6.41**
 with optical fibers, **9.7–9.8, 9.15–9.17**
 in OTDM networks, **20.7–20.8, 20.12–20.17, 20.14f–20.16f**
 polycapillary x-ray optics, **53.5–53.8, 53.6f, 53.7f**
 in polycapillary x-ray optics, **53.5–53.8, 53.6f, 53.7f**
 for solitons, **22.5–22.7**
 WDM dispersion managed soliton, **22.15–22.17**
 of x-ray tube sources, **54.11–54.12**
- Transmission loss, in fiber optic links, **15.7**
- Transmissive thin-film transistor (TFT) LCDs, **8.29–8.31, 8.30f, 8.31f, 8.32–8.35, 8.34f**
- Transparent fiber systems, **14.1**
- Transreflective LCDs (TR-LCDs), **8.32–8.35, 8.33f, 8.34f**
- Transverse coherence, **55.18–55.20, 55.19f**
- Transverse effective wavelength, for PCFs, **11.10**
- Transverse electric (TE) polarization, **19.7**
- Transverse electric (TE) waveguide mode, of laser diodes, **13.13**
- Transverse electro-optic modulators, **7.16, 7.17, 7.17f**
- Transverse holographic technique, for fiber Bragg gratings, **17.5, 17.5f**
- Transverse magnetic (TM) mode, of laser diodes, **13.13**
- Transverse magnetic (TM) polarization, **19.7**
- Transverse ray aberration (TRA) equations, **45.3–45.8**
- Transverse spatial coherence, **55.16**
- Transverse spatial modulation (TSM), **6.11–6.12, 6.23, 6.30, 6.31**
- Trap-assisted thermal generation current, **13.69**
- Traveling wave electro-optic modulators, **7.28–7.30, 7.29f**
- Traveling wave pumping, by EUV lasers, **58.3**
- Traveling-wave amplification, **19.3, 19.4f**
- Tree topologies, of WDM networks, **21.6, 21.7**
- Trivalent rare-earth ions, **2.11**
- Tube voltage, of x-ray tube sources, **54.9**
- Tunable dispersion compensation, in WDM networks, **21.23–21.26, 21.24f–21.26f**
- Tunable filters, acousto-optic, **6.23t, 6.35–6.45**
 collinear beam, **6.43, 6.43f, 6.45t**
 and longitudinal spatial modulation, **6.12**
 long-infrared, **6.42**
 mid-infrared, **6.42**
 noncritical phase-matching, **6.37, 6.39–6.42, 6.38f, 6.43t**
 principle of operation, **6.36–6.39, 6.37f, 6.38f**
 ultraviolet, **6.42**
- Tunable lasers, for fiber optic systems, **13.32–13.36, 13.33f–13.36f**
- Tungsten silicide/silicon (WSi_2/Si) bilayers, in MLLs, **42.6, 42.7**
- Tuning relation, of NPM AOTFs, **6.39–6.40**
- Turbulence:
 and adaptive optics, **5.5–5.21**
 anisoplanatism, **5.19**
 atmospheric tilt and Strehl ratio, **5.14, 5.14f–5.15f**
 Fried's coherence diameter and spatial scale, **5.9–5.13, 5.11t, 5.12f, 5.13f**
 higher-order phase fluctuations, **5.18–5.19**
 imaging, **4.35–4.36, 5.19–5.21, 5.19f, 5.20f**
 Kolmogorov model, **5.5–5.6**
 tracking requirements, **5.15–5.18, 5.17f**

- Turbulence, and adaptive optics (*Cont.*):
 variation of n and C^2n parameters,
 5.6–5.8, 5.7f, 5.8f
 in atmospheric optics, 3.26, 3.28–3.36
 beam spreading, 3.32–3.33, 3.33f
 beam wander, 3.31–3.32
 imaging and heterodyne detection, 3.34
 parameters for, 3.28–3.31, 3.29f, 3.30f
 scintillation, 3.34–3.36, 3.35f
 Hufnagel model of, 3.29
 Hufnagel-Valley model of, 3.30, 5.7, 5.8
 imaging through atmospheric, 4.1–4.37
 aberration variance and approximate
 Strehl ratio for, 4.27–4.28, 4.28f
 adaptive optics, 4.35–4.36
 angle of arrival fluctuations, 4.23–4.26,
 4.25f, 4.26f
 expansion coefficients, 4.20–4.22, 4.21t,
 4.22f–4.23f
 Kolmogorov turbulence and atmospheric
 coherence length, 4.7–4.10, 4.8f, 4.9f
 long-exposure images, 4.3–4.7
 modal correction of turbulence,
 4.28–4.30, 4.29t, 4.30f
 modal expansion of aberration function,
 4.17–4.20, 4.17f–4.18f, 4.19t, 4.20t
 and resolution of telescopes, 4.2–4.3
 short-exposure image, 4.31–4.35,
 4.31f–4.34f, 4.35t
 systems with annular pupils, 4.10–4.16,
 4.11f–4.15f, 4.15t
 inner scale of, 4.7
 Kolmogorov, 4.3, 4.7–4.10, 4.8f, 4.9f, 4.27,
 4.30, 4.36
 Kolmogorov model of, 5.5–5.6, 5.11
 optical strength of (C^2n), 5.6–5.8, 5.7f, 5.8f
 outer scale of, 4.7
 phase structure function of, 4.5
 Turn-on delay, of laser diodes, 13.13–13.14,
 13.14f
 Twist, of liquid crystals, 8.22, 8.23f
 Twist grain boundary (TGB) phases, of liquid
 crystals, 8.11
 Twisted nematic (TN) cells, 8.16, 8.25–8.26, 8.26f
 2D profilometry, 46.6
 Two-wave interaction (acousto-optic), 6.8
 Tyler frequency, 5.17
 UHURU satellite, 47.1
 Ultrafast nonlinear interferometers (UNIs), 19.32
 Ultrahigh-speed OTDM, 20.23–20.24, 20.24f
 Ultrashort pulses (USPs), for fiber lasers,
 25.32–25.33
 Ultraviolet (UV) AOTFs, 6.42
Umweganregung (multiple-beam Bragg
 diffraction), 43.6–43.8, 43.7f
 Unbalanced nonlinear interferometers (UNI),
 20.22
 Undulators, 55.12–55.14, 55.12f, 55.13f, 58.1
 Unitraveling-carrier (UTC) photodiodes,
 13.68–13.69, 13.68f
 Unity divergence ratio, for AO modulators,
 6.32–6.33
 Unruh radiation, 58.2
 Upper Atmospheric Research Satellite (UARS),
 3.36, 3.37f
 Ur (code), 55.14
 Urgent (code), 55.14
 U.S. Air Force Cambridge Research
 Laboratories, 3.22
 U.S. Department of Defense, 5.27
 U.S. National Institute of Standards and
 Technology (NIST) database, 3.26
 U.S. Standard Atmosphere, 3.5, 3.6, 3.8f,
 3.9f
 Vacuum spark source, of z-pinch radiation,
 57.3, 57.3t
 Vacuum ultraviolet (VUV) region, gratings
 and monochromators in, 38.1–38.8, 38.2f,
 38.3f, 38.4t–38.5t
 Van der Waals theory, 8.23
 Vapor axial deposition (VAD), 25.3, 25.26
 Variable optical attenuators (VOAs), 21.12,
 21.13f
 Verdet constant, 18.7
 Vertical alignment (VA) cells, 8.25, 8.27–8.28,
 8.28
 Vertical cavity surface-emitting lasers
 (VCSELs), 13.1–13.2, 13.43f
 commercial, 13.48
 electrical injection and current confinement
 for, 13.45, 13.45f
 light out vs. current in (L-I curve) of,
 13.46–13.47
 and linear optical amplifiers, 19.14
 and mirror reflectivity, 13.44
 and optical fibers, 9.7n
 polarization of, 13.48

- Vertical cavity surface-emitting lasers (VCSELs)
 (*Cont.*):
 and quantum wells, 13.43
 and spatial characteristics of emitted light, 13.46
 spectral characteristics of, 13.47–13.48
 Vertical coupling, of electroabsorption modulators, 13.60
 Virtual tributaries, of SONET, 23.6
 Viscosities, of liquid crystals, 8.23–8.25, 8.24*f*
 Voigt function, 56.7
 Voigt lineshape profiles, 3.23
 Voltage, of x-ray tube sources, 54.9
 Voltage-controlled oscillators (VCOs), 20.11, 20.11*f*
 Volume Bragg gratings (VBGs), 25.29, 25.30
 Volume diffraction calculations, for MLLs, 42.4–42.5, 42.5*f*
 Von Kármán spectrum, 3.28, 5.6
 V-parameter, of fiber lasers, 25.3
- W. M. Keck Observatory, 5.27
 Wadsworth condition, 38.6
 Wall effect, in gas detectors, 63.32
 Water:
 scattering in, 63.10, 63.10*t*, 63.11*f*
 in standard atmosphere, 3.6, 3.10–3.11, 3.11*f*
 Wave propagation, reciprocity of, 4.9–4.10
 Wave structure function, 4.6, 5.9*n*
 Wavefront correctors, for adaptive optics, 5.37–5.38, 5.38*f*
 Wavefront errors, 4.35, 5.40–5.41
 Wavefront sensing techniques, 5.36–5.37
 Wavefront tilt (*Z*-tilt), 4.3, 4.23–4.25, 5.14, 5.15
 Wavefronts, gradients of, 5.23
 Waveguide confinement factor, 13.4*f*, 13.5
 Waveguide modulators, 7.30–7.32, 7.31*f*–7.33*f*, 13.56, 13.57*f*
 Waveguides:
 attenuated total reflectance, 12.11
 double heterostructure, 19.3*f*
 evanescent-wave coupled pin, 13.68
 of SOAs, 19.15–19.16, 19.15*f*–19.16*f*
 Wavelength(s):
 attenuation vs., 15.7
 Bragg, 17.3, 20.15, 24.7
 of liquid crystals, 8.19–8.22, 8.20*f*
 pump, 14.5–14.6
 transverse effective, 11.10
 Wavelength blockers (WB), 21.12, 21.13*f*
 Wavelength dispersion, 9.6–9.7
 Wavelength dispersive detectors (WDS), 62.2
 Wavelength division multiplexing (WDM):
 dense
 and AOTFs, 6.43, 6.44
 and optical fiber amplifiers, 14.1
 and SOAs, 19.25–19.27, 19.25*f*, 19.26*f*
 and dispersion-managed solitons, 22.15–22.17
 and ESCON standard, 23.2
 and fiber-based couplers, 16.1, 16.4
 and solitons, 22.2, 22.8–22.12, 22.15–22.17
 (*See also* Wavelength-division multiplexing networks)
 Wavelength scans, for scatterometers, 1.14
 Wavelength selective switches (WSS), 21.12, 21.13*f*
 Wavelength-dispersive x-ray fluorescence (WDXRF), 29.2–29.3, 29.2*f*
 Wavelength-division multiplexing (WDM)
 networks, 18.4, 18.6, 21.1–21.44
 architecture of
 circuit and packet switching, 21.7–21.11, 21.8*f*–21.11*f*
 network reconfigurability, 21.12–21.13, 21.12*f*, 21.13*f*
 point-to-point links, 21.4
 star, ring, and mesh topologies, 21.5–21.7, 21.5*f*–21.7*f*
 wavelength-routed networks, 21.5, 21.5*f*
 fiber bandwidth, 21.2–21.3, 21.2*f*
 fiber system impairments, 21.13–21.26
 chromatic dispersion, 21.14–21.16, 21.15*f*, 21.16*f*
 dispersion and nonlinearities management, 21.20–21.26, 21.21*f*, 21.23*f*–21.27*f*
 fiber attenuation and optical power loss, 21.13–21.14
 fiber nonlinearities, 21.18–21.20, 21.19*f*, 21.20*f*
 polarization-mode dispersion, 21.16–21.18, 21.17*f*–21.18*f*
 history of, 21.1–21.2
 optical amplifiers in, 21.37–21.44, 21.37*f*
 EDFA, 21.38–21.41, 21.38*f*–21.42*f*
 Raman, 21.42–21.44, 21.42*f*–21.44*f*
 and optical fibers, 9.8, 9.12, 9.13, 9.15

- Wavelength-division multiplexing (WDM)
 - networks (*Cont.*):
 - optical modulation formats for, 21.27–21.36
 - basic concepts, 21.27–21.29, 21.28*f*–21.30*f*
 - carrier-suppressed return-to-zero and duobinary, 21.30–21.33, 21.31*f*, 21.32*f*
 - comparisons of, 21.36, 21.36*t*, 21.37*t*
 - DPSK and DQSK, 21.33–21.36, 21.33*f*–21.35*f*, 21.36*t*, 21.37*t*
 - in real systems, 21.3–21.4, 21.3*f*, 21.4*f*
- Wavelength-routed networks, 21.5, 21.5*f*
- Wavelength-selective couplers (WSCs), 14.2
- Waviness, in polycapillary x-ray optics, 53.4, 53.4*f*
- Wedge multilayer Laue lenses (wMLLs), 42.12–42.13, 42.13*f*, 42.14*f*
- Well and barrier intermixing, 13.60
- Wide area networks (WANs), 9.14, 21.7
- Wideband AO Bragg cells, 6.27, 6.30, 6.31*t*
- Wide-field lobster-eye telescopes, 48.4
- Wigglers, 55.11–55.12
- Window structure, of SOAs, 19.8, 19.8*f*
- Windowing, in x-ray mirror metrology, 46.9, 46.10*f*
- Wire array sources, of z-pinch radiation, 57.3–57.4, 57.3*t*, 57.4*f*
- Wollaston prism, 46.4
- Wolter geometries, 63.21
- Wolter mirror configurations, 52.4, 64.6, 64.6*f*
- Wolter optics, 26.10, 48.1, 48.2, 49.4–49.6, 49.5*f*–49.6*f*
- Wolter telescopes, 44.4, 44.5*f*, 44.6–44.10, 44.7*f*, 44.10*f*, 45.1–45.5, 45.2*f*
- Wolter x-ray optics, 47.2–47.7, 47.2*f*, 47.4*f*, 47.6*f*, 47.7*f*
- Wolter-Schwarzschild (WS) telescopes, 44.7, 44.11, 45.1
- XEUS/IXO mission, 49.7
- XMM mission, 44.6
- XMM-Newton observatory, 33.3, 33.3*t*, 47.2, 47.4*f*, 47.6, 47.6*f*
- X-pinch sources of radiation, 57.3, 57.3*t*, 57.4
- X-ray absorption fine structure (EXAFS), 29.3, 60.4
- X-ray absorption near edge spectroscopy (XANES), 30.2, 30.2*f*, 30.4*f*
- X-ray absorption near-edge structure (XANES), 29.3, 60.4
- X-ray absorption spectroscopy (XAS), 30.1–30.5, 30.2*f*–30.5*f*
- X-ray astronomy, 33.1–33.4, 33.3*t*
- X-ray astronomy satellite (SAX), 44.6
- X-ray detectors, 60.3–60.10
 - cryogenic, 60.8–60.9, 60.9*t*, 60.10*t*
 - film, 60.8, 60.9*t*, 60.10*t*
 - ionization, 60.3–60.7, 60.9*t*, 60.10*t*
 - scintillation, 60.7–60.8, 60.9*t*, 60.10*t*
- X-ray diffraction (XRD), 28.1–28.7, 28.3*f*–28.6*f*
- X-ray fluorescence (XRF), 29.1–29.11, 54.8
 - energy-dispersive, 29.3–29.11, 29.5*f*, 29.6*f*, 29.8*f*–29.11*f*
 - history of, 29.1
 - and polycapillary x-ray optics, 53.10–53.11, 53.11*f*
 - wavelength-dispersive, 29.2–29.3, 29.2*f*
 - and x-ray diffraction, 28.1
 - and x-ray imaging, 62.5, 62.6*f*
- X-ray imaging detectors, 61.1–61.8, 61.2*f*
 - CCD detectors, 61.7–61.8, 61.8*f*
 - flat panel detectors, 61.3–61.7, 61.4*f*, 61.6*t*, 61.7*f*
- X-ray lasers, 58.1–58.4, 58.3*f*
- X-ray mapping, 62.4, 62.5, 62.6*f*
- X-ray microscopes, 37.6
- X-ray mirrors:
 - and grazing incidence optics, 44.3–44.6, 44.3*f*–44.6*f*
 - metrology of, 46.1–46.12
 - history of, 46.1–46.2
 - profile analysis considerations, 46.6–46.12, 46.7*f*, 46.10*f*
 - surface figure metrology, 46.3–46.6, 46.5*f*
 - surface finish metrology, 46.2
- X-ray monochromators, 30.1–30.4, 50.6–50.7
- X-ray observatories, 33.1–33.4, 33.3*t*
- X-ray optics:
 - adaptive, 50.1–50.8
 - hard vs. soft x-ray telescopes, 50.2
 - history of, 50.1, 50.2*f*
 - synchrotron and lab-based sources, 50.2–50.8, 50.4*f*, 50.6*f*, 50.8*f*
 - astronomical, 47.1–47.11
 - angular resolution of, 47.10–47.11, 47.11*f*
 - hard, 47.9–47.10

- X-ray optics, astronomical (*Cont.*):
 history of, 47.1–47.2
 Kirkpatrick-Baez optics, 47.7–47.8, 47.8f, 47.9f
 Wolter, 47.2–47.7, 47.2f, 47.4f, 47.6f, 47.7f
 coherent, 27.1–27.5, 27.3f–27.5f
 gratings and monochromators in, 38.1–38.8
 diffraction properties, 38.1–38.3, 38.2f
 dispersion properties, 38.6–38.7
 efficiency, 38.8
 focusing properties, 38.3–38.6, 38.3f, 38.4t–38.5t
 resolution properties, 38.7
 and inverse Compton x-ray sources, 59.1–59.4, 59.3t
 and medical imaging, 31.1–31.4, 31.2f–31.4f
 monocrapillary, 52.1–52.6, 52.2f–52.5f, 52.2t
 multifoil, 48.1–48.4, 48.2f, 48.3f
 and neutron optics, 26.5–26.11, 26.8f, 26.9f, 26.11f, 36.2f
 polycapillary, 53.1–53.19, 53.2f
 alignment and measurement, 53.5–53.8, 53.6f, 53.7f
 collimation, 53.8–53.9, 53.8f–53.9f
 energy filtering, 53.10
 focusing, 53.9–53.10, 53.10t
 monochromatic imaging, 53.16–53.17, 53.17f
 powder diffraction, 53.14, 53.14f
 radiation resistance, 53.5
 scatter rejection in imaging, 53.14–53.16, 53.15f–53.16f
 scintigraphy, 53.17, 53.18, 53.18f
 simulations and defect analysis, 53.3–53.5, 53.3f–53.5f
 single crystal diffraction, 53.12–53.14, 53.12f, 53.13f
 therapy, 53.18–53.19
 x-ray fluorescence, 53.10–53.11, 53.11f
 and pore optics, 49.1–49.7
 ray tracing for, 35.1–35.6, 35.4f, 35.5f
 and Schwarzschild objective, 51.3, 51.3f–51.4f
 spectral detection and imaging in, 62.1–62.5, 62.2f–62.6f
 and x-ray properties of materials, 36.1–36.9
 Auger energies, 36.3t, 36.9t
 electron binding energies, 36.3t–36.6t
- X-ray optics, and x-ray properties of materials (*Cont.*):
 photoabsorption and scattering, 36.1
 photon energies, 36.7t–36.8t
 x-ray and neutron optics, 36.2f
- X-ray spectral detection and imaging, 62.1–62.5, 62.2f–62.6f
- X-ray tube sources, 54.3–54.16
 brightness and intensity of, 54.11–54.15, 54.13f–54.15f
 cathode design and geometry, 54.10–54.11
 characteristics of, 54.3, 54.4f
 optimization of, 54.15–54.16
 spectra of, 54.4–54.10, 54.5f, 54.8f
- X-rays:
 circularly polarized soft, 55.6–55.7
 nanofocusing of hard, 42.1–42.17
 history of, 42.2–42.4, 42.2f, 42.3f
 instrumental beamline arrangement and measurements for, 42.9–42.10, 42.9f–42.12f
 limitations of, 42.15–42.17, 42.16f–42.17f
 with magnetron-sputtered MLLs, 42.5–42.7, 42.6f–42.8f
 on MLLs with curved interfaces, 42.14, 42.15f
 Takagi-Taupin calculations for, 42.12–42.14
 volume diffraction calculations for, 42.4–42.5, 42.5f
 with wedged MLLs, 42.12–42.13, 42.13f, 42.14f
 polarization of, 43.1–43.2
- Y-branch interferometric modulators, 13.51–13.52
- Young's modulus, of infrared optical fibers, 12.3t, 12.9
- Ytterbium-doped fiber amplifiers (YDFAs), 14.2, 14.7
- Ytterbium-doped fibers, 25.23t, 25.24–25.25, 25.31, 25.33
- ZBLAN (fluorozirconate glass), 12.5, 12.5f
 and fiber lasers, 25.3, 25.24, 25.27t, 25.28
 fluoraluminate glass vs., 12.3t, 12.4, 12.4t
 Zeeman effect, 2.21
 Zeeman spectroscopy, 2.23–2.24, 2.24f
 Zemax (program), 35.1
 Zernike annular expansion coefficients, 4.25

- Zernike coefficients, variance of, **4.22**, **4.22f**,
 4.23f, **4.36**
- Zernike decomposition elements, **5.37**
- Zernike modes of aberrations, **5.11**, **5.11t**, **5.12f**
- Zernike polynomials, **4.17–4.20**, **4.20t**, **5.10**, **46.6**
- Zernike tilts, **4.23–4.26**
- Zero order, of radiation, **40.1**
- Zerodur glass shells, **47.5**
- Zero-phonon lines, in solids, **2.13–2.14**, **2.14f**
- Zonal approach, to wavefront error correction,
 4.35
- Zone plate law, **42.4**
- Zone plates, **40.1–40.10**
 - amplitude, **40.4–40.5**, **40.5f**
 - and Bragg-Fresnel lenses, **40.9–40.10**, **40.9f**
 - diffraction efficiencies of, **40.4–40.8**, **40.5f**,
 40.7f, **40.8f**
 - Fresnel, **40.2**, **42.3**, **42.3f**, **55.16**
 - geometry of, **40.1–40.3**, **40.2f**
 - neutron, **63.25**
 - phase, **40.5–40.7**, **40.7f**
 - as thin lenses, **40.3–40.4**
- Z-pinch plasma, **57.1–57.5**, **57.2f**, **57.3t**, **57.4f**
- Z-tilt (wavefront tilt), **4.3**, **4.23–4.25**, **5.14**, **5.15**

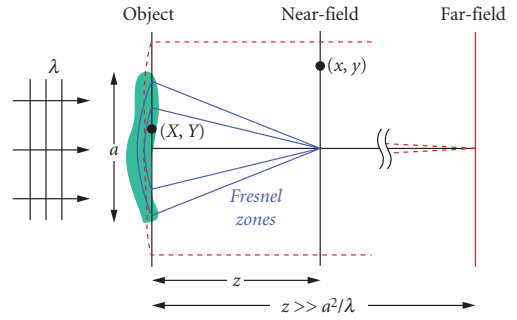


FIGURE 27.1 Schematic illustration of coherent x-ray wave propagation with a distorted object approach both for nearfield Fresnel diffraction, where an object extends into multiple Fresnel zones (solid lines), and for far-field Fraunhofer diffraction, where an object occupies only the center of the first Fresnel zone (dashed lines).

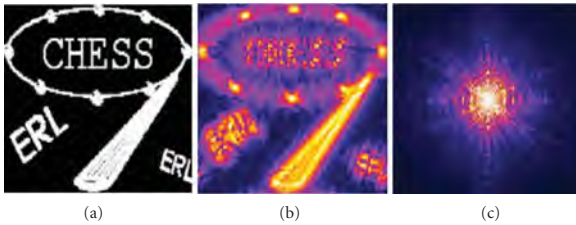


FIGURE 27.2 Simulated diffraction amplitudes $|F(x, y)|$, of an amplitude object (a) of $10 \mu\text{m} \times 10 \mu\text{m}$, with $l = 1 \text{ \AA}$ x rays, at image-to-object distance (b) $z = 2 \text{ mm}$ and (c) $z = \infty$, using the unified distorted object approach Eq. (3) with $Nz = 500$ zones in (b) and $Nz = 0$ in (c). Notice that the diffraction pattern changes from noncentrosymmetric in the near-field (b) to centrosymmetric in the far-field (c).

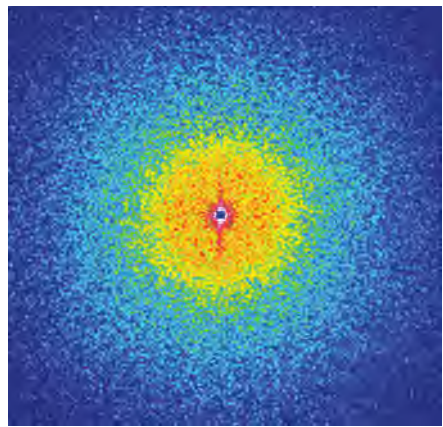


FIGURE 27.3 Example of a coherent x-ray diffraction pattern from a gold nanofoam specimen of $\sim 2 \mu\text{m}$ in size, using 7.35-keV coherent x rays. The corner of the image corresponds to $\sim 8 \text{ nm}$ spatial frequency.

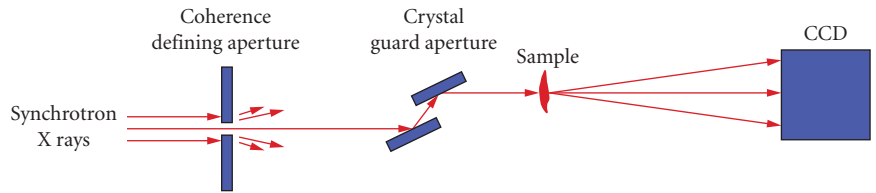


FIGURE 27.4 Concept of a perfect-crystal guard aperture in coherent diffraction imaging experiments for the purpose of eliminating unwanted parasitic scattering background in order to achieve high signal-to-noise in a diffraction pattern.

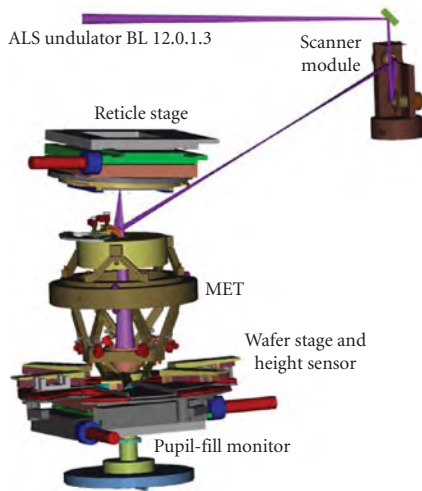


FIGURE 34.1 EUV exposure tool. The design includes 4 mirrors, and the mask (reticle) and wafer location. The whole system is under vacuum, and the condenser optical system is not shown. The whole system is enclosed in a vacuum chamber, not shown. The overall size is of several meters. This tool is installed at the Advanced Light Source, Berkeley.



FIGURE 34.2 ASML alpha demo tool during installation. The whole system is under vacuum, and a plasma source is used to generate the EUV radiation (not shown).

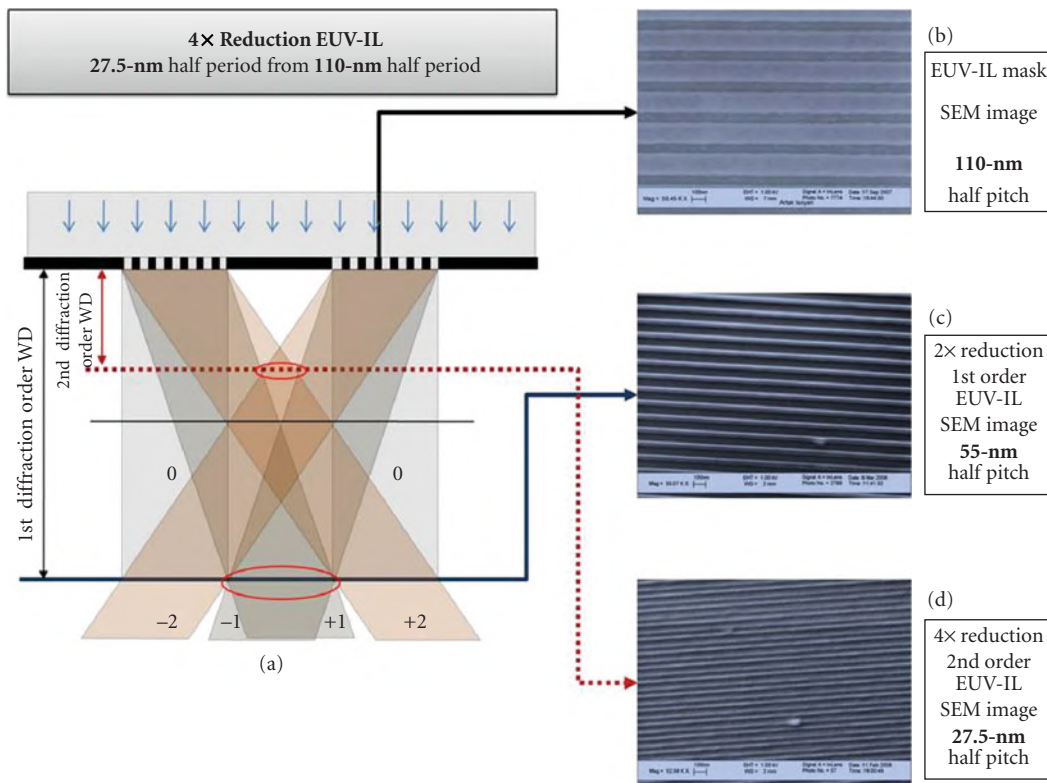


FIGURE 34.3 EUV interferometric lithography. The diffraction gratings are illuminated by a synchrotron, and the diffracted beams interfere as shown. The beams overlap creating 1st and 2nd order interference patterns of excellent visibility. Right, SEM images of the grating, and of the first- and second-order exposures. Notice the relative period of the images—the 1× period is half of that of the diffracting grating, and the 2× is 1/4.

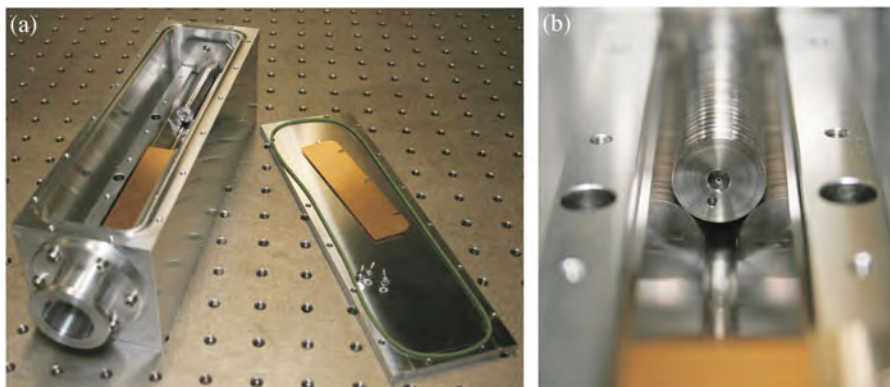


FIGURE 37.2 (a) Housing with partly assembled Be lens, (b) stack of Be lenses. Each individual lens is centered inside of a hard metal coin. They are aligned along the optical axis by stacking the coins in a high precision v-groove.

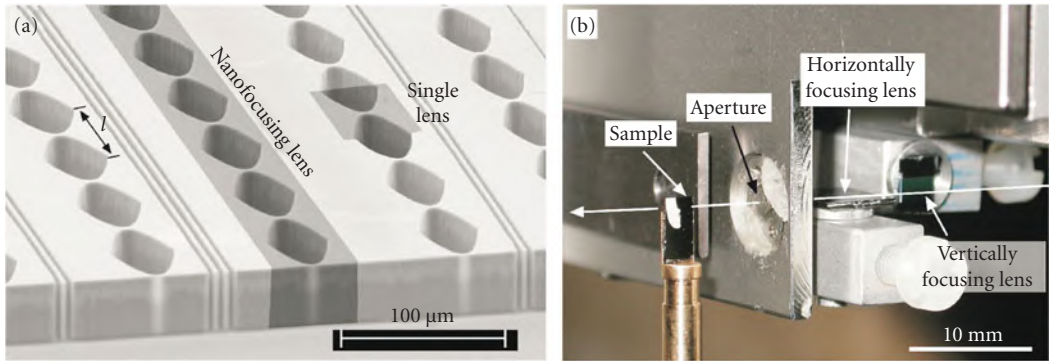


FIGURE 37.5 (a) Array of nanofocusing lenses made of silicon. A large number of single lenses are aligned behind each other to form a nanofocusing lens. Several nanofocusing lenses with different radius of curvature R are placed in parallel onto the same substrate. (b) Scanning microprobe setup with two crossed nanofocusing lenses. An aperture defining pinhole is placed behind the second lens.

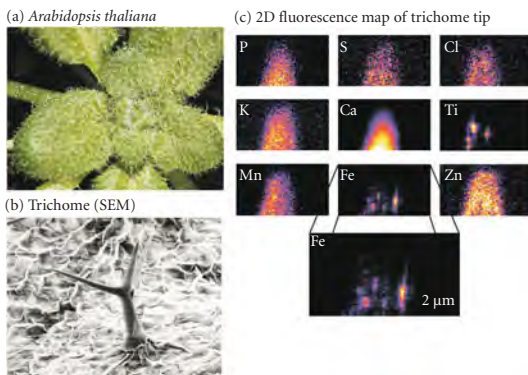


FIGURE 37.7 (a) Photograph of the plant *Arabidopsis thaliana*, (b) secondary electron micrograph of a leaf hair (trichome), (c) two-dimensional fluorescence map of the tip of the trichome at 100-nm spatial resolution. While most elements are homogeneously distributed, iron (Fe) and titanium (Ti) are localized on the level of 100 nm.

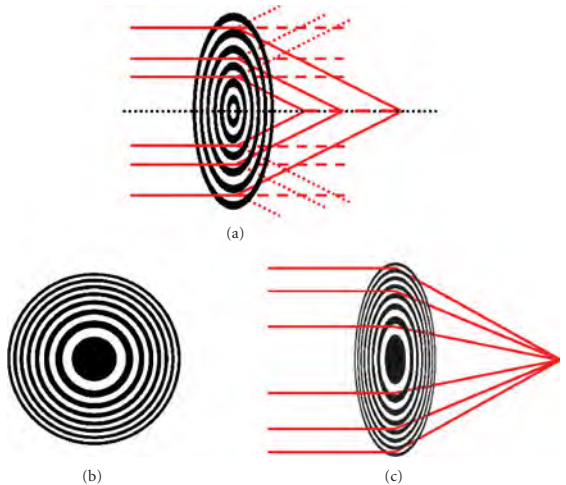


FIGURE 40.1 Diffraction by (a) a circular grating of constant period and (b, c) a zone plate.

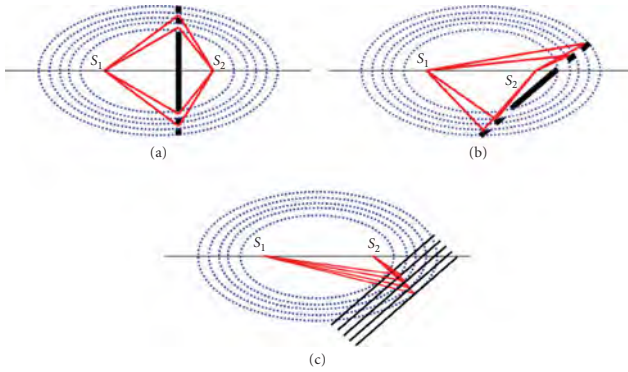


FIGURE 40.6 Structure of Bragg-Fresnel lenses.

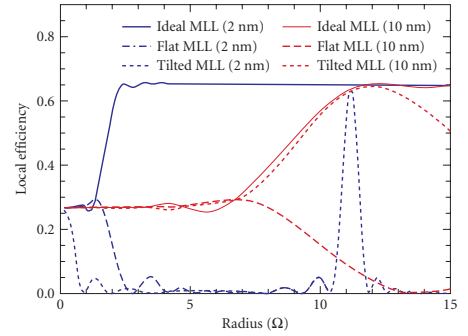


FIGURE 42.3 Calculated diffraction efficiency at 0.064 nm (19.5 keV) for outermost zone widths of both 10 nm (red) and 2 nm (blue) as a function of radius for ideal (wedged), flat, and tilted MLLs. For flat MLLs efficiencies do not exceed 26 percent and only very low-order zones diffract in the 2-nm case. For an outermost zone width of 10 nm, the ideal and tilted cases have almost the same performance, but for a 2-nm outermost zone, the ideal case is far superior. For the 10-nm tilted case, only a sharp Bragg peak is seen at the radius for which a Bragg condition is satisfied. This figure shows that meeting the Bragg condition everywhere becomes increasingly important for outermost zones less than 10 nm in order to ensure a high efficiency throughout most of the MLL.

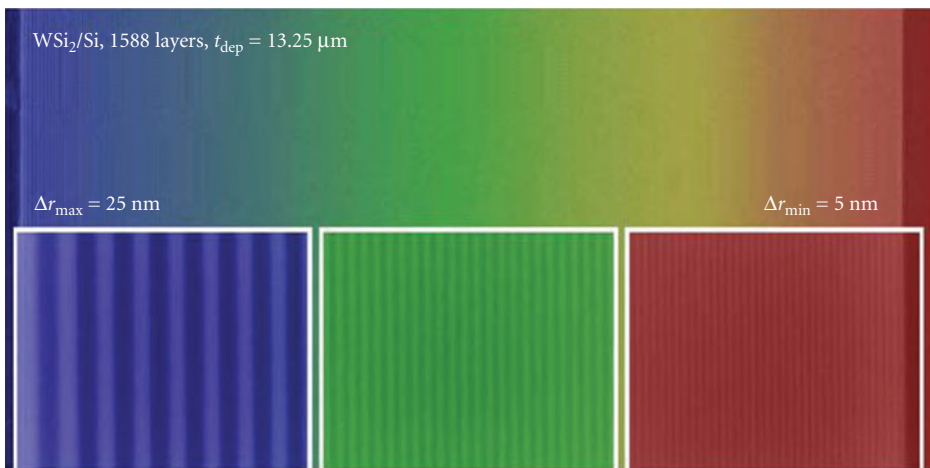


FIGURE 42.5 A scanning electron microscope image of a partial MLL having a 5-nm-thick outermost zone. The structure was used to obtain a focus of 16 nm for x rays of 0.064-nm (19.5-keV) wavelength.

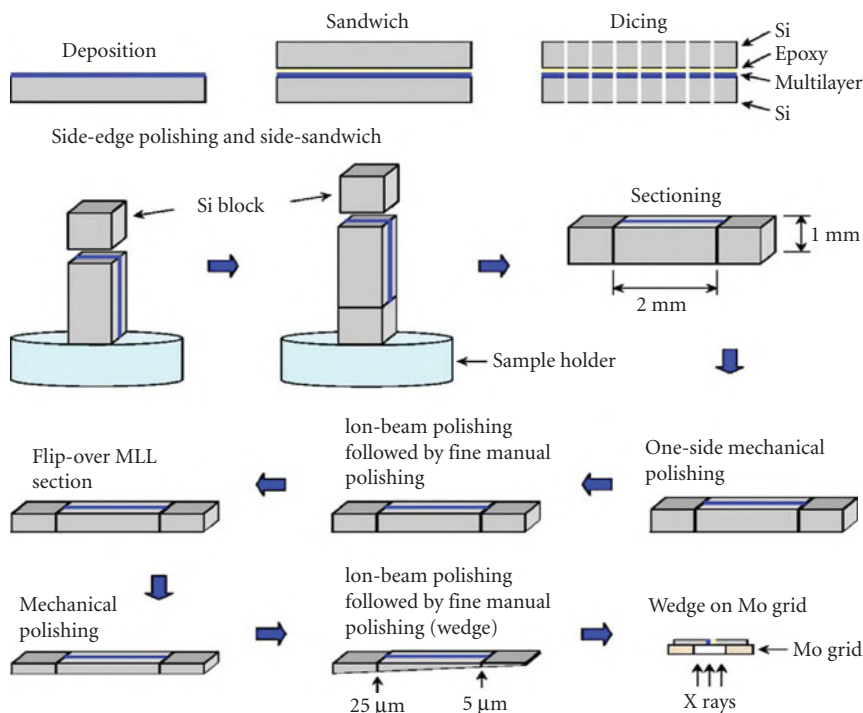


FIGURE 42.7 Cartoon illustrating the sequence of steps used to process a lens suitable for use with x rays, starting from an as-sputtered MLL wafer.

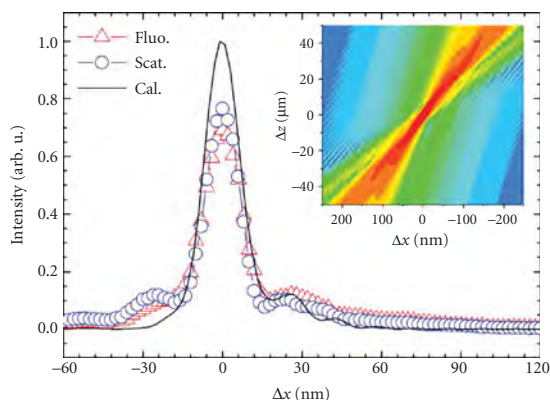


FIGURE 42.9 Measured and calculated intensity profiles for the focus of the MLL shown in Fig. 42.5. The FWHM values are 17.6 nm from fluorescence data and 15.6 nm from far-field scattering data. These should be compared to a calculated value of 15.0 nm. The calculated intensities were scaled according to measured and calculated efficiencies of 30 and 32 percent, respectively. The calculated results do not include the effects of vibrations, finite analyzer width, and finite sources size. The inset shows the calculated isophote pattern near the focal plane.

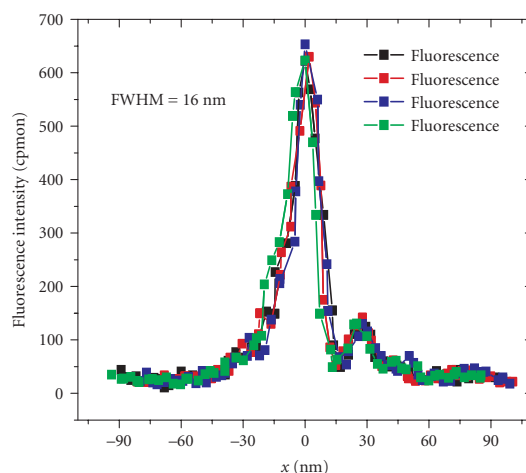


FIGURE 42.12 Measured line focuses of the MLL in Fig. 42.5 with 0.042-nm (19.25-keV) x rays. Results from four scans of the fluorescence intensity are shown. The efficiency was measured to be 17 percent. A FWHM value of 16 nm applies.

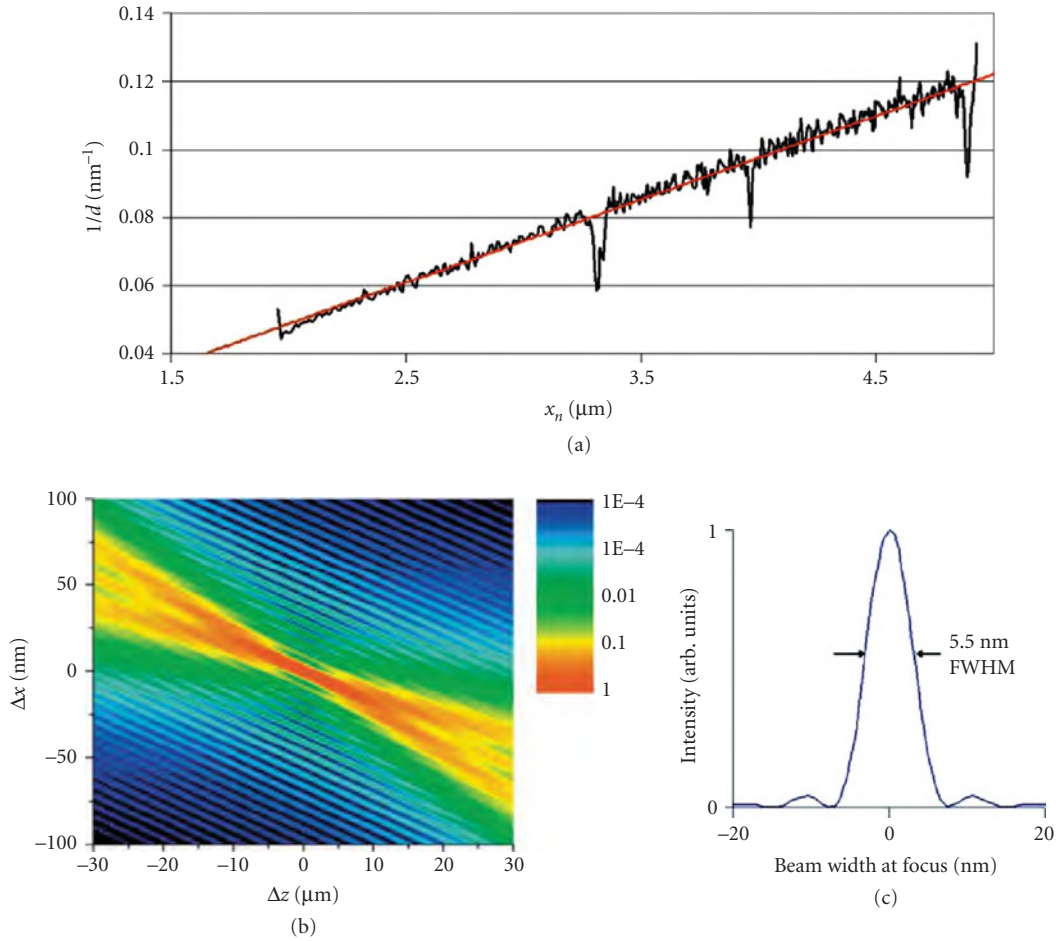


FIGURE 42.14 Analyses for ideal (wedged) MLL structure of Fig. 42.13a. The inverse layer thickness vs. radius is shown in (a). The calculated isophote pattern and intensity in the focal plane are shown in (b) and (c), respectively.

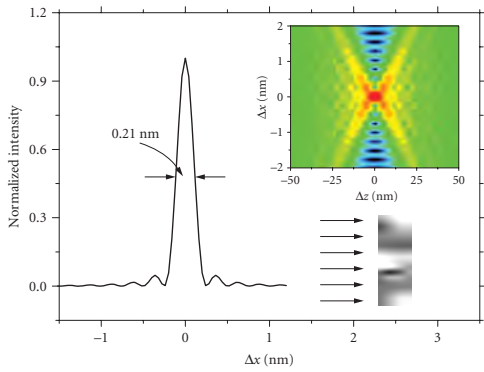


FIGURE 42.15 Calculated intensity in the focal plane for an MLL having parabolically shaped interfaces and an outermost zone width of 0.25 nm. The lower inset shows a cartoon of the MLL, and the upper inset shows an isophote pattern around the focus.

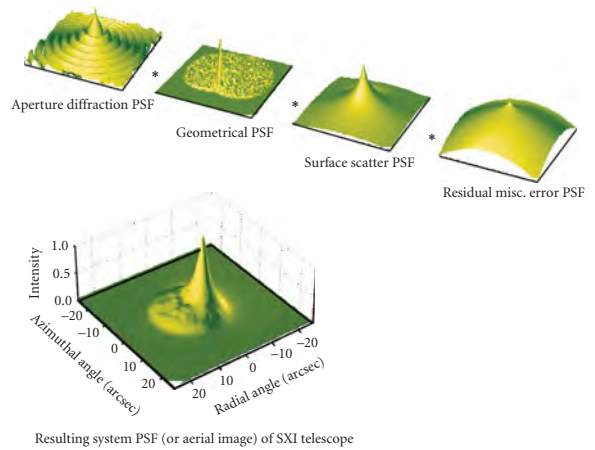


FIGURE 44.10 Illustration of the PSFs for the individual image degradation mechanisms and their resulting convolution (*aerial image*) of the SXI telescope for a field angle of 15 arcmin and a wavelength of 44.7 Å.

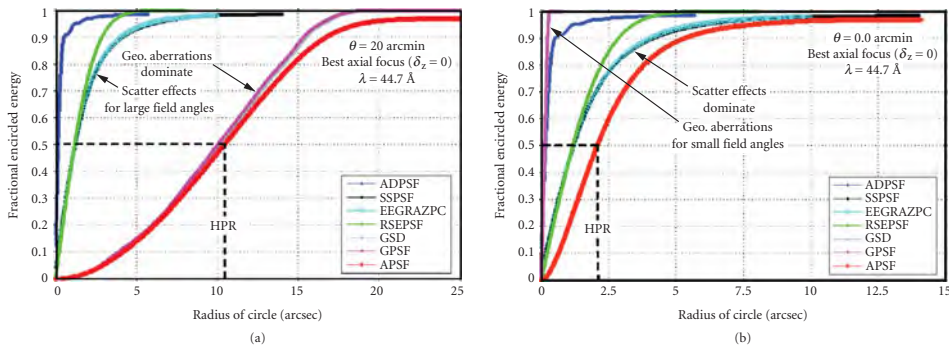


FIGURE 44.11 The fractional encircled energy of the aerial image and the four functions contributing to it provide insight into the image quality of a grazing incidence x-ray telescope. Note that scatter effects dominate geometrical aberrations for small field angles and geometrical aberrations dominate scatter effects for large field angles.

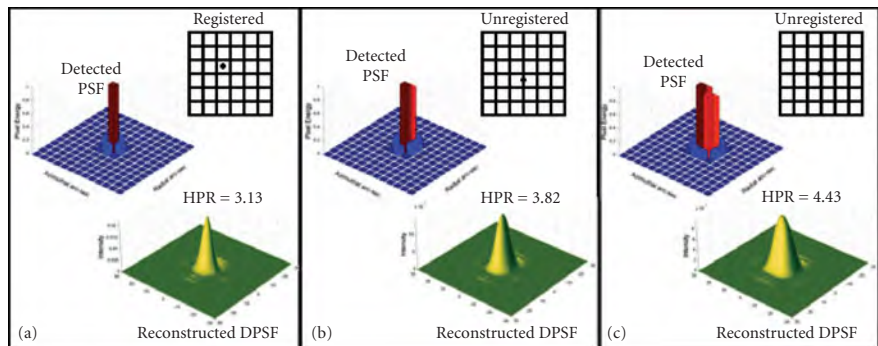


FIGURE 44.12 The DPSF and the reconstructed DPSF for (a) precisely “registered” aerial PSF, (b) aerial PSF centered on boundary between two pixels, and (c) aerial PSF positioned where four pixels meet.

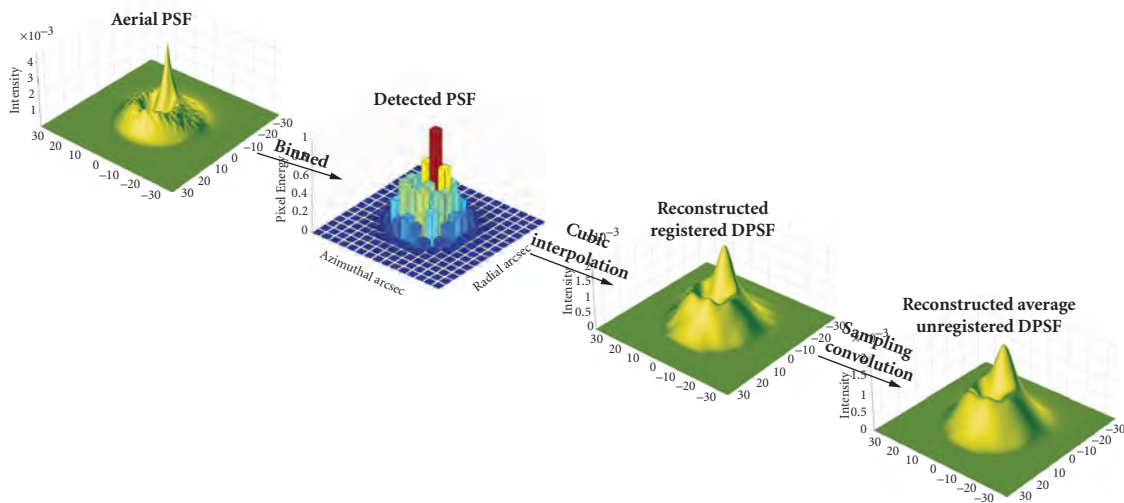


FIGURE 44.13 A graphical illustration of the numerical computation technique for modeling both the reconstructed “registered” DPSF and the reconstructed “average unregistered” DPSF is indicated.

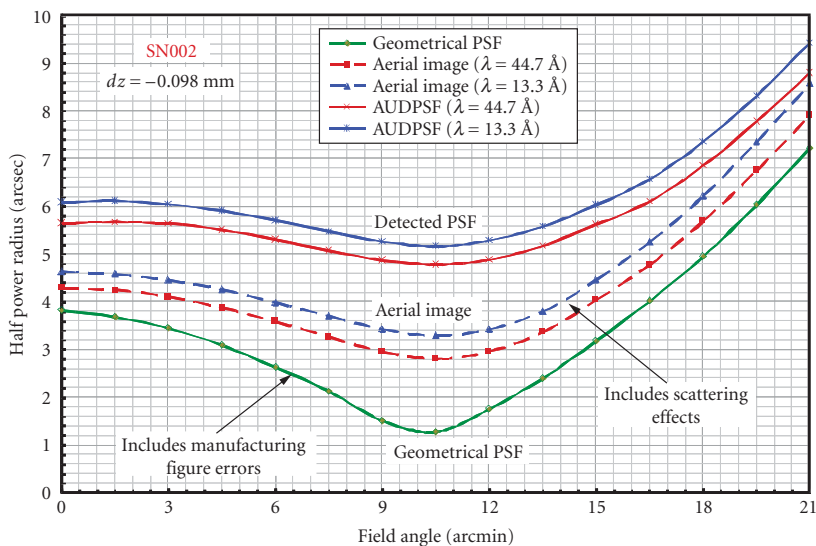


FIGURE 44.14 Comparison of the predicted HPR versus field angle of the geometrical PSF, the aerial image, and the average unregistered detected PSF for the SXI telescope at two different wavelengths (44.7 and 13.3 Å).

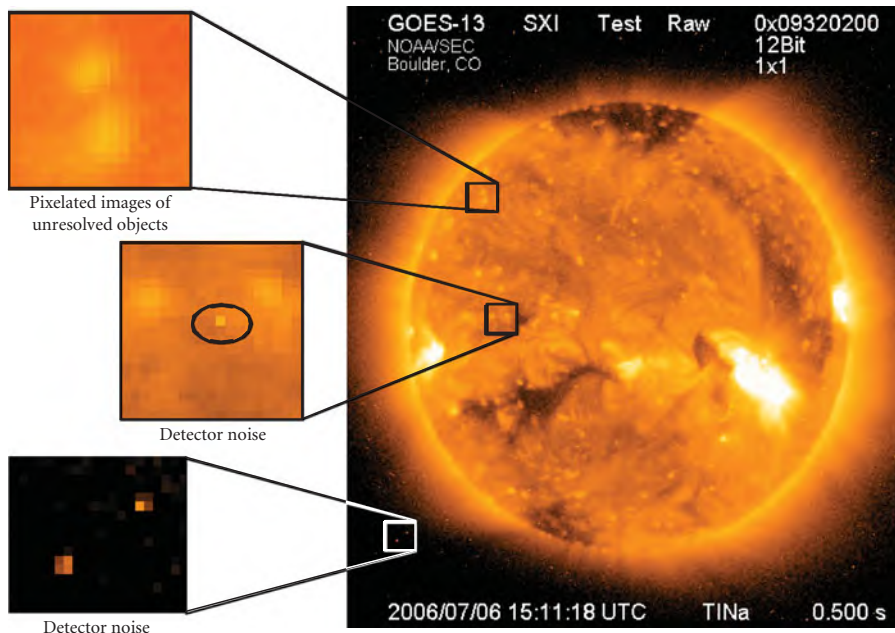


FIGURE 44.15 On-orbit solar image with three small regions highlighted and magnified for detailed inspection. This allows one to distinguish between images of unresolved bright features on the sun and merely detector noise.

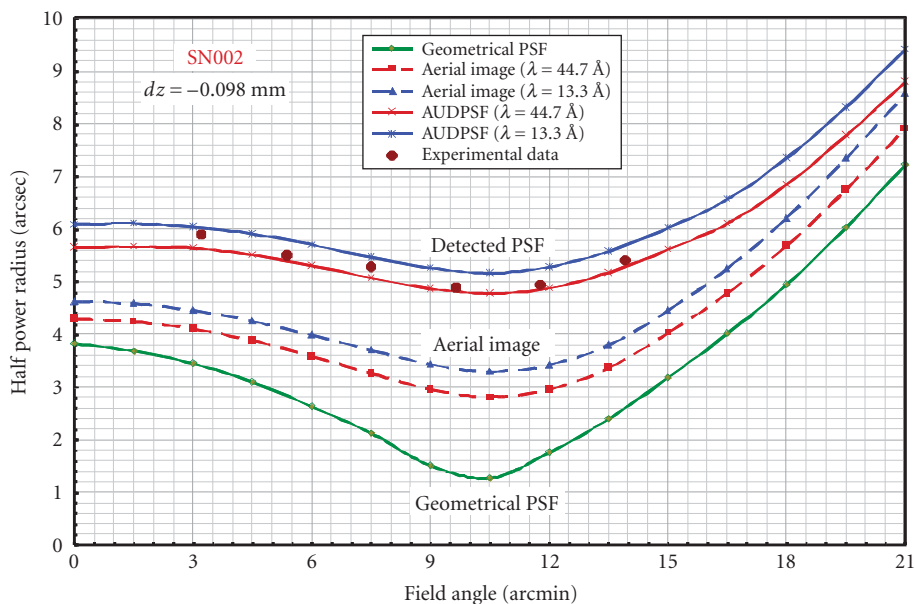


FIGURE 44.16 Experimental validation of an exhaustive systems engineering analysis of image quality for grazing incidence x-ray telescopes, including the modeling of surface scatter and detector effects.

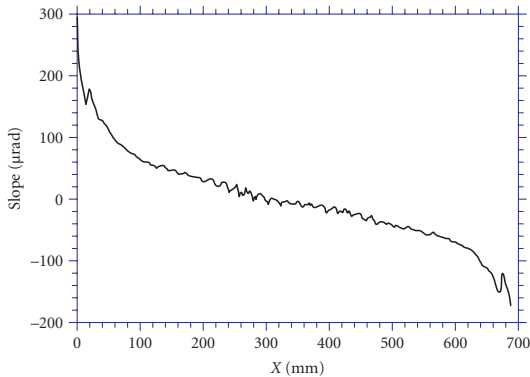


FIGURE 46.2 Slope profile measurement on a 700-mm long silicon cylinder mirror made with an LTP. Sampling step size is 2 mm. Mean has been subtracted from the data (detrend 0). Profile shows that the surface has an overall convex curvature (tilted profile down to the right) with significant edge roll-off (change in slope at each end). Also, a polishing defect with a 20-mm period is evident in the center of the surface. The slope profile emphasizes high-frequency surface defects.

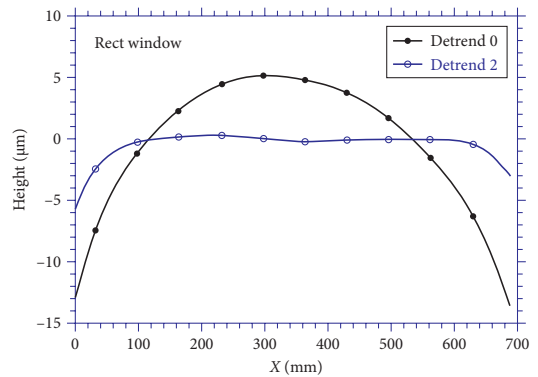


FIGURE 46.3 Height profile calculated by integrating the slope profile of Fig. 46.1. Solid circles: mean height subtracted (detrend 0); open circles: second order polynomial subtracted (detrend 2). The radius of curvature extracted from the second order term coefficient is 3.572 km. The residual profile shows that the surface has a “kink” in the center that separates it into two distinct segments with slightly different slopes. This low frequency defect is not evident in the slope profile of Fig. 46.1.

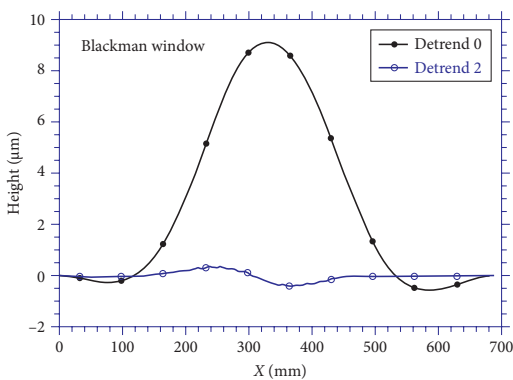


FIGURE 46.4 The height profiles of Fig. 46.3 with a Blackman window applied. The edge discontinuities are minimized by this function. Although the shape of the profile is distorted, the average statistical properties of the underlying function are not changed.

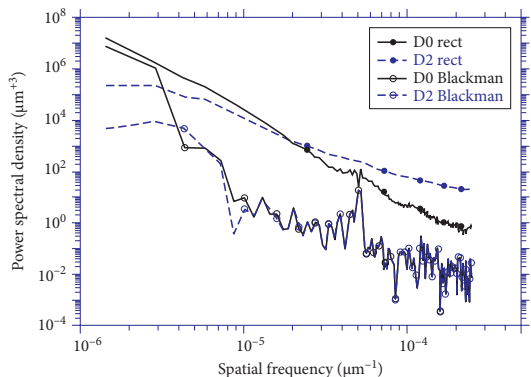
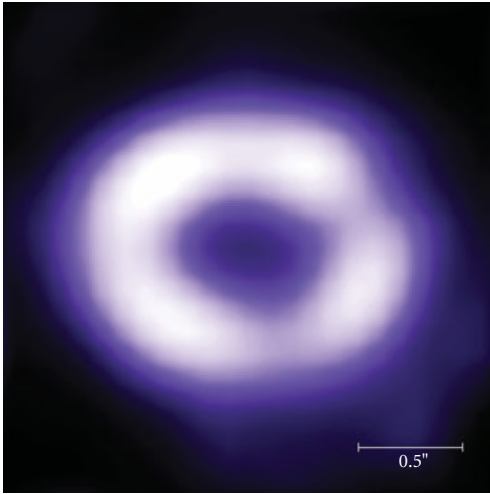
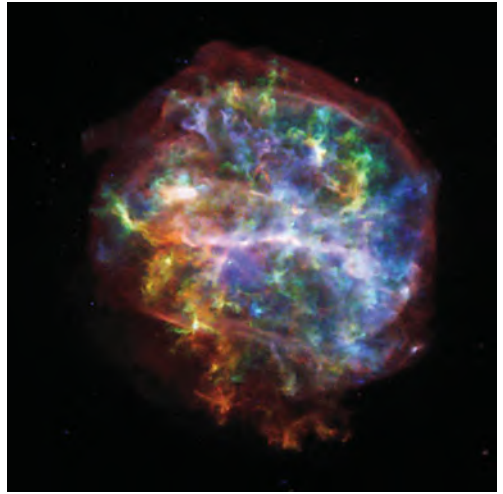


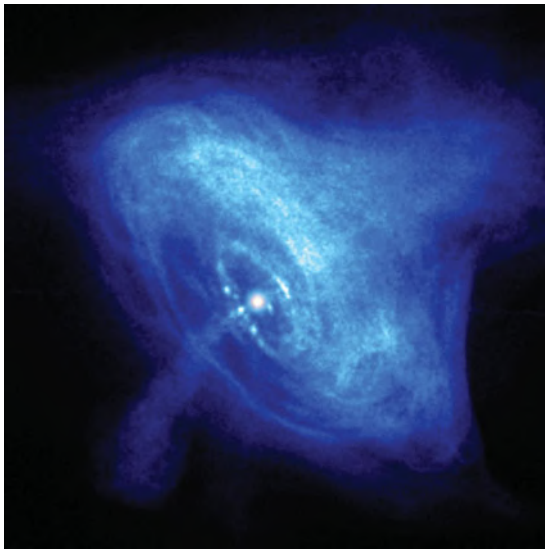
FIGURE 46.5 PSD curves computed from the four profiles in Figs. 46.3 and 46.4. The two upper curves from the unwrapped data show severe contamination effects due to the strong edge discontinuities that introduce spurious power into all frequencies. The lower curves show how the Blackman filter eliminates the discontinuity contamination, allowing the underlying surface spectral characteristics to become visible.



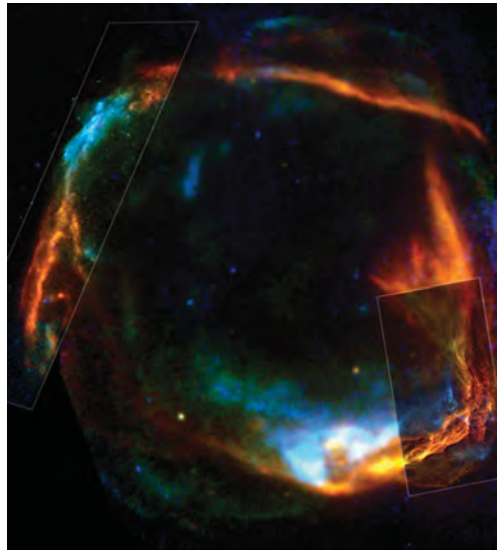
(a)



(b)



(c)



(d)

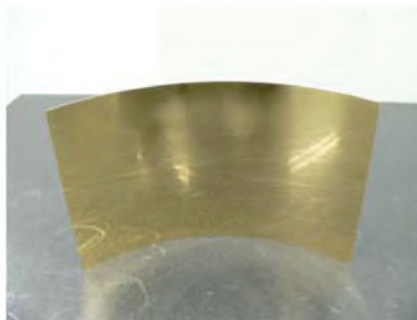
FIGURE 47.2 (a) A Chandra x-ray image of the fiery ring surrounding the 1987 supernova explosion in the Dorado constellation. Subarcsecond angular resolution is required to resolve the structure surrounding the supernova remnant (<http://chandra.harvard.edu/photo/2005/sn87a/>). (NASA/CXC/PSU/S.Park & D. Burrows.) (b) A Chandra x-ray image of the supernova remnant G292.0+1.8. The colors in the image encode the x-ray energies emitted by the supernova remnant; the center of G292.0+1.8 contains a region of high energy x-ray emission from the magnetized bubble of high-energy particles that surround the pulsar, a rapidly rotating neutron star that remained behind after the original, massive star exploded (<http://chandra.harvard.edu/photo/2007/g292/>). (NASA/CXC/Penn State/S.Park et al.) (c) Chandra x-ray image of the Crab Nebula—the remains of a nearby supernova explosion which was seen on Earth in 1054 AD. At the center of the bright nebula is a rapidly spinning neutron star, or pulsar, that emits pulses of radiation 30 times a second (<http://chandra.harvard.edu/photo/2002/0052>). (NASA/CXC/ASU/J. Hester et al.) (d) XMM-Newton and Chandra x-ray images RCW 86, an expanding ring of debris that was created after a massive star in the Milky Way collapsed onto itself and exploded. Both the XMM-Newton and Chandra images show low-energy x-rays in red, medium energies in green and high energies in blue. The Chandra observations focused on the northeast (left-hand) and southwest (lower right) side of RCW 86, and show that x-ray radiation is produced both by high-energy electrons accelerated in a magnetic field (blue) as well as heat from the blast itself (red). These images demonstrate the large field of view and moderate angular resolution of XMM-Newton, compared to the smaller field of view and high angular resolution provided by Chandra (<http://chandra.harvard.edu/photo/2007/2snr/>). (NASA/CXC/ESA/Univ. of Utrecht/J. Vink et al.)



FIGURE 47.3 An x-ray optics module for the XMM observatory. Fifty-eight electroformed nickel Wolter I optics are nested to increase the effective x-ray collecting area.



(a)



(b)

FIGURE 47.4 The segmented foil mirrors aboard the SUZAKU spacecraft. (a) A complete mirror module and (b) a single aluminum foil reflector coated with gold.

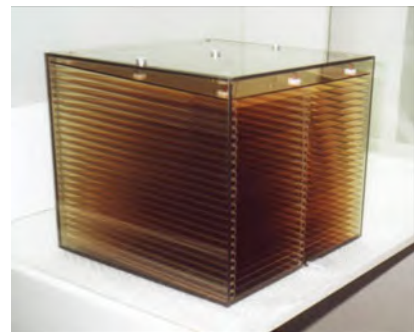


FIGURE 48.3a Kirkpatrick-Baez test Au coated glass foils system studied for XEUS project.

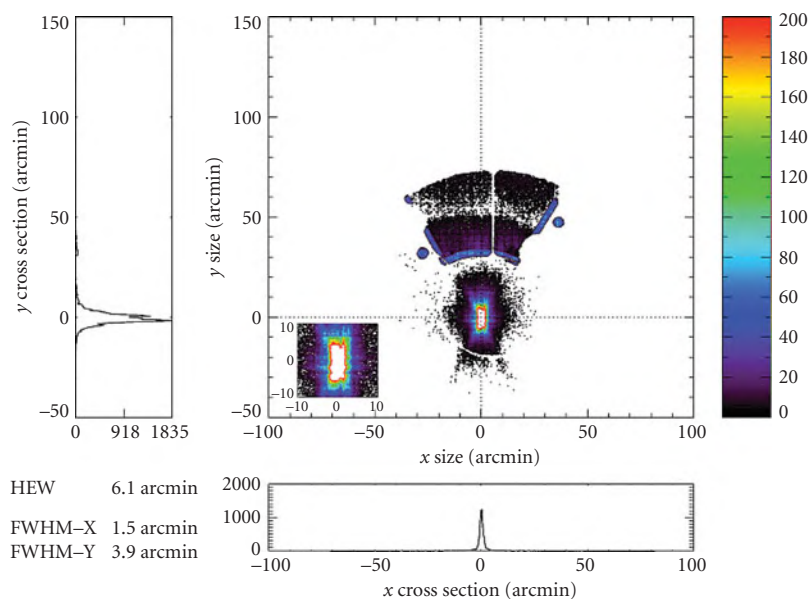


FIGURE 49.7 The focus of a single plate radially packed microchannel plate optic.

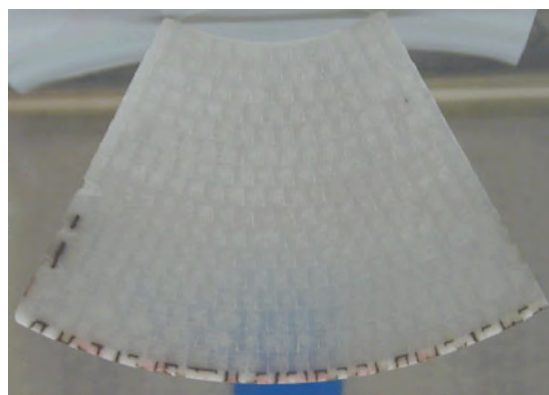


FIGURE 49.9 A segment of a radially stacked micropore optic. Two such plates behind each other act as a conical approximation to a Wolter-I optic. The width of the segment is 35 mm.

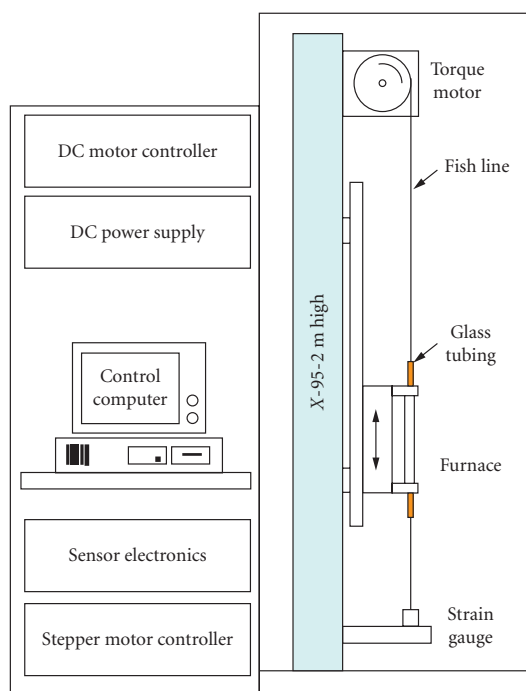


FIGURE 52.4 A glass tube is suspended in an electric furnace from a piece of fish line that is attached to a strain gauge at the bottom. The torque motor keeps a constant tension as the glass yields during drawing. The furnace is programmed to move based on the amount of glass yielding to make the desired elliptical, parabolic, etc., shape.

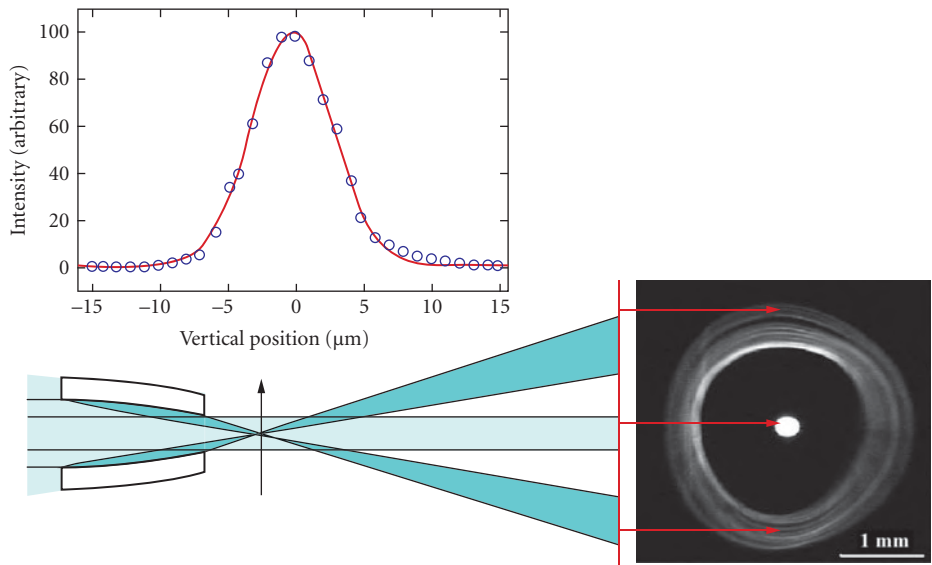


FIGURE 52.5 Upper panel: Profile of intensity versus $5\ \mu\text{m}$ vertical pinhole position at focus of a 9 milliradians capillary producing a spot size of $5\ \mu\text{m}$ FWHM at a distance of 20 mm beyond the tip of the capillary. Lower panel: The far-field image shows the direct beam (center dot) passing through the capillary and the once-reflected beam forming the outer ring of intensity. The structure in the ring is due to slope-error imperfections arising from the pulling process.

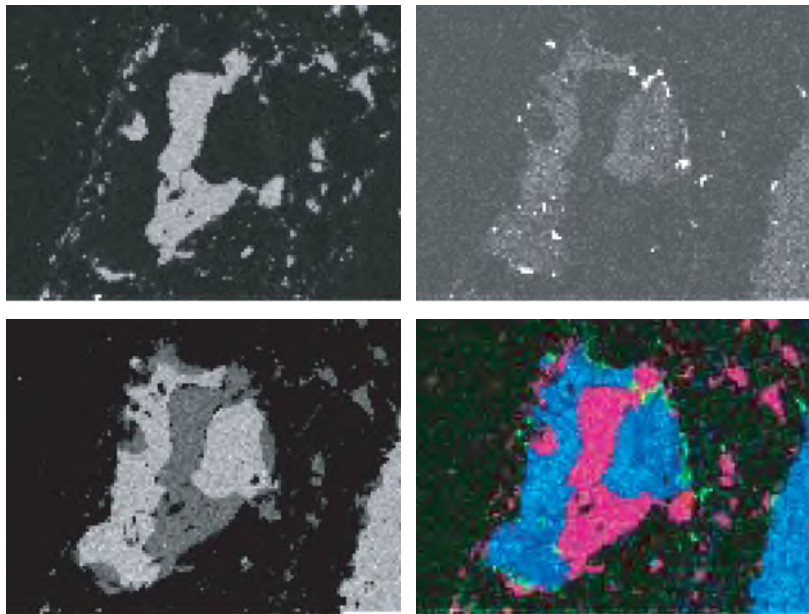


FIGURE 62.3 A 12-s x-ray spectrum image (128×96 pixels; 1 ms) obtained with a silicon drift detector.

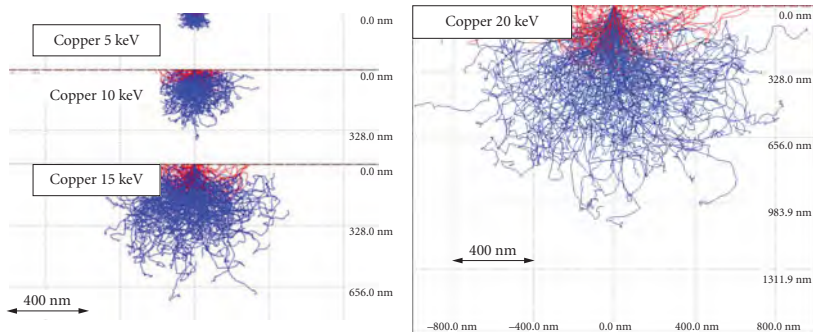


FIGURE 62.4 Monte Carlo simulations of electron trajectories in copper at various beam energies. Note the significant increase in the scattering volume with Increasing energy.

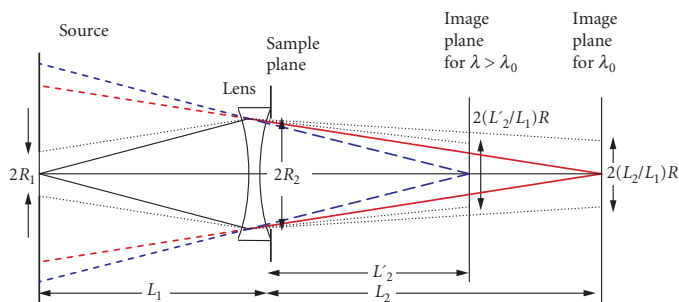


FIGURE 63.6 A schematic diagram of a focusing lens arrangement with source of radius R_1 at a distance L_1 from the sample with an aperture of radius R_2 . The focusing lens with a focal length f_0 is placed in front of the sample such that the source is imaged (continuous lines) with a radius $(L_2/L_1) R_1$ at a distance L_2 from the lens for a wavelength λ_0 . For another wavelength $\lambda (> \lambda_0)$ the source is imaged (long dashed lines) with a radius $(L'_2/L_1) R_1$ at a distance $L'_2 (< L_2)$ such that $1/L'_2 - 1/L_2 - 1/f_0[(\lambda/\lambda_0)^2 - 1]$.

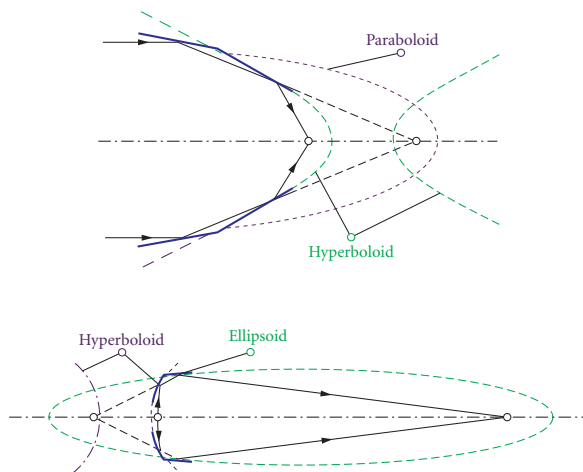


FIGURE 64.5 Commonly used Wolter-1 mirror configurations.