

SPIE
PRESS

Optical System Design

SECOND EDITION

**Robert E. Fischer
Biljana Tadic-Galeb
Paul R. Yoder**

OPTICAL SYSTEM DESIGN

About the Authors

Robert E. Fischer is the president and founder of OPTICS 1, Incorporated, of Westlake Village, California. He has a BS and MS in Optics from the Institute of Optics at the University of Rochester. He has been involved in optical design and engineering since 1967 when he joined the Itek Corporation. Prior to founding OPTICS 1, he was with Hughes Aircraft. His primary areas of technical interest and expertise include lens design, optical engineering, optical system manufacturing and testing, illumination systems, and related engineering technologies. In addition to chairing many conferences with the SPIE, Fischer has held several positions with the society, including president in 1984, and treasurer from 2001 through 2005. His popular short course "Practical Optical System Design" has been widely attended over the years and forms the basis for this book. Mr. Fischer is a Fellow of SPIE and the Optical Society of America. He was awarded the Albert M. Pezzuto award from SPIE in 1986, and the Gold Medal of SPIE in 2000.

Biljana Tadic-Galeb is a Senior Staff Optical Engineer at Panavision of Woodland Hills, California. She has a BS in Physics from the University of Sarajevo, Yugoslavia, and an MS in Optics from Reading University in England. She also holds an MS in Metrology from the University of Beograd, Yugoslavia. Ms. Tadic-Galeb has 30 years of experience as an Optical Systems Engineer with specialties in the development of complex visible, IR, and UV optical systems, projections systems, laser systems, and hybrid systems with diffractive elements and fibers. She was recognized by the National Academy of Engineering in their featured Women Engineers Program.

Paul R. Yoder, Jr. has, since 1951, designed and managed the development of optical systems and instruments for military, aerospace, and commercial applications at the US. Army's Frankford Arsenal and at the Perkin Elmer Corporation, as well as ophthalmic laser systems at Taunton Technologies, Inc. Since 1983, he has also served many clients as an independent consultant in optical engineering and optomechanical design. He has published widely in those fields with 66 technical papers, 2 handbook chapters, and 4 reference books to his credit. Yoder received his BS and MS degrees in physics from Juniata College (1947) and Penn State University (1950). Yoder is a Fellow of the SPIE, a Fellow of the OSA, a member of Sigma Xi and a co-founder of the SPIE's Optomechanical/Instrument Working Group. He has taught numerous short courses on optomechanical design and engineering for industry, government agencies, and SPIE in the USA and Europe as well as graduate-level courses for the University of Connecticut. In recognition of his contributions to optomechanical design and engineering, Yoder received the *Engineering Excellence Award* from the OSA in 1997 and the *George W. Goddard Award* from the SPIE in 1999 as well as the SPIE Director's Awards for outstanding service to the Society in 1996 and 2006.

Optical System Design

Robert E. Fischer
CEO, OPTICS 1, Incorporated

Biljana Tadic-Galeb
Panavision

Paul R. Yoder
Consultant

With contributions by
Ranko Galeb
Bernard C.Kress, Ph.D.
Stephen C. McClain, Ph.D.
Tom Baur
Rick Plympton
Bob Wiederhold
Alastair J. Grant

Second Edition



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

Copyright © 2008 by The McGraw-Hill Companies, Inc. All rights reserved. Manufactured in the United States of America. 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.

0-07-159358-6

The material in this eBook also appears in the print version of this title: 0-07-147248-7.

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. For more information, please contact George Hoare, Special Sales, at george_hoare@mcgraw-hill.com or (212) 904-4069.

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.

DOI: 10.1036/0071472487

For more information about this title, [click here](#)

CONTENTS

Preface	xiii	
Acknowledgments	xv	
Chapter 1.	Basic Optics and Optical System Specifications	1
	The Purpose of an Imaging Optical System	1
	How to Specify Your Optical System: Basic Parameters	4
	Basic Definition of Terms	11
	Useful First-Order Relationships	15
Chapter 2.	Stops and Pupils and Other Basic Principles	29
	The Role of the Aperture Stop	29
	Entrance and Exit Pupils	31
	Vignetting	32
Chapter 3.	Diffraction, Aberrations, and Image Quality	35
	What Image Quality Is All About	35
	What Are Geometrical Aberrations and Where Do They Come From?	36
	What Is Diffraction?	40
	Diffraction-Limited Performance	43
	Derivation of System Specifications	45
Chapter 4.	The Concept of Optical Path Difference	49
	Optical Path Difference (OPD) and the Rayleigh Criteria	49
	Peak-to-Valley and RMS Wavefront Error	52
	The Wave Aberration Polynomial	55
	Depth of Focus	56
Chapter 5.	Review of Specific Geometrical Aberrations and How to Get Rid of Them	59
	Spherical Aberration	60
	Coma	72
	Astigmatism	75

Contents

Field Curvature and the Role of Field Lenses	78
Distortion	85
Axial Color	89
Lateral Color	90
Parametric Analysis of Aberrations Introduced by Plane Parallel Plates	91
Chapter 6. Glass Selection (Including Plastics)	95
Material Properties Overview	95
The Glass Map and Partial Dispersion	96
Parametric Examples of Glass Selection	102
How to Select Glass	106
Plastic Optical Materials	109
A Visual Aid to Glass Selection	111
Chapter 7. Spherical and Aspheric Surfaces	115
Definition of an Aspheric Surface	115
Conic Surfaces	117
Application of Aspheric Surfaces in Reflective and Refractive Systems	119
Guidelines in the Use of Aspheric Surfaces	124
Specification of Aspheric Surfaces	126
Chapter 8. Design Forms	129
Introduction	129
System Configurations for Refractive Systems	131
System Configurations for Reflective Systems	138
Reflective Systems, Relative Merits	144
Refractive Systems, Relative Merits	146
Mirrors and Prisms	147
Design of Visual Systems	155
Chapter 9. The Optical Design Process	167
What Do We Do When We Optimize a Lens System?	168
How Does the Designer Approach the Optical Design Task?	171
Sample Lens Design Problem	176

Contents

Chapter 10.	Computer Performance Evaluation	179
	What Is Meant by Performance Evaluation	179
	What Is Resolution?	180
	Ray Trace Curves	181
	Spot Diagrams	187
	Optical Path Difference	189
	Encircled Energy	189
	MTF	191
Chapter 11.	Gaussian Beam Imagery	199
	Beam Waist and Beam Divergence	201
	Collimation of Laser Beams	203
	Propagation of Gaussian Beams and Focusing into a Small Spot	204
	Truncation of a Gaussian Beam	205
	Application of Gaussian Beam Optics in Laser Systems	208
	F- θ Lenses in Laser Scanners	211
Chapter 12.	Basics of Thermal Infrared Imaging in the 3- to 5- and 8- to 12- μm Spectral Bands (Plus UV Optics)	213
	The Basics of Thermal Infrared Imaging	213
	The Dewar, Cold Stop, and Cold Shield	217
	Cold Stop Efficiency	219
	Scanning Methods	222
	IR Materials	229
	Reduced Aberrations with IR Materials	236
	Image Anomalies	239
	Athermalization	246
	System Design Examples	250
	Optical Systems for the UV	255
Chapter 13.	Diffractive Optics	259
	Introduction	259
	The Many Faces of Diffractive Optics	262
	What Design and Modeling Tools Should I Use?	277
	How Are Diffractives Fabricated?	287
	Where Are Diffractives Used?	308
	References	318

Chapter 14.	Design of Illumination Systems	321
	Introduction	321
	Köhler and Abbe Illumination	322
	Optical Invariant and Etendue	324
	Other Types of Illumination Systems	329
Chapter 15.	Performance Evaluation and Optical Testing	333
	Testing with the Standard 1951 U.S. Air Force Target	333
	The Modulation Transfer Function	337
	Interferometry	340
	Other Tests	344
Chapter 16.	Tolerancing and Producibility	347
	Introduction	347
	What Are Testplates and Why Are They Important?	348
	How to Tolerance an Optical System	353
	How Image Degradations from Different Tolerances Are Summed	356
	Forms of Tolerances	359
	Adjusting Parameters	364
	Typical Tolerances for Various Cost Models	366
	Example of Tolerance Analysis	367
	Surface Irregularities	374
	How Does Correlation Relate to Performance?	376
	Effect to Spot Diameter	377
	Effect to MTF: The Optical Quality Factor	379
	Beam Diameter and Surface Irregularity	383
	The Final Results	384
Chapter 17.	Optomechanical Design	389
	Environmental Considerations	389
	Applicable Design Guidelines	393
	Environmental Testing Methods	393
	Mechanical Parameters and Properties	393
	Typical Mechanical Property Values for Selected Materials	394
	Structural Design	396
	Vibration, Self-Weight Deflection, and Fundamental Frequency	398
	Shock	400

Contents

Rigid Housing Configurations	400
Modular Construction	401
Support Structure Configurations	405
Establishing Axial and Lateral Preload Requirements	414
Spherical and Crowned Lens Rims	415
Interfaces for Other Optical Components	416
Individual Lens Mounting Techniques	419
Surface Contact Interface Shapes	426
Mounting Windows, Shells, and Domes	429
Stress Consequences of Axial Preload	434
Temperature Effects on Axial Preload	436
Radial Stresses and Their Variations with Temperature	439
Bending Effects in Rotationally Symmetric Optics	439
Multiple-Component Lens Assemblies	441
Incorporating Prisms into the Design	452
Mirror Mountings	459
Mechanical Athermalization Techniques	467
References	476
Chapter 18. Optical Manufacturing Considerations	479
Material	480
Manufacturing	485
Special Fabrication Considerations	492
Relative Manufacturing Cost	502
Sourcing Considerations	502
Conclusion	504
Chapter 19. Polarization Issues in Optical Design	507
Introduction	507
Introduction to Polarization	508
The Mathematical Description of Polarized Light	513
Some Polarization Phenomena	523
Polarization Control Nuts and Bolts	535
Polarization Analysis of an Optical System	555
Minimizing Polarization Problems in Optical Design	559
Polarization as a Tool in Optical System Design	560
Summary	565
Bibliography	567

Chapter 20.	Optical Thin Films	569
	Introduction	569
	Designing Optical Coatings	570
	Various Categories of Optical Coatings	571
	Optical Coating Process	578
	Coating Performance Versus Number of Layers	582
	Specifying Coating Requirements	583
	Relationship Between Production Cost, Tolerances, and Quality	584
	Bibliography	585
Chapter 21.	Hardware Design Issues	587
	Off-the-Shelf Optics	587
	How to Effectively Work with Off-the-Shelf Optics	589
	Working with Off-the-Shelf Singlets and Doublets	590
	Example of Lens Used at Conjugates Different from What It Was Designed	591
	Pupil Matching	594
	Development of a Lab Mockup Using Off-the-Shelf Optics	595
	Stray Light Control	595
	Optomechanical Design	600
Chapter 22.	Lens Design Optimization Case Studies	603
	Error Function Construction	603
	Achromatic Doublet Lens Design	605
	Double Gauss Lens Design	610
	Digital Camera Lens	632
	Binocular Design	642
	Parametric Design Study of Simple Lenses Using Advanced Manufacturing Methods	646
	Design Data for Double Gauss	655
Chapter 23.	Optical Sensor Systems Modeling and Analysis	659
	Introduction	659
	Image Formation	660
	Detector Arrays	663
	Optical System Noise Characteristics	669

Contents

Color Sensors	691
Electronic Correction	696
Camera Connectivity	697
Bibliography	701
Chapter 24. Stray Light and Optical Scattering	703
Introduction	703
Stray Light Scatter Sources	703
Types of Scatter	711
Modeling and Analysis Techniques	713
Veiling Glare	715
Cleanliness	716
Suppression Techniques	717
Bright Field and Dark Field	731
How to Avoid Unwanted Stray Light	736
Bibliography	737
Chapter 25. Bloopers and Blunders in Optics	739
Distortion in a 1:1 Imaging Lens	739
Zoom Periscope	740
Sign of Distortion	742
Lens Elements That Are Not Necessary	744
Pupil Problems	744
Not Enough Light	745
Athermalization Using Teflon	746
Athermalization Specifications	746
Bad Glass Choice	747
Elements in Backward	747
Insufficient Sampling of Fields of View or Aperture	748
Images Upside Down or Rotated	749
The Hubble Telescope Null Lens Problem	750
Wrong Glass Type in a Precision Lens System	755
Single Use Camera with a Diffractive Achromat	755
Wrong Image Handedness	756
Cemented Triplet as Part of an Imaging System	757
Total Internal Reflection in a Cube Beamsplitter	758
Diffractive Optics Issues	760
Case of the Miscoated Mangin	763
Telescopes and Polarization	765

Contents

Chapter 26.	Rule of Thumb and Hints	767
General Optical Design Topics	767	
Optomechanical Topics	770	
Diffractive Optics	772	
Glossary	775	
Bibliography	785	
Index	787	



PREFACE

The design of imaging optical systems is an engineering discipline which has been practiced and written about for many years. In many ways, optical design is both a science and an art, and for this reason it is a technology that can cause problems if it is not done properly. Furthermore, most books on the subject tend to be complex and difficult to follow and to understand. With this book, we hope to bring the understanding of our discipline to everyone.

We are all aware of cameras, binoculars, and other optical systems and instruments. In the past several years, the field of optics and photonics has seen a tremendous surge in both technology and in applications. This is fueled by a closer association with electronics in devices such as digital cameras, enhanced machine vision systems, MEMS and microoptical systems for telecommunications and other related applications, many of which have yet to be invented.

With this surge in the applications of optics, the educational process of training experienced optical designers and engineers becomes extremely important if not critical.

We realize that it is difficult to be an expert in everything. We also realize that in addition to optical design which is the core of the book, important topics including optical manufacturing, polarization, and optical coatings are important subjects that need to be covered in this book, and the first edition included these topics. With this new second edition, other critical technologies including optomechanical design, systems modeling and analysis, and stray light suppression are now included. Further, completely revised chapters on diffractive optics and polarization are also included.

We are honored to have contributed chapters written by experts in their fields: Paul Yoder on optomechanical design, Rick Plympton and Bob Wiederhold on optical manufacturing, Steve McClain and Tom Baur on polarization in optical systems, Ranko Galeb on thin films and optical coatings, Bernard Kress on diffractive optics, and Alastair J. Grant on systems modeling and analysis, as well as stray light suppression.

The ultimate goal of this book is to teach optical design and engineering in a fully unintimidating way using clear and easy to understand graphics and explanations. Many authors feel an obligation to

include complex mathematical derivations. We have taken a very different approach. We will make this book clear and easy to understand with the goal that you will learn the subject matter with a combination of complete graphics, easy to follow explanations, and just enough math to be useful, but not too much math to make the book hard to follow or difficult to understand.

This book *Optical System Design* is largely based on the firm foundation of the short course by the same title taught by Bob Fischer to over thousands of students over the past 20 years. The course has been honed, polished, and expanded over the years. It is available on CD ROM and videotape, and finally via this book. Typical comments have been:

"This course was just what I had hoped it would be. It condensed the vast optical world into the key elements necessary for a broad understanding of the subject and an excellent foundation for future study. Good job!"

"Excellent presentation! This is an invaluable course for those who are engaged in optical systems efforts and have a minimum training in optics."

*"A fast paced, well-prepared study, presented by a hands-on instructor."
"I learned what I came to learn, thanks."*

*"Excellent! Wonderful presentation and technique. Material was well covered."
"Excellent in explaining and answering questions. Very useful rules of thumb, great presentation. Thanks!"*

"Very professional and excellent presentation."

This book is for everyone from program managers to seasoned optical designers and engineers, mechanical engineers, electrical engineers, and others. You will find that it is like reading *Gulliver's Travels*. We all read *Gulliver's Travels* in elementary school, some of us again in high school, and some scholars wrote their Ph.D theses on the book. *Gulliver's Travels* can be read at multiple levels, just like this book.

ROBERT E. FISCHER

ACKNOWLEDGMENTS

Welcome to the second edition of *Optical System Design*. When first published in the year 2000 the goal was to create a book in the field of optical system design and engineering that was clear and easy to understand, and at the same time highly useful to the reader. Far too many books are filled with complex mathematics and other “mumbo jumbo” that often makes it difficult to find what you are really looking for. The overriding goal was thus to create a book that was both extremely useful as well as readable, and all indications are that this goal was met.

I recall talking to a professor at a leading university in optics just after the first edition was published who not only liked the book, but he told me that he was reading the book *backwards*, that is to say from the last chapter forward. “Why” I asked with a puzzled look? Because he really liked the “Bloopers and Blunders in Optics” chapter and that was the last chapter in the book!

I am pleased to welcome four new contributing authors for the second edition. Bernard Kress has a new and fully revised chapter on “Diffractive Optics,” and Steve McClain along with Tom Baur have a new and fully revised chapter on “Polarization in Optical Systems.” In addition to the above, I am honored to welcome to the book a new chapter “Optomechanical Design” by Paul Yoder. Optics and mechanics are both critical technologies that truly go hand-in-hand to make for a complete system. We must never underestimate the importance of mechanics to an optical system. Paul Yoder is one of those true icons in the field! Thanks Paul!

Also welcome to Alastair Grant who has contributed chapters on “Optical sensor systems Modeling and Analysis” and also “Stray Light and Optical scattering.”

Finally, a deep and sincere thanks is due to my wife Emilia who so graciously supports my many extracurricular activities in optics. Also thanks to Antonia Petrusse Surd, my administrative assistant, who played a major role in coordinating the many inputs from authors.

ROBERT E. FISCHER

This page intentionally left blank

OPTICAL SYSTEM DESIGN

This page intentionally left blank

CHAPTER

1

Basic Optics and Optical System Specifications

This chapter will discuss what a lens or mirror system does and how we specify an optical system. You will find that properly and completely specifying a lens system early in the design cycle is an imperative ingredient required to design a good system.

The Purpose of an Imaging Optical System

The purpose of virtually all image-forming optical systems is to resolve a specified minimum-sized object over a desired field of view. The *field of view* is expressed as the spatial or angular extent in object space, and the minimum-sized object is the smallest *resolution element* which is required to identify or otherwise understand the image. The word “spatial” as used here simply refers to the linear extent of the field of view in the plane of the object. The field of view can be expressed as an angle or alternatively as a lateral size at a specified distance. For example, the field of view might be expressed as $10^\circ \times 10^\circ$, or alternatively as 350×350 m at a distance of 2 km, both of which mean the same thing.

A good example of a resolution element is the dot pattern in a dot matrix printer. The capital letter E has three horizontal bars, and hence five vertical resolution elements are required to resolve the letter. Horizontally, we would require three resolution elements. Thus, the minimum number of resolution elements required to resolve capital letters is in the vicinity of five vertical by three horizontal. Figure 1.1 is an example of this. Note that the capital letter B and the number 8 cannot be distinguished in a 3×5 matrix, and the 5×7 matrix of dots will do just fine. This applies to telescopes, microscopes, infrared systems, camera lenses, and any other form of image-forming optics. The generally accepted guideline is that approximately three resolution elements or 1.5 line pairs over the object's spatial extent are required to acquire an object. Approximately eight resolution elements or four line pairs are required to recognize the object and 14 resolution elements or seven line pairs are required to identify the object.

There is an important rule of thumb which says that this smallest desired resolution element should be matched in size to the minimum detector element or pixel in a pixelated charged-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS)—type sensor. While not rigorous, this is an excellent guideline to follow for an optimum match between the optics and the sensor. This will become especially clear when we learn about the Nyquist Frequency in Chap. 22, where we show a digital camera design example. In addition, the aperture of the system and transmittance of the optics must be sufficient for the desired sensitivity of the sensor or detector. The detector can be the human eye, a CCD chip or film in your 35-mm camera. If we do not have enough photons to record the imagery, then what good is the imagery?

The preceding parameters relate to the optical system performance. In addition, the design *form* or *configuration* of the optical system must be capable of meeting this required level of performance. For example, most of us will agree that we simply cannot use a single magnifying

Figure 1.1
Illustration of Number of Resolution Elements Required to Resolve or Distinguish Alphanumerics



Basic Optics and Optical System Specifications

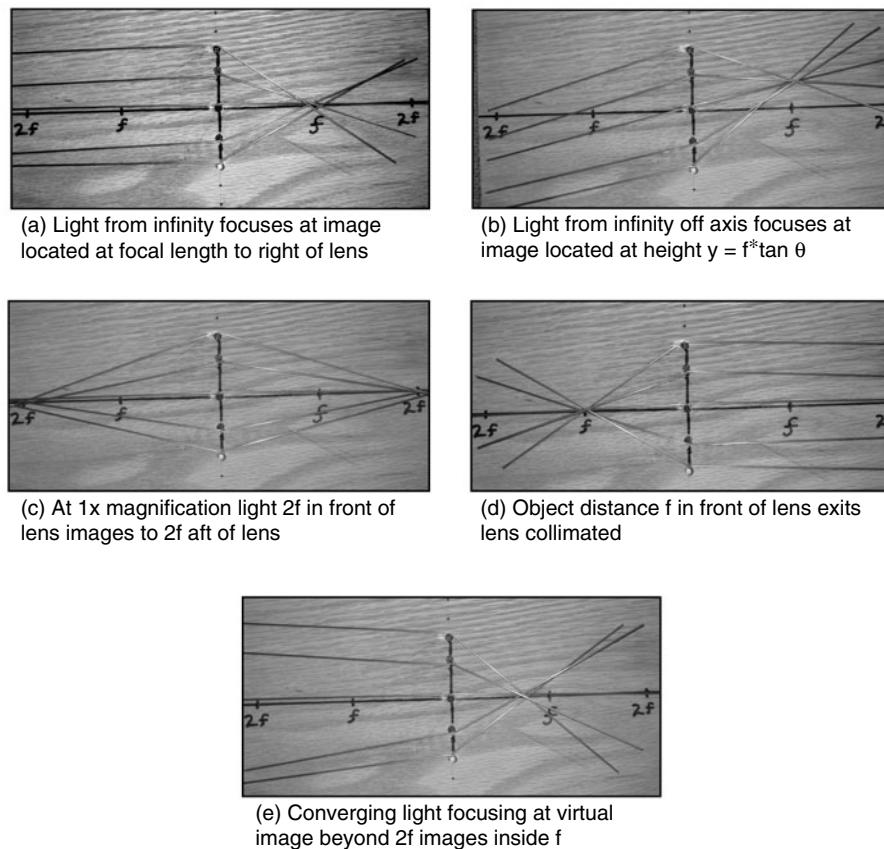
glass element to perform optical microlithography where submicron line-width imagery is required, or even lenses designed for 35-mm or digital photography for that matter. The form or configuration of the system includes the number of lens or mirror elements along with their relative position and shape within the system. We discuss design configurations in Chap. 8 in detail.

Furthermore, we often encounter special requirements, such as cold stop efficiency, in infrared systems, scanning systems, and others. These will be addressed later in this book.

Finally, the system design must be producible, meet defined packaging and environmental requirements, weight and cost guidelines, and satisfy other system specifications.

Figure 1.2 shows what we like to call a “wire lens” and what is shown are photographs of a real construction of a wire lens. If you are new to

Figure 1.2
A Simple “Wire Model” of a Lens Shows how Lenses Work, Where Wires Pivot about Vertical Line Representing Lens



optics, this simple demonstration is an excellent way to learn quickly just how a lens works. Consider (a) where we will show how the wire lens is constructed and how it works. The horizontal centerline that connects the two foci (f) represents the optical axis, and the vertical line represents our lens. If there are no aberrations, light from infinity will have parallel or collimated rays incident onto the lens from the left, and after passing through the lens each of these rays (wires) refract (bend) so as to all pass through the focal point at (f). Each of the five wires is free to pivot about the five black dots on the lens. It is clear that the rays (wires) further from the optical axis bend at a greater angle so that each ray passes through the focus. This is our basic paraxial lens.

Now in (b) we gather the wires at the focus and keeping them together, move the focal point where the wires converge vertically so that they are displaced in the image plane a vertical distance we will call " y ." Note that the rays entering the lens will be parallel or collimated as expected and all be going up to the right by angle θ . This yields the paraxial equation $y = f * \tan \theta$.

Let us now gather the wires together left of the lens and move what is now the object to a position $2f$ in front of the lens. After refraction by the lens the wires will converge at $2f$ behind the lens. It is a well-known property of a lens that an object $2f$ in front of a lens will image $2f$ behind the lens.

In (d) we bring our wires to all pass through (f) in front of the lens, and as expected the exiting wires are parallel, the reverse of (a).

Finally we show in (e) how if the image is inside (f), the wires will be converging entering the lens.

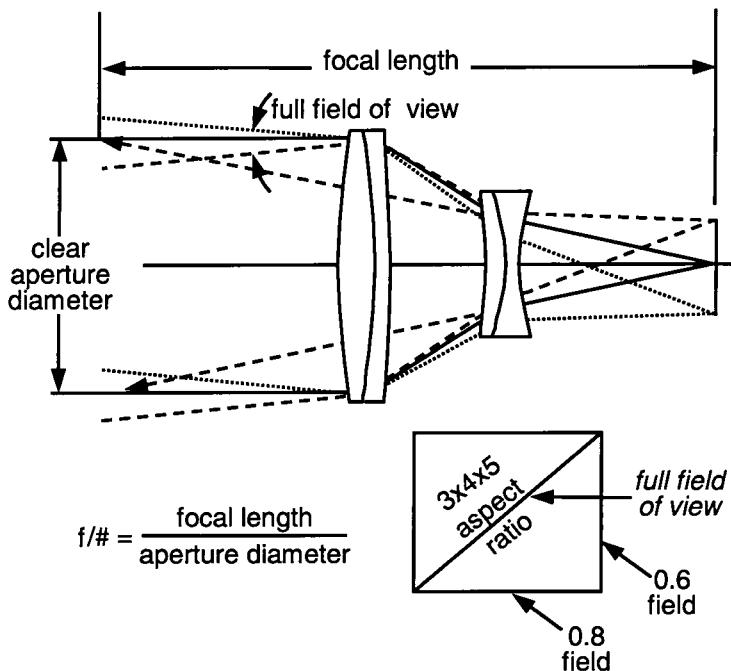
The *wire lens* is very simple to construct, and it is both educational and fun to work with. All you need is piano wire, some small screws, a felt tip Sharpie pen, a piece of wood, and a little bit of time.

How to Specify Your Optical System: Basic Parameters

Consider the lens shown in Fig. 1.3 where light from infinity enters the lens over its *clear aperture diameter*. If we follow the solid ray, we see that it is redirected by each of the lens element groups and components until it comes to focus at the image. If we now extend this ray *backward* from the image toward the front of the system as if it were not bent or

Basic Optics and Optical System Specifications

Figure 1.3
Typical Specifications



refracted by the lens groups, it intersects the entering ray at a distance from the image called the *focal length*. The final imaging cone reaching the image at its center is defined by its *f/number* or *f/#*, where

$$f/\text{number} = \frac{\text{focal length}}{\text{clear aperture diameter}}$$

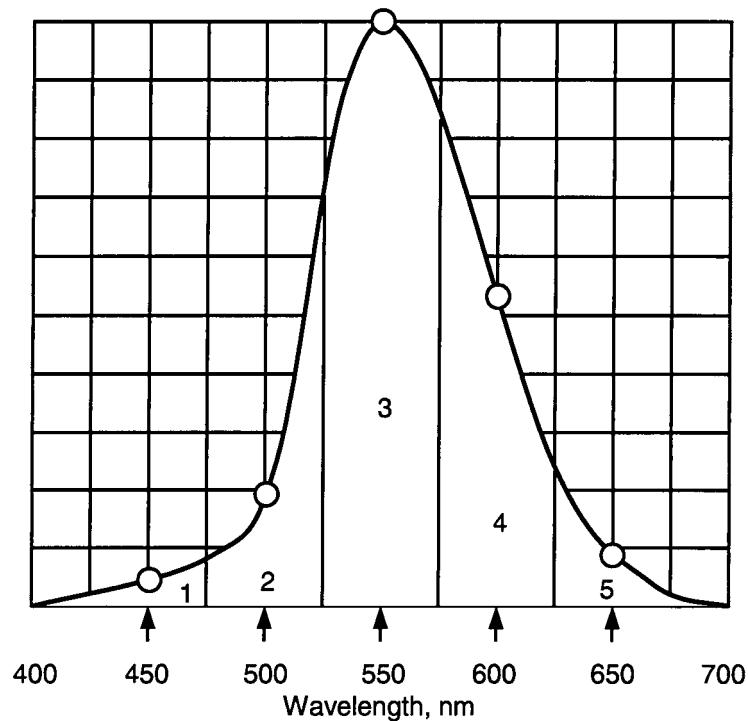
You may come across two other similar terms, *effective focal length* and *equivalent focal length*, both of which are often abbreviated EFL. The effective focal length is simply the focal length of a lens or a group of lenses. Equivalent focal length is very much the same; it is the overall focal length of a group of lens elements, some or all of which may be separated from one another.

The lens is used over a *full field of view*, which is expressed as an angle, or alternatively as a linear distance on the object plane. It is important to express the total or full field of view rather than a subset of the field of view. This is an extremely critical point to remember. For example, assume we have a CCD camera lens covering a sensor with a $3 \times 4 \times 5$ aspect ratio. We could specify the *horizontal field of view*, which is often

done in video technology and cinematography. However, if we do this, we would be ignoring the full diagonal of the field of view. If you do specify a field of view less than the full or total field, you absolutely must indicate this. For example, it is quite appropriate to specify the field of view as $\pm 10^\circ$. This means, of course, that the total or full diagonal field of view is 20° . Above all, do not simply say "field of view 10° " as the designer will be forced to guess what you really mean!

System specifications should include a defined *spectral range* or wavelength band over which the system will be used. A visible system, for example, generally covers the spectral range from approximately 450 nm to 650 nm. It is important to specify from three to five specific wavelengths and their corresponding relative weights or importance factors for each wavelength. If your sensor has little sensitivity, say, in the blue, then the image quality or performance of the optics can be more degraded in the blue without perceptible performance degradation. In effect, the spectral weights represent an importance factor across the wavelength band where the sensor is responsive. If we have a net spectral sensitivity curve, as in Fig. 1.4, we first select five representative

Figure 1.4
Example of Spectral Sensitivity Curve



Basic Optics and Optical System Specifications

wavelengths distributed over the band, $\lambda_1 = 450$ nm through $\lambda_5 = 650$ nm, as shown. The circular data points represent the relative sensitivity at the specific wavelengths, and the relative weights are now the normalized area or integral within each band from band 1 through band 5, respectively. Note that the weights are not the ordinate of the curve at each wavelength as you might first expect but rather the integral within each band. Table 1.1 shows the data for this example.

Even if your spectral band is narrow, you must work with its bandwidth and derive the relative weightings. You may find some cases where you think the spectral characteristics suggest a monochromatic situation but in reality, there is a finite bandwidth. Pressure-broadened spectral lines emitted by high-pressure arc lamps exhibit this characteristic. Designing such a system monochromatically could produce a disastrous result. In most cases, laser-based systems only need to be designed at the specific laser wavelength.

System *packaging constraints* are important to set at the outset of a design effort, if at all possible. These include length, diameter, weight, distance or clearance from the last surface to the image, location and space for fold mirrors, filters, and/or other components critical to the system operation.

Sets of specifications often neglected until it is too late are the *environmental parameters* such as *thermal soak* conditions (temperature range) that the system will encounter. Also, we may have *radial thermal gradients*, which are changes in temperature from the optical axis outward; *diametral thermal gradients*, which are thermal gradients across the diameter of the system in a nonaxially symmetrical profile; and *axial gradients*, which are thermal gradients from the front to the rear of the system. You may also be provided with a set of *operational* specifications and a set of *storage* specifications with respect to temperature.

TABLE 1.1

Example of Spectral Sensitivity and Relative Wavelength Weights

	Wavelength, nm	Relative sensitivity	Relative weight
	450	0.05	0.08
	500	0.2	0.33
	550	1.0	1.0
	600	0.53	0.55
	650	0.09	0.16

System *transmittance*, or throughput, as well as *relative illumination*, or brightness, uniformity over the image format are also often specified.

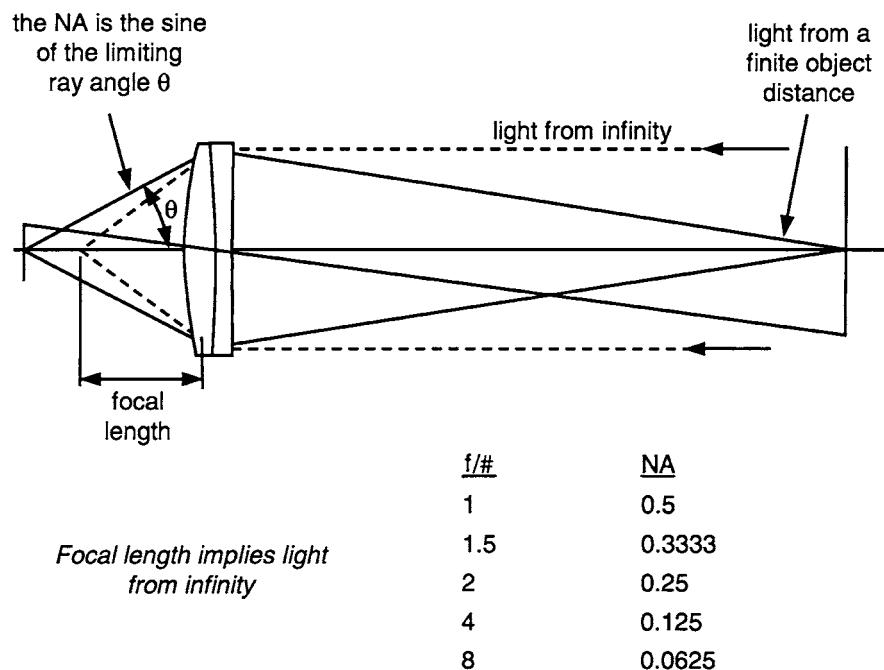
One of the most important specifications is the optical performance or *image quality*. The following list contains some of the more common ways of specifying the image quality, along with simple definitions. Each of these will be discussed in more detail in Chap. 10.

- *Modulation transfer function (MTF)*. The modulation (think of the word contrast) versus the number of line pairs per millimeter in the image
- *RMS blur diameter*. The diameter of a circle containing approximately 68% of the energy imaged from a point source
- *Encircled energy (or ensquared energy)*. The diameter of a circle (or side of a square such as a pixel) containing a given percent of energy, for example, 80%
- *Root-mean-square (rms) wavefront error*. The rms departure of the real wavefront from a perfect wavefront
- *Other*. Depending on the functional requirements of the system, there may be other performance requirements relating to image quality, for example, point spread function (PSF), control of specific aberrations, etc.

The most fundamental set of first-order specifications are *focal length*, *clear aperture diameter* (more properly called *entrance pupil diameter*, as will be explained in Chap. 2), and *f/number*. As we know, the *f/number* is the focal length divided by the clear aperture diameter (the entrance pupil diameter). There is, however, another important and related quantity called the *numerical aperture* (NA), which is often used. The numerical aperture is simply the sine of the half cone angle of the limiting edge ray coming to the axial image, or the sine of the half cone angle coming from the axial object point, as shown in Fig. 1.5. Why would we want or need yet another term to remember? The reason is that the definition of focal length is based on light from infinity entering the system. What if we have a so-called finite conjugate system where neither the object nor the image is at infinity? The traditional definition of focal length and *f/#* would be misleading since the system really is not being used with collimated light input. Numerical aperture is the answer. The *numerical aperture* is simply the sine of the image cone half angle, regardless of where the object is located. We can also talk about the numerical aperture at the object, which is the sine of the half cone angle from

Basic Optics and Optical System Specifications

Figure 1.5
Numerical Aperture
and $f/\#$



the optical axis to the limiting marginal ray emanating from the center of the object. Microscope objectives are routinely specified in terms of numerical aperture. Some microscope objectives reimagine the object at a finite distance, and some have collimated light exiting the objective. These latter objectives are called *infinity corrected* objectives, and they require a “tube lens” to focus the image into the focal plane of the eyepiece or alternatively onto the CCD or other sensor.

As noted earlier, the definition of focal length implies light from infinity. And similarly, f/number is focal length divided by the clear aperture diameter. Thus, f/number is also based on light from infinity. Two terms commonly encountered in finite conjugate systems are “ f/number at used conjugate” and “working f/number .” These terms define the equivalent f/number , even though the object is not at infinity. The f/number at used conjugate is $1/(2 \cdot \text{NA})$, and this is valid whether the object is at infinity or at a finite distance.

It is important at the outset of a design project to compile a specification for the desired system and its performance. The following is a candidate list of specifications:

Optical system basic operational and performance specifications and requirements

Basic system parameters:	
Object distance	
Image distance	
Object to image total track	
Focal length	
f/number (or numerical aperture)	
Entrance pupil diameter	
Wavelength band	
Wavelengths and weights for 3 or 5 λ s	
Full field of view	
Magnification (if finite conjugate)	
Zoom ratio (if zoom system)	
Image surface size and shape	
Detector type	
Optical performance:	
Transmission	
Relative illumination (vignetting)	
Encircled energy	
MTF as a function of line pairs/mm	
Distortion	
Field curvature	
Lens system:	
Number of elements	
Glass versus plastic	
Aspheric surfaces	
Diffractive surfaces	
Coatings	
Sensor:	
Sensor type	
Full diagonal	
Number of pixels (horizontal)	
Number of pixels (vertical)	
Pixel pitch (horizontal)	
Pixel pitch (vertical)	
Nyquist frequency at sensor, line pairs/mm	
Packaging:	
Object to image total track	
Entrance and exit pupil location and size	
Back focal distance	
Maximum diameter	
Maximum length	
Weight	
Environmental:	
Thermal soak range to perform over	
Thermal soak range to survive over	
Vibration	
Shock	
Other (condensation, humidity, sealing, etc.)	

Basic Optics and Optical System Specifications

Optical system basic operational and performance specifications and requirements
(Continued)

Illumination:	
Source type	_____
Power, in watts	_____
Radiometry issues, source:	
Relative illumination	_____
Illumination method	_____
Veiling glare and ghost images	_____
Radiometry issues, imaging:	
Transmission	_____
Relative illumination	_____
Stray light attenuation	_____
Schedule and cost:	
Number of systems required	_____
Initial delivery date	_____
Target cost goal	_____

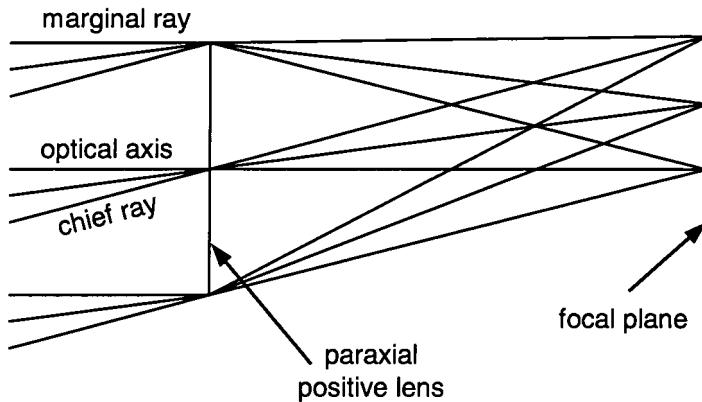
Basic Definition of Terms

There is a term called *first-order optics*. In first-order optics the bending or refraction of a lens or lens group happens at a specific plane rather than at each lens surface. In first-order optics, there are no aberrations of any kind and the imagery is perfect, by definition.

Let us first look at the simple case of a perfect thin positive lens often called a *paraxial lens*. The limiting aperture that blocks the rays beyond the lens clear aperture is called the *aperture stop*. The rays coming from an infinitely distant object that passes through the lens clear aperture focus in the image plane. A paraxial positive lens is shown in Fig. 1.6. The rays coming from an infinitely distant point on the optical axis approach the lens as the bundle parallel to the optical axis. The ray that goes along the optical axis passes through the lens without bending. However, as we move away from the axis, rays are bent more and more as we approach the edge of the clear aperture. The ray that goes through the edge of the aperture parallel to the optical axis is called the *marginal ray*. All of the rays parallel to the optical axis focus at a point on the optical axis in the focal plane. The rays that are coming from a nonaxial object point form an angle with the optical axis. One of these rays is called a *chief ray* and it goes through the center of the lens (center of the *aperture stop*) without bending.

Figure 1.7 shows a *paraxial lens* (a), and a real lens (b). Both lenses have the same focal length and *f*/number; however the paraxial lens has zero spherical aberration while the real lens has significant spherical aberration.

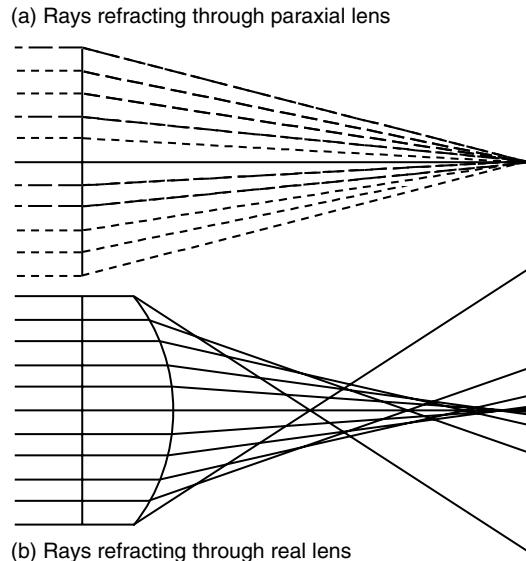
Figure 1.6
Paraxial Positive Lens



A common first-order representation of an optical system is shown in Fig. 1.8. What we have here is the representation of any optical system, yes, any optical system! It can be a telescope, a microscope, a submarine periscope, or any other imaging optical system.

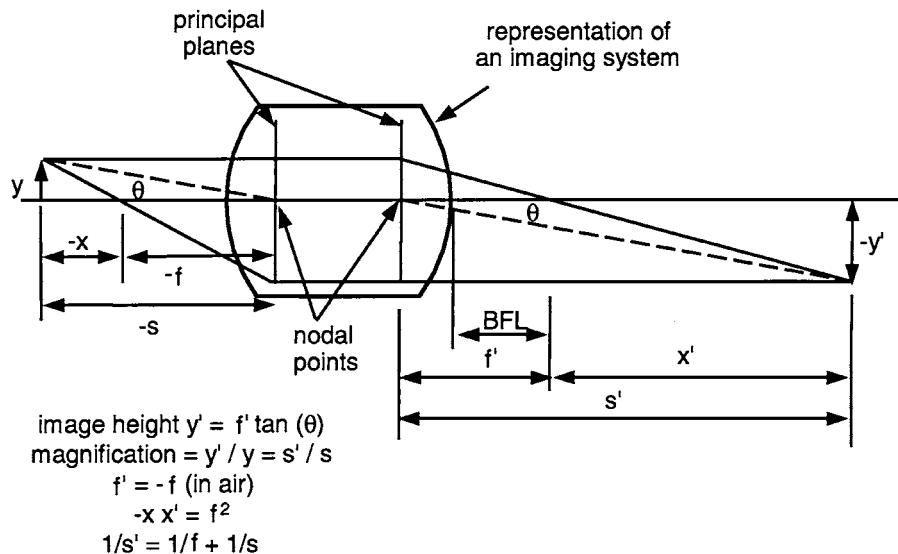
The easiest way to imagine what we have here is to think of having a shoebox with a 2-in-diameter hole in each end and inside is some arbitrary optical system (or perhaps nothing at all!). If we send a laser beam into the shoebox through the center of the left-hand hole normal to the hole, it will

Figure 1.7
Paraxial Lens (a) and
Real Lens (b)



Basic Optics and Optical System Specifications

Figure 1.8
Cardinal Points of an Optical System

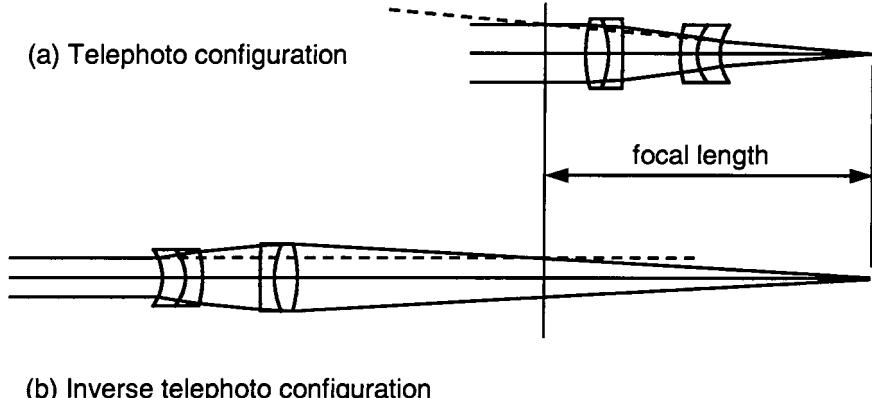


likely exit through the center of the hole at the other end of the shoebox. The line going through the center of each of the holes is the optical axis.

If we now send the laser beam into the shoebox displaced nearly 1 in vertically, it may exit the shoebox on the other end exactly the same and parallel to how it entered, in which case there is probably nothing in the shoebox. Alternately, the laser beam may exit the shoebox either descending or ascending (going downhill or uphill). If the laser beam is descending, it will cross the optical axis somewhere to the right of the shoebox, as shown in Fig. 1.8. If we connect the entering laser beam with the exiting laser beam, they will intersect at a location called the *second principal plane*. This is sometimes called the *equivalent refracting surface* because this is the location where all of the rays appear to bend about. In a high-performance lens, this equivalent refracting surface is spherical in shape and is centered at the image. The distance from the second principal plane to the plane where the ray intersects the optical axis is the *focal length*.

If we now send a laser beam into the hole on the right parallel to the optical axis and in a direction from right to left, it will exit either ascending or descending (as previously), and we can once again locate the principal plane, this time the *first principal plane*, and determine the focal length. Interestingly, the focal length of a lens system used in air is identical whether light enters from the left or the right. Figure 1.9a shows a telephoto lens whose focal length is labeled. Recall that we can compute

Figure 1.9
The Identical Lens
Showing How the
Focal Length Is Iden-
tical When the Lens Is
Reversed



the focal length by extending the marginal ray back from the image until it intersects the incoming ray, and this distance is the focal length. In the telephoto lens the focal length is longer than the physical length of the lens, as shown. Now consider Fig. 1.9b, where we have taken the telephoto lens and simply reversed it with no changes to radii or other lens parameters. Once again, the intersection of the incoming marginal ray with the ray extending forward from the image is the focal length. The construction in Fig. 1.9b shows clearly that the focal lengths are identical with the lens in either orientation!

The center of the principal planes (where the principal planes cross the optical axis) are called the *nodal points*, and for a system used in air, these points lie on the principal planes. These nodal points have the unique property that light directed at the front nodal point will exit the lens from the second nodal point at exactly the same angle with respect to the optical axis. This, too, we can demonstrate with our laser beam and shoebox.

So far, we have not talked about an object or an image at all. We can describe or represent a cone of light leaving an object (at the height, y , in Fig. 1.8.) as including the ray parallel to the optical axis, the ray aimed at the front nodal point, and lastly the ray leaving the object and passing through the focal point on the left side of the lens. All three of these rays (or laser beams) will come together once again to the right of the lens a distance, y' , from the optical axis, as shown. We will not bore you with the derivation, but rest assured that it does happen this way.

What is interesting about this little example is that our shoebox could contain virtually any kind of optical system, and all of the preceding will hold true. In the case where the laser beam entering parallel to the

Basic Optics and Optical System Specifications

optical axis exits perhaps at a different distance from the axis but parallel to the axis, we then have what is called an *afocal lens* such as a laser beam expander, an astronomical telescope, or perhaps a binocular. An afocal lens has an infinite focal length, meaning that both the object and the image are at infinity.

Useful First-Order Relationships

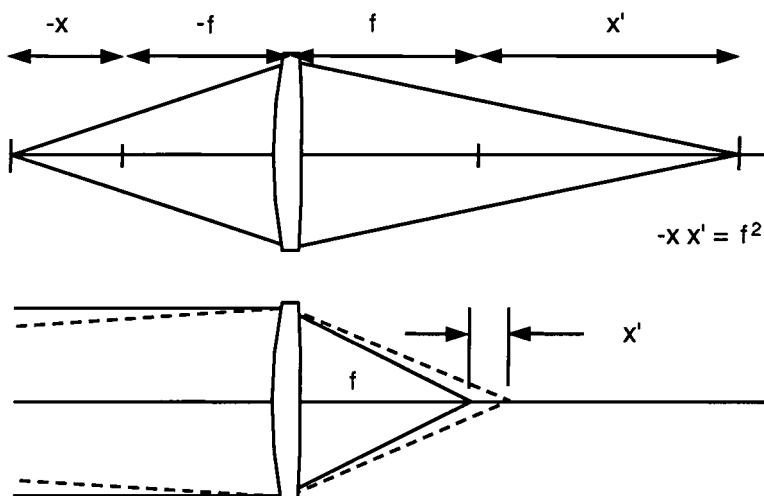
As discussed earlier, in first-order optics, lenses can be represented by planes where all of the bending or refraction takes place. Aberrations are nonexistent in first-order optics, and the imagery is by definition absolutely perfect. There are a series of first-order relationships or equations, which come in very handy in one's everyday work, and we will discuss the most useful ones here.

Consider the simple lens system shown in Fig. 1.10. Newton's equation says:

$$(-x)(x') = f^2$$

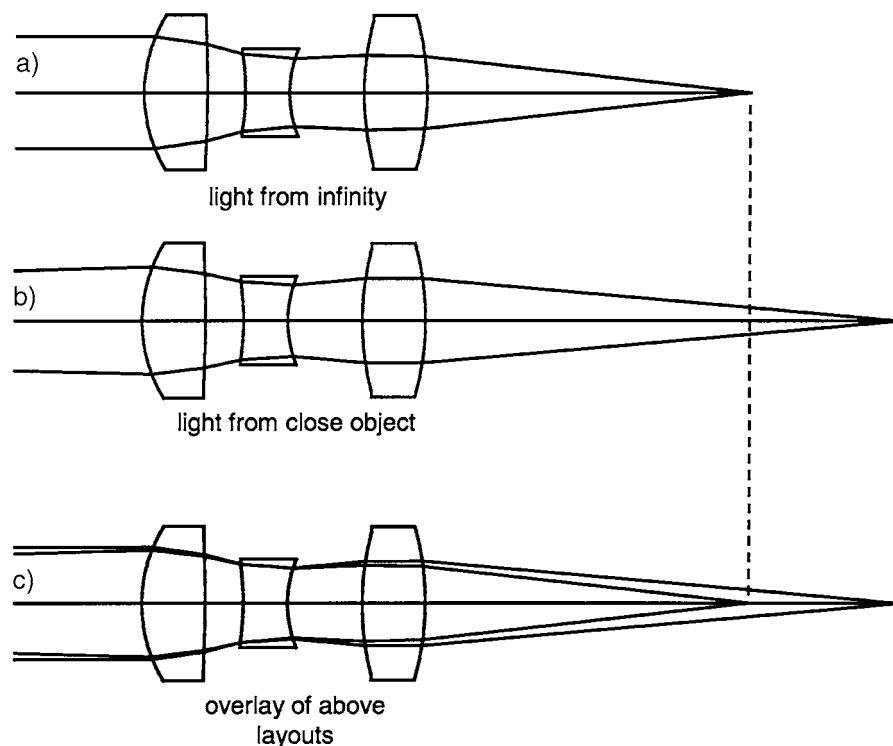
where x is the distance from the focal point on the front side of the lens to the object, and x' is the distance from the rear focal point to the image. Note that x is negative according to the sign convention, since the distance from the image to the object is in a direction to the

Figure 1.10
Newton's Equation



left. This is an interesting equation in that, at first glance, it seems to be of marginal use. However, consider the example where we need to determine how far to refocus a 50-mm focal length lens for an object at a distance of 25 m. The result is 0.1 mm, and this is, in all likelihood, a very reliable and accurate answer. We must always remember, however, that first-order optics is an approximation and assumes no aberrations whatsoever. For small angles and large $f/\#$ s the results are generally reliable; however, as the angles of incidence on surfaces increase, the results become less reliable. Consider Fig. 1.11a where we show how light proceeds through a three-element lens known as a Cooke triplet, with the object at infinity. If we were to use Newton's equation to determine how far to refocus the lens for a relatively close object distance, as shown in Fig. 1.11b, the resulting amount of refocusing may not be reliable. This is because the ray heights and angles of incidence are different from the infinite object condition, especially at the outer positive elements, as shown in Fig. 1.11c, which is an overlay of the infinite and close object distance layouts. These different ray heights and angles of

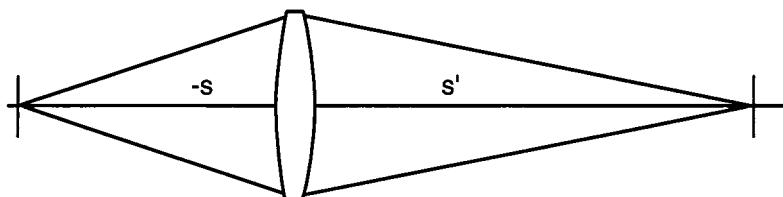
Figure 1.11
Light Imaging
through a Cooke
Triplet for Two Object
Distances



Basic Optics and Optical System Specifications

Figure 1.12

Basic Relationship of Object and Image:
The "Lens Makers" Equation



incidence will cause aberrations, and the net effect is that the result determined by Newton's equation might not be reliable for predicting where best focus is located. Consider a typical $f/5$ 50-mm focal length Cooke triplet lens used at an object distance of 0.5 m. Newton's equation gives a required refocusing of 2.59 mm from infinity focus, versus 3.02 mm based on optimum image quality, a difference of 0.43 mm. However, for a 10-m object distance, the difference between Newton's equation and best focus reduces to 0.0008 mm, which is negligible.

The important message here is to use first-order optics with caution. If you have any question as to its accuracy in your situation, you really should perform a computer analysis of what you are modeling. If you then find that your first-order analysis is sufficiently accurate, continue to use it with confidence in similar situations. However, if you find inaccuracies, you may need to work with real rays in your computer model.

Another useful and commonly used equation is

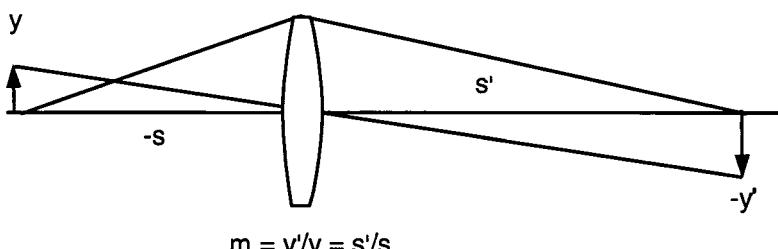
$$\frac{1}{s'} = \frac{1}{f} + \frac{1}{s}$$

where s and s' are the object and image distances, respectively, as shown in Fig. 1.12.

Consider now the basic definitions of magnification from an object to an image. In Fig. 1.13, we show how *lateral magnification* is defined. Lateral implies in the plane of the object or the image, and lateral magnification

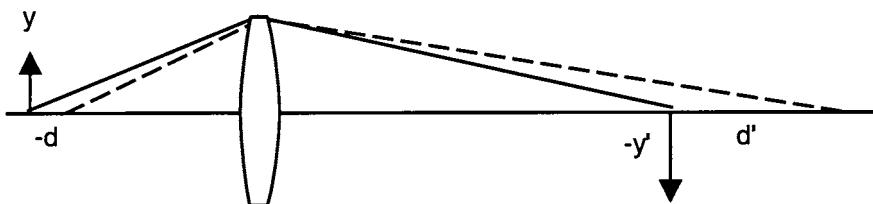
Figure 1.13

Lateral Magnification



$$m = y'/y = s'/s$$

Figure 1.14
Longitudinal
Magnification



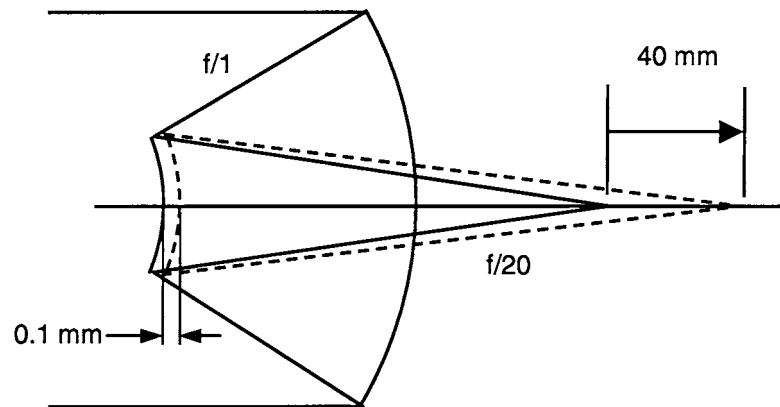
$$\text{longitudinal magnification} = d'/d = (\text{lateral magnification})^2 = (y'/y)^2$$

is therefore the image height, y' divided by the object height, y . It is also the image distance, d' , divided by the object distance, d .

There is another form of magnification: the *longitudinal magnification*. This is the magnification *along the optical axis*. This may be a difficult concept to visualize because the image is always in a given plane. Think of longitudinal magnification this way: if we move the object a distance, d , we need to move the image, d' , where d'/d is the longitudinal magnification. It can be shown that the longitudinal magnification is the square of the lateral magnification, as shown in Fig. 1.14. Thus, if the lateral magnification is $10\times$, the longitudinal magnification is $100\times$. A good example is in the use of an overhead projector where the viewgraph is in the order of 250 mm wide and the screen is in the order of 1 m wide, giving a lateral magnification of $4\times$. If we were to move the viewgraph 25 mm toward the lens, we would need to move the screen outward by $16 \times 25 = 400$ mm.

As a further example of the concept, consider Fig. 1.15 where we show a two-mirror reflective system called a *Cassegrain*. Let us assume

Figure 1.15
Cassegrain Reflective
System



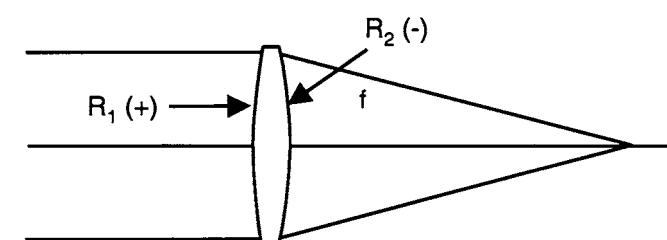
Basic Optics and Optical System Specifications

that the large or *primary mirror* is 250 mm in diameter and is $f/1$. Also, assume that the final image is $f/20$. The small, or *secondary*, mirror is, in effect, magnifying the image, which would be formed by the primary mirror by $20\times$ in lateral magnification. Thus, the longitudinal magnification is $400\times$, which is the square of the lateral magnification. Now let us move the secondary mirror 0.1 mm toward the primary mirror. How far does the final image move? The answer is $0.1 \times 400 = 40$ mm to the right. This is a very large amount and it illustrates just how potent the longitudinal magnification really can be.

While we are on the subject, how can we easily determine which way the image moves if we move the secondary mirror to the right as discussed previously? Indeed there is an easy way to answer this question (and similar questions). The approach to follow when presented by a question of this kind is to consider moving the component a very large amount, perhaps even to its limit, and ask "what happens?" For example, if we move the secondary mirror to a position approaching the image formed by the primary, clearly the final image will coincide with the secondary mirror surface when it reaches the image formed by the primary. This means that the final image will move in the same direction as the secondary mirror motion. In addition, if you take the secondary and move it a large amount toward the primary eventually the light will reflect back to the primary when the rays are incident normal to the secondary mirror surface. Moreover, at some intermediate position, the light will exit to the right collimated or parallel. The secret here, and for many other similar questions, is to *take the change to the limit*. Take it to a large enough magnitude so that the direction of the result becomes fully clear and unambiguous.

Figure 1.16 shows how the *optical power* of a single lens element is defined. The optical power is given by the Greek letter, Φ , and Φ is the

Figure 1.16
Optical Power and
Focal Length of a
Single Lens Element



$$\Phi = \text{Optical power} = \frac{1}{f} = (n - 1) \cdot \left(\frac{1}{r_1} - \frac{1}{r_2} \right)$$

reciprocal focal length or 1 divided by the focal length. In optics we use a lot of reciprocal relationships. Power = $\Phi = 1/\text{focal length}$, and curvature = (1/radius) is another.

If we know the radii of the two surfaces, r_1 and r_2 , and the refractive index, n , we find that

$$\Phi = \frac{1}{\text{focal length}} = (n - 1) \left(\frac{1}{r_1} - \frac{1}{r_2} \right)$$

In addition, if we have two thin lenses separated by an air space of thickness d , we find that

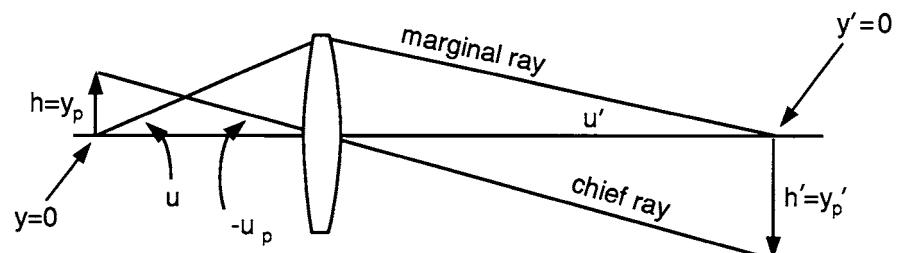
$$\Phi = \frac{1}{\text{focal length}} = \Phi_a + \Phi_b - d(\Phi_a \Phi_b)$$

One very important constant in the optical system is the *optical invariant* or *Lagrange invariant* or *Helmholtz invariant*. It has a constant value throughout the entire system, on all surfaces and in the spaces between them. The optical invariant defines the system throughput. The basic characteristic of an optical system is known when the two main rays are traced through the system: the *marginal ray* going from the center of the object through the edge of the aperture stop and the *chief* or *principal ray* going from the edge of the object through the center of the aperture stop. These rays are shown in Fig. 1.17. The optical invariant defines the relationship between the angles and heights of these two rays through the system, and in any location in the optical system it is given as

$$I = y_p n u - y n u_p$$

where the subscript p refers to the principal ray, no subscript refers to the marginal ray, and n is the refractive index.

Figure 1.17
The Optical Invariant



Basic Optics and Optical System Specifications

The optical invariant, I , once computed for a given system, remains constant everywhere within the system. When this formula is used to calculate the optical invariant in the object plane and in the image plane where the marginal ray height is zero, then we get the commonly used form of the optical invariant

$$I = hnu = h'n'u'$$

where h , n , and u are the height of the object, the index of refraction, and angle of the marginal ray in the object plane, and h' , n' , and u' are the corresponding values in the image space. Although this relationship is strictly valid only in the paraxial approximation, it is often used with sufficient accuracy in the form

$$nh \sin u = n'h' \sin u'$$

From this form of optical invariant we can derive the magnification of the system $M = h'/h$ as

$$M = \frac{n \sin u}{n' \sin u'}$$

In simple terms these relationships tell us that if the optical system magnifies or increases the object M times, the viewing angle will be decreased M times.

In systems analysis, the specification of the optical invariant has a significant importance. In the radiometry and photometry of an optical system, the total light flux collected from a uniformly radiating object is proportional to I^2 of the system, commonly known as *etendue*, where I is the optical invariant. For example, if the optical system is some kind of a projection system that uses a light source, then the projection system with its optical invariant defines the light throughput. It is useful to compare the optical invariant of the light source with the invariant of the system to see how much light can be coupled into the system. It is not necessarily true that the choice of a higher-power light source results in a brighter image. It can happen that the light-source optical invariant is significantly larger than the system optical invariant, and a lot of light is stopped by the system. The implications of the optical invariant and etendue on radiometry and photometry will be discussed in more depth in Chap. 14.

The *magnification* of a visual optical system is generally defined as the ratio of the angles subtended by the object *with* or looking through

the optical system to the angle subtended by the object *without* the optical system or looking at the object directly with unaided vision. In visual optical systems where the human eye is the detector, a nominal viewing distance without the optical system when the magnification is defined as unity is 250 mm. The reason that unity magnification is defined at a distance of 250 mm is that this is the closest distance that most people with good vision can focus. As you get closer to the object, it subtends a larger angle and hence looks bigger or magnified.

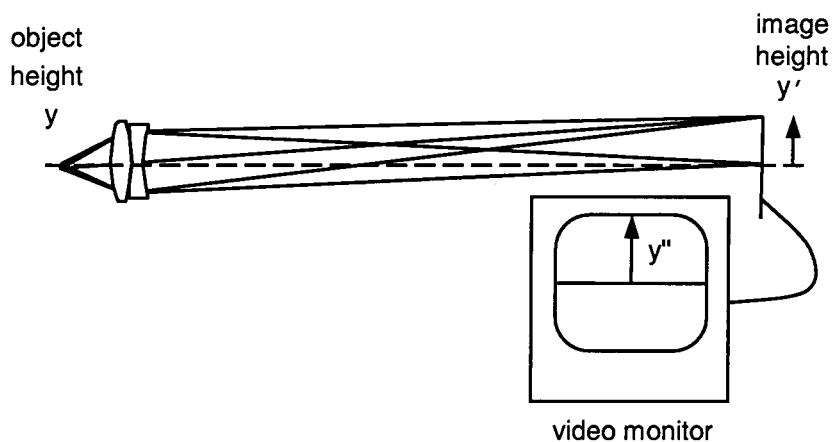
This general definition of magnification takes different forms for different types of optical systems. Let us look first at the case of a microscope objective with a CCD camera, as shown in Fig. 1.18. The image from the CCD is brought to the monitor and the observer is located at the distance, D , from the monitor. The question is what is the magnification of this system. In the first step, a microscope objective images the object with the height, y , onto the CCD camera, with the magnification

$$\frac{y'}{y}$$

where y' is the image height at the CCD camera. In the next step, the image from the CCD is brought to the monitor with the magnification

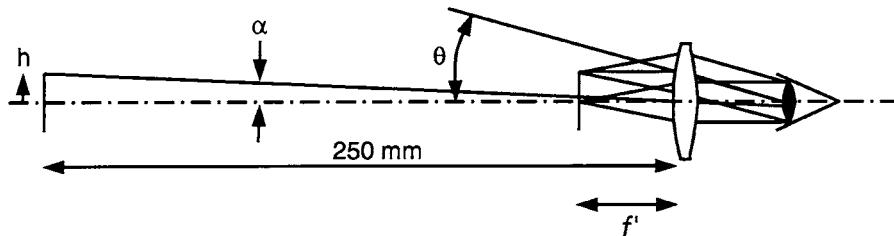
$$\frac{y''}{y'}$$

Figure 1.18
Magnification of
a Microscope



Basic Optics and Optical System Specifications

Figure 1.19
Magnification
of a Magnifier
or Eyepiece



where y'' is the image height at the monitor. In the third step the observer watches the monitor from the distance, D , with the magnification

$$\frac{250 \text{ mm}}{D}$$

Overall, the magnification of this system is

$$M = \frac{y'}{y} \frac{y''}{y'} \frac{250}{D}$$

$$M = \frac{y''}{y} \frac{250}{D}$$

The second example is a magnifier or an eyepiece, as shown in Fig. 1.19. The object with height h at a distance of 250 mm is seen to subtend an angle, α . When the same object is located in the first focal plane of the eyepiece, the eye sees the same object at an angle, θ , where

$$\alpha = \frac{h}{250} \quad \theta = \frac{h}{f}$$

Therefore, magnification M is given by

$$M = \frac{\theta}{\alpha} = \frac{250}{f}$$

The next example is the visual microscope shown in Fig. 1.20. A microscope objective is a short focal length lens, which forms a highly

Figure 1.20
Magnification of a
Visual Microscope

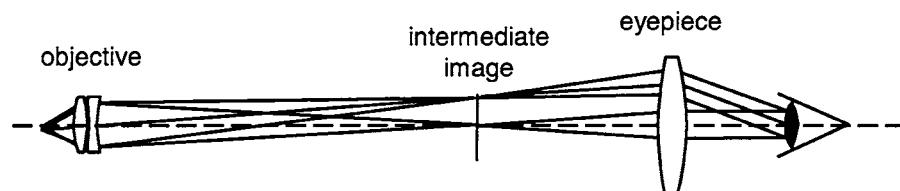
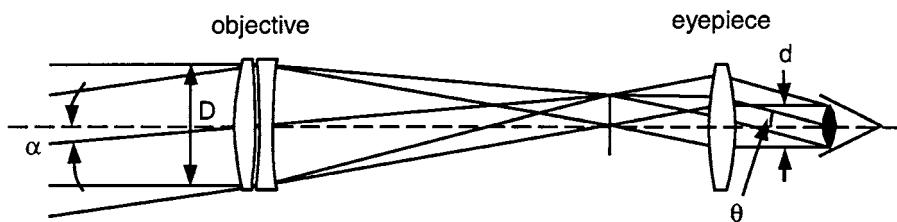


Figure 1.21
Magnification of a Visual Telescope



magnified image of the object. A visual microscope includes an eyepiece which has its front focal plane coincident with the objective image plane. The image formed by the objective is seen through the eyepiece, which has its magnification defined as

$$M_e = \frac{250}{f}$$

where f is the focal length of the eyepiece. The magnification of the microscope is the product of the magnification of the objective times the magnification of the eyepiece. Thus

$$M_m = M_o M_e$$

A visual telescope is shown in Fig. 1.21. A distant object is seen at an angle, α , without the telescope and at an angle, θ , with the telescope. The angular magnification of the telescope is

$$M = \frac{\theta}{\alpha}$$

Using the similarity of triangles, it can also be shown that the telescope magnification is

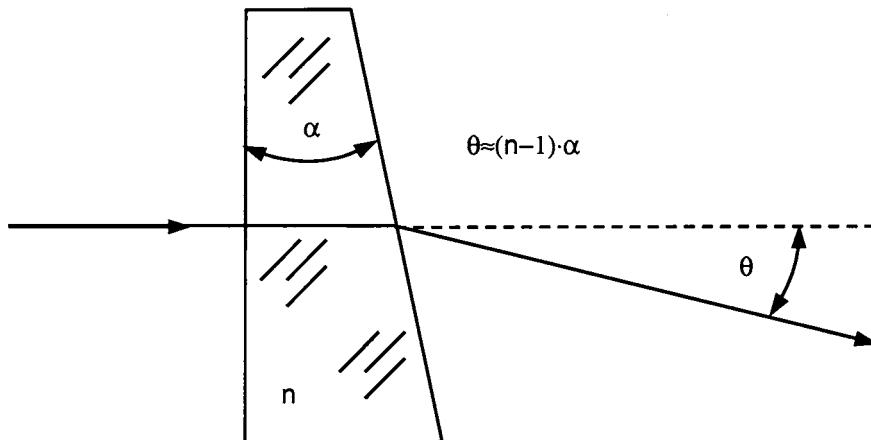
$$M = \frac{f_o}{f_e} = \frac{D}{d}$$

where f_o is the focal length of the objective, f_e is the focal length of the eyepiece, D is the diameter of the entrance pupil, and d is the diameter of the exit pupil.

There are several useful first-order relationships regarding plane parallel plates in an optical system. The first relates to what happens in an optical system when a wedge is added to a plane parallel plate. If a ray, as shown in Fig. 1.22, goes through the wedged piece of material of index

Basic Optics and Optical System Specifications

Figure 1.22
Light Deviation
through Wedged
Material



of refraction n and a small wedge angle α , the ray deviates from its direction of incidence by the angle, θ , according to

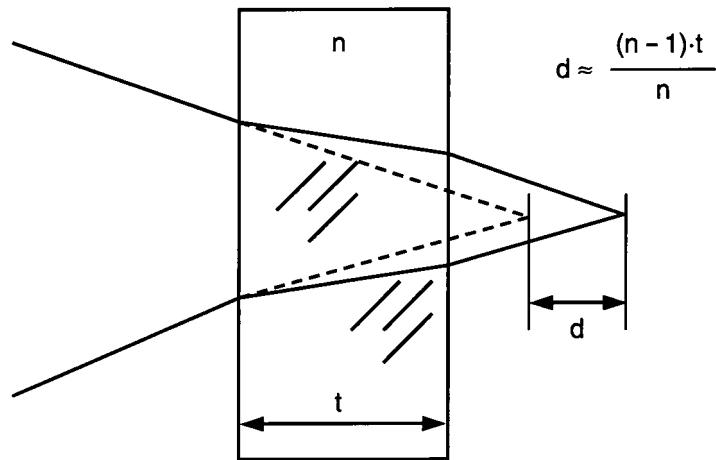
$$\theta \approx (n - 1)\alpha$$

The angle of deviation depends on the wavelength of light, since the index of refraction is dependent on the wavelength. It is important to understand how the wedge can affect the performance of the optical system. When a parallel beam of white light goes through the wedge, the light is dispersed into a rainbow of colors, but the rays of the individual wavelengths remain parallel. Therefore, the formula that gives the angle of deviation through the wedge is used to quickly determine the allowable wedge in protective windows in front of the optical system. However, if the wedge is placed into a converging beam, not only will the different colors be focused at different distances from the optical axis, but also the individual colors will be blurred. There is a term called *boresight error*, which means the difference between where you think the optical system is looking and where it really is looking. A wedged window with a wedge angle, α , will cause a system to have a boresight error of angle θ .

A plane parallel plate in a converging beam moves the image plane further along the optical axis, as shown in Fig. 1.23. If the thickness of the plate is t , the image displacement, d , along the optical axis is

$$d = (n - 1) \frac{t}{n}$$

Figure 1.23
Focus Shift
Introduced by a
Plane Parallel Plate

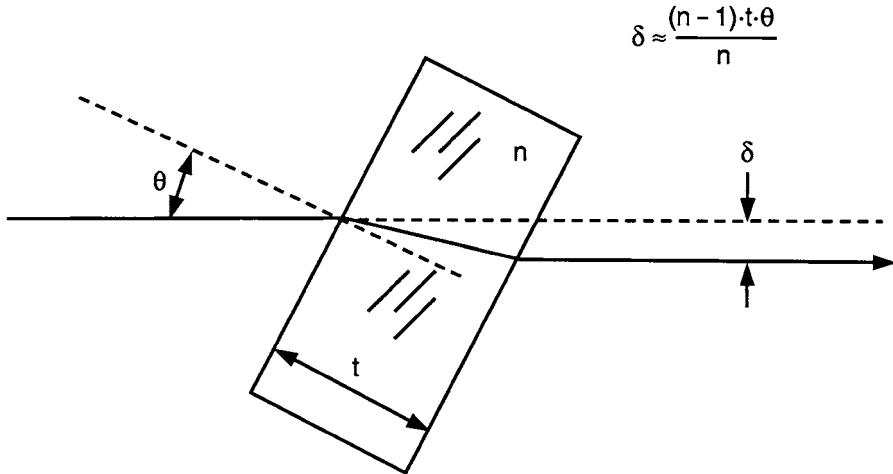


When a plane parallel plate is tilted in the optical system, as in Fig. 1.24, then the ray incident at an angle, θ , is displaced laterally by the amount, δ , given by

$$\delta = (n - 1) \frac{t\theta}{n}$$

Note that if we look through a telescope at an infinitely distant object and we put a tilted plane parallel plate in front of the telescope,

Figure 1.24
Lateral Displacement
of a Ray Introduced
by a Tilted Plane
Parallel Plate



Basic Optics and Optical System Specifications

there will be no change in the image. One would think that color fringes would be seen because the different wavelengths are displaced differently. However, because the parallel bundle of rays going through the tilted plate is only laterally displaced, it remains parallel to itself after transmission through the plate, and therefore there is no color fringing. If there is a wedge in the plate, however, chromatic dispersion will, of course, cause the appearance of color fringing.

This page intentionally left blank

CHAPTER
2

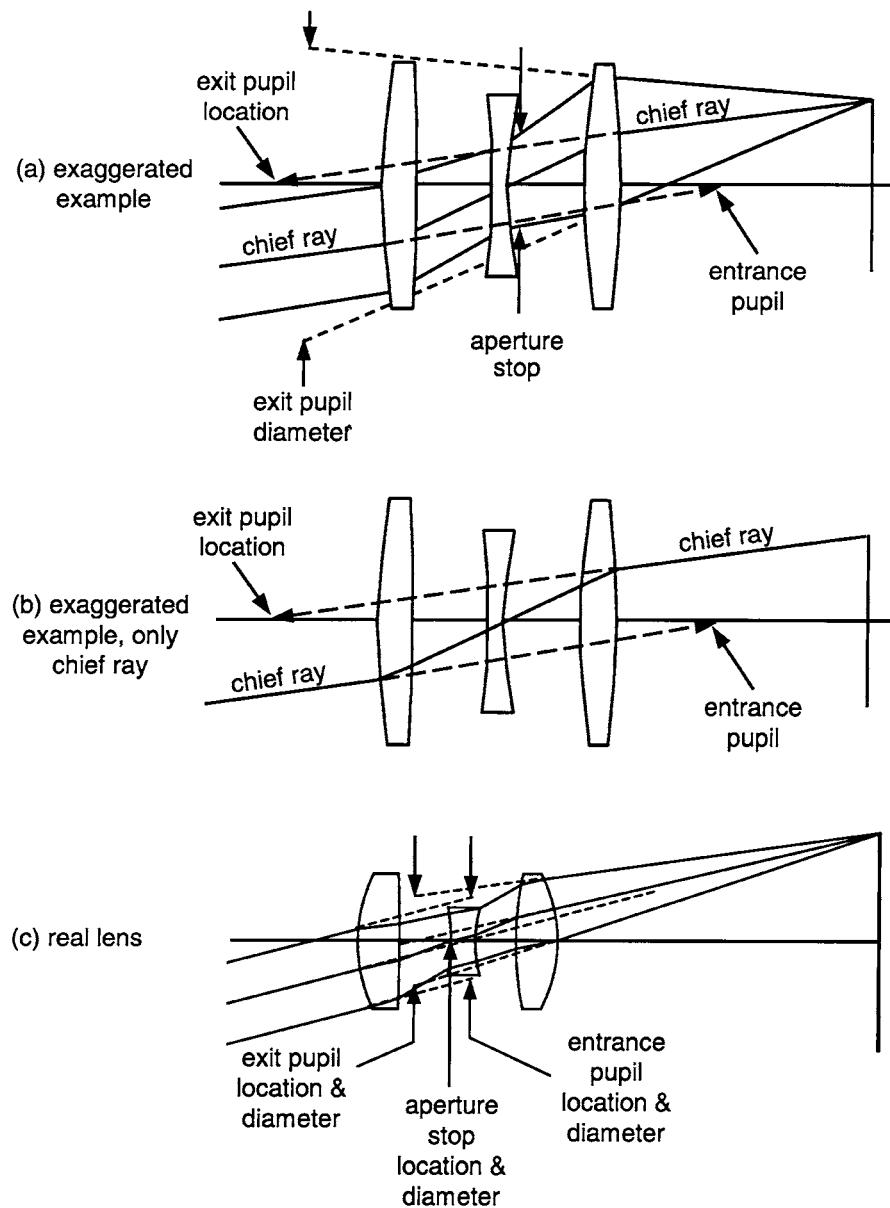
Stops and Pupils and Other Basic Principles

The Role of the Aperture Stop

In an optical system, there are apertures which are usually circular that limit the bundles of rays which go through the optical system. In Fig. 2.1 a classical three-element form of lens known as a Cooke triplet is shown as an example. Take the time to compare the exaggerated layout (Figs. 2.1a and b) with an actual computer optimized design (Fig. 2.1c). From each point in the object only a given group or bundle of rays will go through the optical system. The *chief ray*, or *principal ray*, is the central ray in this bundle of rays. The *aperture stop* is the surface in the system where all of the chief rays from different points in the object cross the optical axis and appear to pivot about. There is usually an iris or a fixed mechanical diaphragm or aperture in the lens at this location. If your lens has an adjustable iris at the stop, its primary purpose is to change the brightness of the image. The chief ray is, for the most part, a mathematical convenience; however, there definitely is a degree of symmetry that makes its use valuable. We generally refer to the specific height of the chief ray on the image as the image height.

Figure 2.1

Aperture Stop and Pupils in an Optical System



Entrance and Exit Pupils

The *entrance pupil* is the image of the aperture stop when viewed from the front of the lens, as shown in Fig. 2.1. Indeed, if you take any telescope, such as a binocular, and illuminate it from the back and look into the optics from the front, you will see a bright disk which is formed, in most cases, at the objective lens at the front of the binocular. In the opposite case, if you illuminate the system from the front, there will be a bright disk formed behind the eyepiece. The image of the aperture stop in the image space is called the *exit pupil*. If you were to write your initial with a grease pencil on the front of the objective lens and locate a business card at the exit pupil, you would see a clear image of the initial on the card.

There is another way to describe entrance and exit pupils. If the chief ray entering the lens is extended without bending or refracting by the lens elements, it will cross the optical axis at the entrance pupil. This is shown in Figs. 2.1a and *b* where only the chief ray and the pupil locations are shown for clarity. Clearly, it is the image of the aperture stop since the chief ray crosses the optical axis at the aperture stop. In a similar way, the exit pupil will be at the location where the chief ray appears to have crossed the optical axis. The location of the exit pupil can be obtained if the chief ray that exits the optical system is extended backwards until it crosses the optical axis. Both definitions are synonymous, and it will be valuable to become familiar with each.

Let us assume that we have an optical system with a lot of optical components or elements, each of them having a known clear aperture diameter. There are also a few mechanical diaphragms in the system. The question is, which of all these apertures is the aperture stop? In order to answer this question, we have to image each aperture into object space. The aperture whose image is seen from the object plane at the smallest angle is the aperture stop. It is the limiting aperture within the lens.

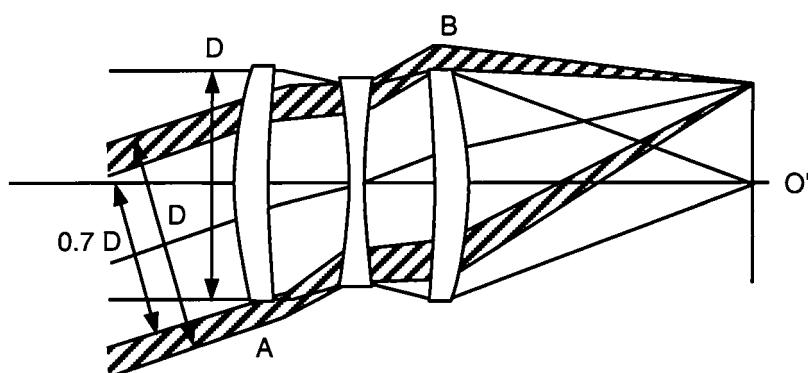
There are many systems such as stand-alone camera lenses, where the location of the entrance and exit pupils are generally not important. The exit pupil location of a camera lens will, of course, dictate the angle of incidence of the off-axis light onto the sensor. However, the specific pupil locations are generally not functionally critical. When multiple groups of lenses are used together, then the pupil locations become very important since the exit pupil of one group must match the entrance pupil location of the following group. This will be discussed later in this chapter.

Vignetting

The position of the aperture stop and the entrance and exit pupils is very important in optical systems. Two main reasons will be mentioned here. The first reason is that the correction of aberrations and image quality very much depends on the position of the pupils. This will be discussed in detail later in the book. The second reason is that the amount of light or throughput through the optical system is defined by the pupils and the size of all elements in the optical system. If ray bundles from all points in the field of view fill the aperture stop entirely and are not truncated or clipped by apertures fore or aft of the stop, then there is no *vignetting* in the system.

For a typical lens, light enters the lens on axis (the center of the field of view) through an aperture of diameter D in Fig. 2.2, and focuses down to the center of the field of view. As we go off axis to the maximum field of view, we are now entering the lens at an angle. In order to allow the rays from the entire diameter, D , to proceed through the lens, in which case the aperture stop will be filled with the ray bundle from the edge of the field, the rays at the edge of the pupil have to go through points A and B . At these positions, A and B , the rays undergo severe bending which means that they contribute significantly to the image aberrations of the system, as will be discussed in Chaps. 3 and 5. At the same time, mounting of the lenses with larger diameters is more expensive. Further, the lens will be heavier and thicker. So why don't we truncate the aperture in the plane of Fig. 2.2 to $0.7D$? We will lose approximately 30% of the energy at the edge of the field of view compared to the center of the field; however, the positive elements in our Cooke triplet example will be smaller in diameter, which means that they can also be thinner and the housing can be smaller

Figure 2.2
Vignetting



Stops and Pupils and Other Basic Principles

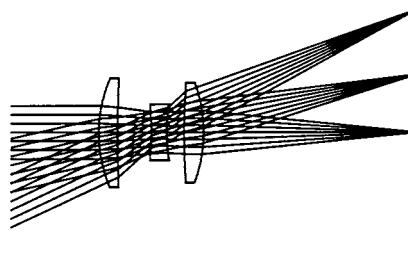
and lighter in weight. Telescopes, projectors, and other visual optical systems can have vignetting of about 30 to 40%, and the eye can generally “tolerate” this amount of vignetting. When we say that the eye can tolerate 30 to 40% vignetting, what we mean is that a slowly varying brightness over an image of this magnitude is generally not noticed. A good example is in overhead viewgraph and slide projectors where this amount of brightness falloff is common, yet unless it is pointed out, most people simply will not notice. If the film in a 35-mm camera has a large dynamic range, then this magnitude of vignetting is also acceptable in film-based photography. In digital cameras the vignetting can be calibrated out since the vignetting is a known function.

In Fig. 2.3 a triplet lens example is shown first in its original form without vignetting (Fig. 2.3a). In the next step the elements are sized for 40% vignetting, but with the rays traced as if there is no vignetting (Fig. 2.3b). In the last step the lens is shown with the vignetted bundle of rays at the edge of the field (Fig. 2.3c).

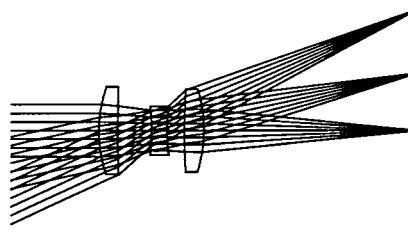
Figure 2.3

Example of
Vignetting

(a) Basic design $f/5 \pm 20$
degree field, no
vignetting



(b) Elements sized for
40% vignetting



(c) Final design with 40%
vignetting

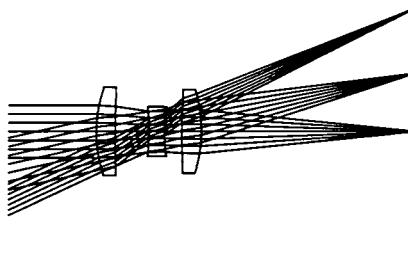
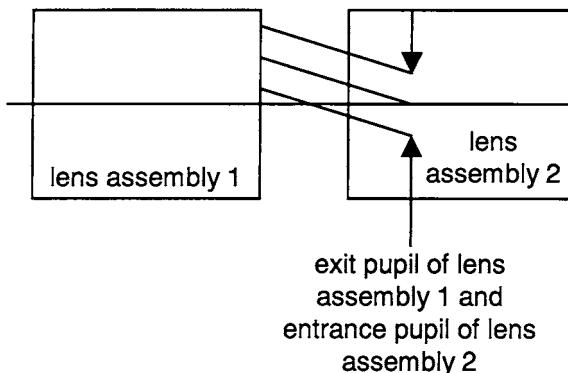


Figure 2.4
Matching of Pupils



Although vignetting is acceptable and often desirable in visible optical systems, it can be devastating in thermal infrared optical systems because of image anomalies, as will be discussed in Chap. 12. One must also be very careful when specifying vignetting in laser systems, as will be discussed in Chap. 11.

When a system is designed using off-the-shelf components with a combination of two or more modules or lens assemblies, it is very important to know the positions of the entrance and exit pupils of these modules. The exit pupil of the first module *must* coincide with the entrance pupil of the second module, etc. This is shown in Fig. 2.4.

There can be a very serious pupil-matching problem when using off-the-shelf (or even custom) zoom lenses as modules in optical systems. Zoom lenses have a given size and position of their pupils which change as a function of zoom position or focal length. It is very easy to make a mistake when the exit pupil of the first module is matched to the entrance pupil of the second module for only one zoom position. When the pupils move with respect to one another through zoom and do not image from one to another, we can lose the entire image.

CHAPTER
3

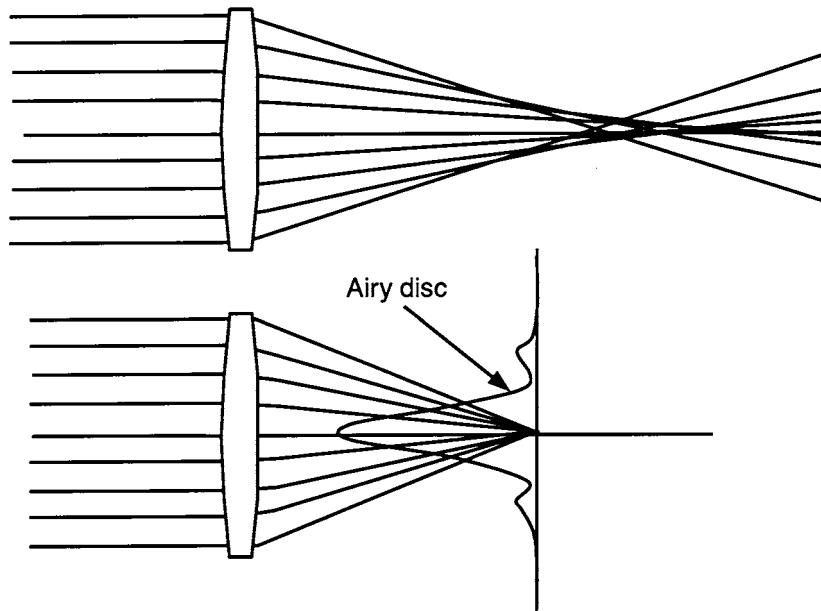
Diffraction, Aberrations, and Image Quality

What Image Quality Is All About

Image quality is never perfect! While it would be very nice if the image of a point object could be formed as a perfect point image, in reality we find that image quality is degraded by either geometrical aberrations and/or diffraction. Figure 3.1 illustrates the situation. The top part of the figure shows a hypothetical lens where you can see that all of the rays do not come to a common focus along the optical axis. Rather, the rays entering the lens at its outer periphery cross the optical axis progressively closer to the lens than those rays entering the lens closer to the optical axis. This is one of the most common and fundamental aberrations, and it is known as *spherical aberration*. Geometrical aberrations are due to the failure of the lens or optical system to form a perfect geometrical image. These aberrations are fully predictable to many decimal places using standard well-known ray trace equations.

If there were no geometrical aberrations of any kind, the image of a point source from infinity is called an *Airy disk*. The profile of the Airy

Figure 3.1
Image Quality, Geometrical Aberrations
(Top) and Diffraction Limited (Bottom)



disk looks like a small gaussian intensity function surrounded by low-intensity rings of energy, as shown in Fig. 3.1, exaggerated.

If we have a lens system in which the geometrical aberrations are significantly larger than the theoretical diffraction pattern or blur, then we will see an image dominated by the effect of these geometrical aberrations. If, on the other hand, the geometrical aberrations are much smaller than the diffraction pattern or blur, then we will see an image dominated by the effect of the Airy disk. If we have a situation where the blur diameter from the geometrical aberration is approximately the same size as the theoretical diffraction blur, we will see a somewhat degraded diffraction pattern or Airy disk. Figure 3.1, while exaggerated, does show a situation where the resulting image would, in fact, be a somewhat degraded Airy disk.

What Are Geometrical Aberrations and Where Do They Come From?

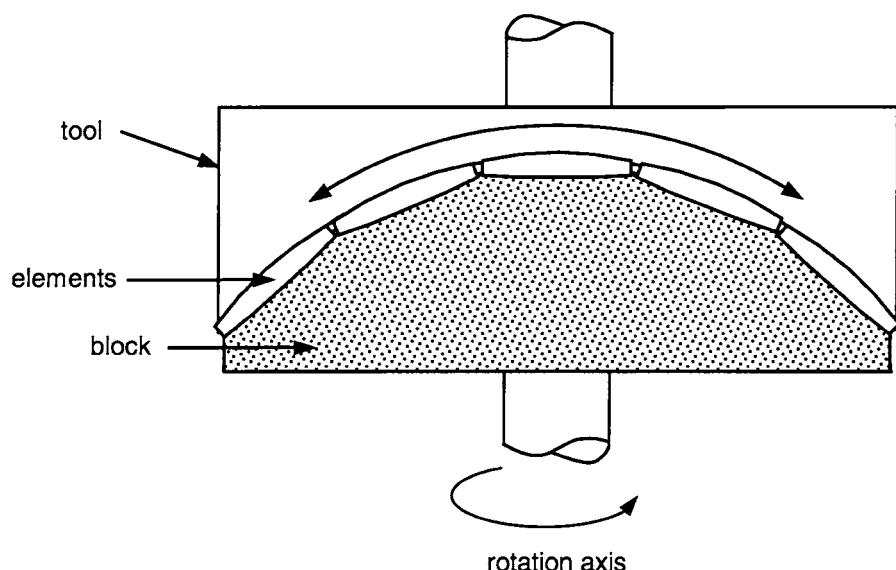
In the previous section, we have shown the distinction between geometrical aberrations and diffraction. The bottom line is that imagery formed by lenses with spherical surfaces simply is not perfect! We use

Diffraction, Aberrations, and Image Quality

spherical surfaces primarily because of their ease of manufacture. In Fig. 3.2, we show how a large number of elements can be ground and polished using a common or single tool. The elements are typically mounted to what is called a *block*. Clearly, the smaller the elements and the shallower the radius, the more elements can be mounted on a given block. The upper tool is typically a spherical steel tool. The grinding and polishing operation consists of a rotation about the vertical axis of the blocked elements along with a swinging motion of the tool from left to right, as indicated by the arrows. The nature of a sphere is *that the rate of change of slope is constant everywhere on a sphere*, and because of this mathematical definition, the tool and lens surfaces will only be in perfect contact with one another over the full range of motions involved when both are perfectly spherical. Due to asymmetries in the process, the entire surface areas of the elements and tool are not in contact the same period of time. Hence this process is not perfect. However, the lens surfaces are driven to a near-spherical shape in reasonable time by a skilled optician. This is the reason we use spherical surfaces for most lenses. Chapter 18 discusses optical component manufacturing in more detail.

We will discuss the use of nonspherical or aspheric surfaces in Chap. 8.

Figure 3.2
Manufacture of
Spherical Lens
Surfaces



Earlier we said that geometrical aberrations are due entirely to the failure of the lens or optical system to form a perfect geometrical image. Maxwell formulated three conditions that have to be met for the lens to form a perfect geometrical image:

1. All the rays from object point O after passing through the lens, must go through the image point O' .
2. Every segment of the object plane normal to the optical axis that contains the point O must be imaged as a segment of a plane normal to the optical axis, which contains O' .
3. The image height, h' , must be a constant multiple of the object height, h , no matter where O is located in the object plane.

Violation of the first condition results in the image degradation, or image aberrations. Violation of the second condition results in the presence of image curvature, and violation of the third condition in image distortion. A different way to express the first condition is that all the rays from the object point, O must have the same *optical path length* (OPL) to the image point, O' .

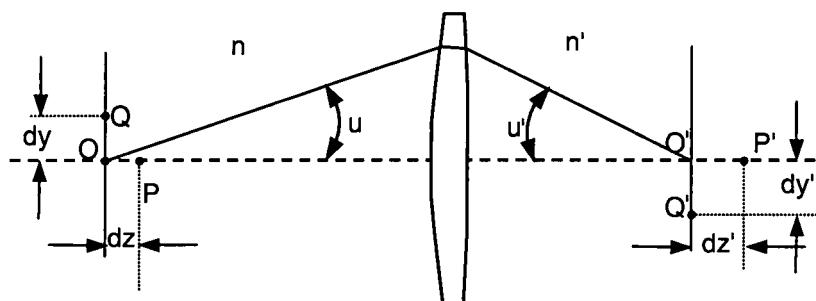
$$\text{OPL} = \int_{O(x,y)}^{O'(x',y')} n(s)ds$$

where $n(s)$ is the index of refraction at each point along the ray path, s .

The lenses that meet the first Maxwell condition are called stigmatic. Perfect stigmatic lenses are generally stigmatic only for one pair of conjugate on-axis points. If the lens shown in Fig. 3.3 is to be stigmatic not only for the points, O and O' , but also for the points, P and P' , it must satisfy the *Herschel condition*

$$n dz \sin^2 \frac{u}{2} = n' dz' \sin^2 \frac{u'}{2}$$

Figure 3.3
Definition of Paraxial
Lens



Diffraction, Aberrations, and Image Quality

If the same lens is to be stigmatic at the off-axis conjugate points, Q and Q' , it must satisfy the *Abbe sine condition*

$$n dy \sin u = n' dy' \sin u'$$

Generally these two conditions cannot be met exactly and simultaneously. However, if the angles u and u' are sufficiently small, and we can substitute the sine of the angle with the angle itself

$$\sin u \approx u \quad \text{and} \quad \sin u' \approx u'$$

then both the Herschel and Abbe sine condition are satisfied. We say that the lens works in the *paraxial region*, and it behaves like a perfect stigmatic lens. The other common definition of paraxial optics is that paraxial rays are rays “infinitely close to the optical axis.” This is a fine and correct definition; however, it can become difficult to understand when we consider tracing a paraxial ray through the edge of a lens system, a long way from the optical axis. This creates a dilemma since rays traced through the edge of the system are hardly infinitely close to the optical axis! This is why the first definition of paraxial optics, that is, using the small-angle approximation to the ray tracing equations, as would be the case for rays infinitely close to the optical axis, is easier to understand. Consider Fig. 3.4a, where we show how the rays are refracted at the interface between two optical media, according to Snell’s law.

$$n \sin \theta = n' \sin \theta'$$

Figure 3.4
A Real Ray Trace and
a Paraxial Ray Trace
through a Lens

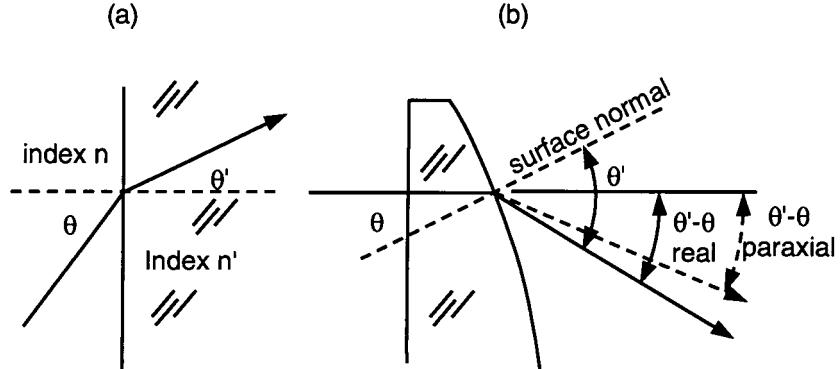


TABLE 3.1

Paraxial Approximation Versus Real Ray Angles of Refraction

θ	$\theta' - \theta$ real (degrees)	$\theta' - \theta$ paraxial (degrees)	Difference (degrees)
1	0.5001	0.5	0
10	5.0981	5	0.0981
20	10.8659	10	0.8659
30	18.5904	15	3.5904
40	34.6186	20	14.6186

In Table 3.1 we show just how a real ray, according to Snell's law, and a paraxial ray, using the small-angle approximation $\sin \theta \sim \theta$, refract or bend after refraction from a spherical surface at a glass-air interface (index of refraction of glass $n = 1.5$). These data use the nomenclature of Fig. 3.4*b*. Note that the difference in angle between the paraxial and the real rays define the resulting image blur. For angles of incidence θ of 10° or less we see that the real refracted ray is descending within 0.1° of the paraxial ray (0.0981° difference at a 10° angle of incidence). However, as the angles of incidence increase, the difference between the real and the paraxial descending angles increases quite significantly. This is where aberrations come from.

Along with this understanding, it is evident that in order to keep aberrations small, it is desirable if not mandatory to keep the angles of incidence as small as possible on the various surfaces within your system.

What Is Diffraction?

Diffraction is a phenomena or effect resulting from the interaction of light (which of course is electromagnetic radiation) with the sharp limiting edge or aperture of an optical system. While we could very easily fill the next few pages with integral signs and Bessel functions, it is not the intention of the authors to provide this level of detail. Rather, the following explanation is easy to follow and should provide a sufficient level of understanding of the causes of diffraction and resultant observable effects.

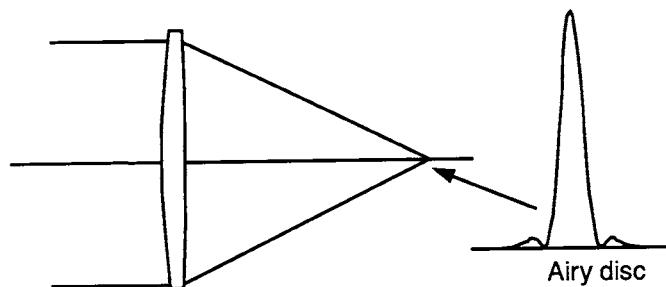
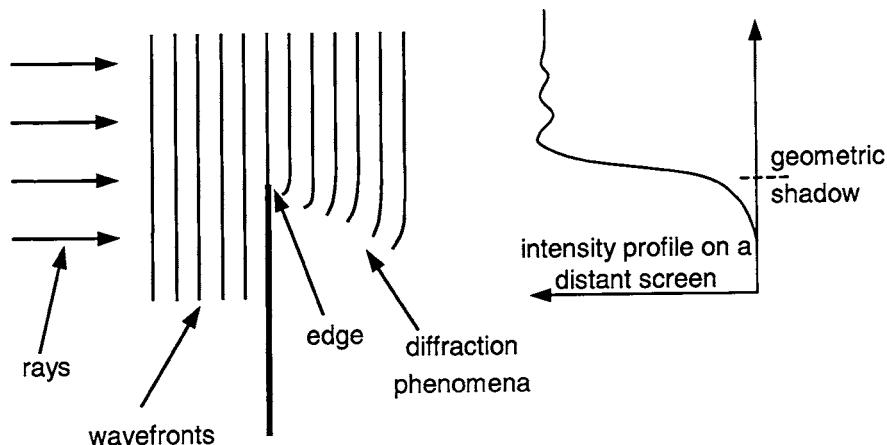
Imagine a swimming pool at 3 o'clock in the morning with no wind present; the water is like a sheet of glass. Imagine throwing a large rock

Diffraction, Aberrations, and Image Quality

into one end of the pool. Water waves will emanate outward from where the rock has entered the water as concentric, ever-expanding circles. Before proceeding onward, note that the physics of water waves is virtually identical to the physics of electromagnetic radiation, and all of the derivations are quite analogous.

Now let us proceed to the other end of the swimming pool. If the pool is large enough, the water waves will be nearly straight and parallel to one another. In reality, of course, they are going to be curved and centered about the point where the rock entered the water. For this discussion, consider the water waves as straight. Let us now immerse a 1- by 2-m sheet of plywood partially into the pool, as shown at the top of Fig. 3.5. What you will see to a reasonable extent above the edge of the board at the top of Fig. 3.5 is that the water waves will continue to propagate left

Figure 3.5
Diffraction Effects



to right undisturbed. Below the edge where the major part of the board is located, you will see to the right of the board virtually no disturbance in the water. To the right of the intersection of the upper edge of the board and the water waves, you will see little curlicues traveling or emanating outward from the edge of the board. These curlicues are, in reality, *diffraction* of water waves. We sincerely hope that none of our readers would think there would be a sharp step in the water to the right of the edge... if you did think this was the case, we urge you to try this little experiment in your backyard swimming pool.

The peaks of the water waves are called *wavefronts*. Perpendicular to the wavefronts are the *rays*. While we rarely if ever talk about "water rays," we certainly do talk all the time about light rays. The important point here is that the rays are perpendicular or normal to the wavefronts and the wavefronts are perpendicular to the rays. Throughout your reading of this book, we would like you to understand the difference between ray optics and wave optics.

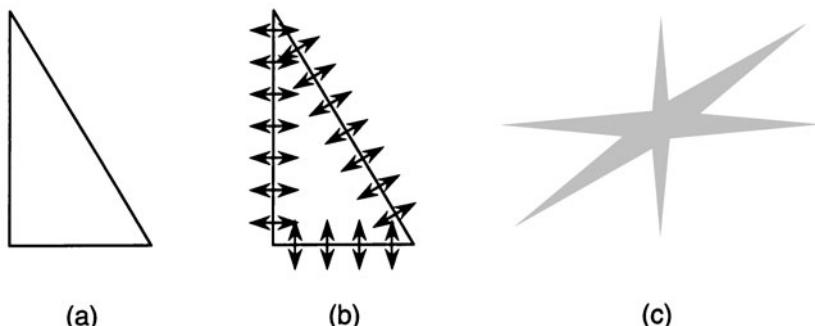
Now, back to our little example. If instead of our swimming pool example we were to have parallel or collimated light incident upon the edge of a razor blade, we would have diffraction of the electromagnetic radiation much in the same way as we showed diffraction of water waves. On a distant screen or card you will not see a very sharp edge or step function, but rather an intensity gradient with slight variations in intensity occurring in a similar fashion to the curlicues of the water waves.

The previous explanation represents our attempt to illustrate how diffraction occurs without the lengthy and messy mathematical derivations. If we now have a typical lens system, it is easy to understand that there will, by definition, be a limiting edge or aperture at which the light effectively stops or is blocked. This edge, which in many cases is the aperture stop of your system, wraps around the optical axis generally in a symmetrical, circular fashion, and the resulting diffraction pattern acquires a rotationally symmetric shape known as the "Airy disk."

It is important to note that diffraction occurs perpendicular to an edge. Since a circular aperture, in effect, wraps around in a full 360° , the resulting diffraction pattern (the Airy disk) is a rotationally symmetrical blur. However, if your aperture were a triangular shape as shown in Fig. 3.6a, the resulting diffraction pattern would be star shaped with three spikes as shown in Fig. 3.6c. The reason there are three notable spikes is that the diffraction spreading has occurred perpendicular to the three straight edges of the aperture, as shown in Fig. 3.6b. Note that the relative length of the spikes is proportional to the length of the edge.

Diffraction, Aberrations, and Image Quality

Figure 3.6
Diffraction from a Triangular Aperture



Diffraction-Limited Performance

As discussed previously, if the geometrical aberrations are significantly smaller than the diffraction blur, the image is, in effect, well represented by the Airy disk. This form of optics is called *diffraction-limited optics*. Understanding the limits of diffraction-limited optics becomes extremely important, especially with today's extremely demanding levels of performance. Figure 3.7 shows two very important principles: (1) the physical diameter of the Airy disk and (2) the angular diameter or subtense of the Airy disk. It can be shown that:

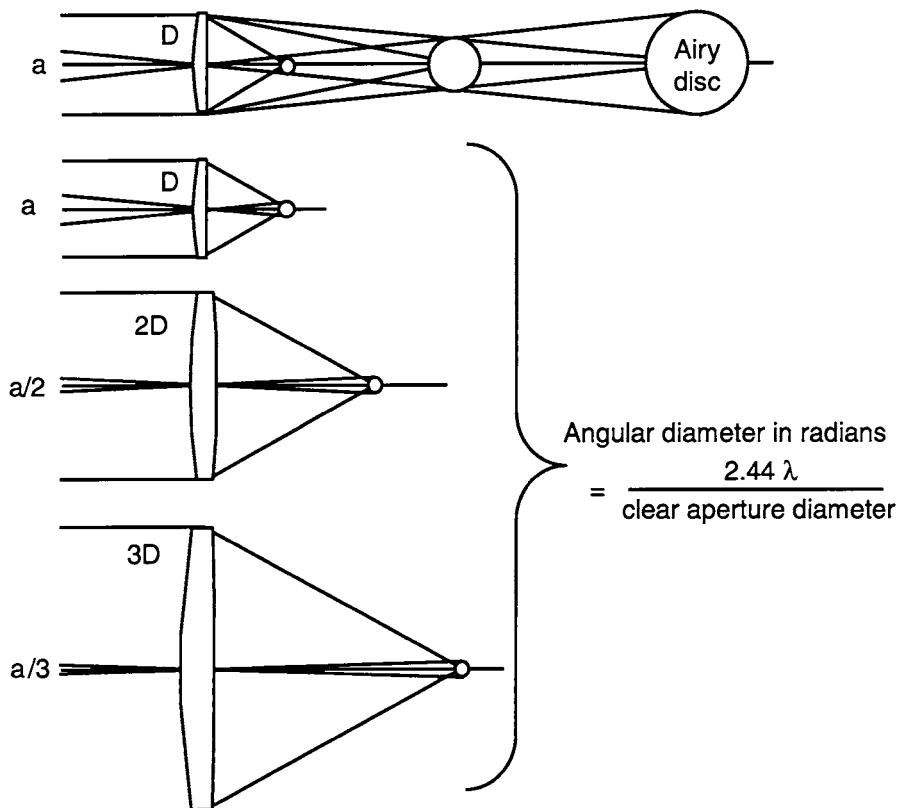
$$\text{Physical diameter of the Airy disk} = 2.44 \lambda f/\text{number}$$

Shown in the top part of Fig. 3.7 are three different lenses, all of diameter D . One lens focuses images fairly close to the lens, the second has a somewhat longer focal length, and the third a still longer focal length. For these three lenses, all of which have the same entrance pupil diameter D , the $f/\#$ increases in proportion to the increase in focal length. From the equation, we see that the Airy disk increases in diameter in direct proportion to the $f/\#$ increase and thus in proportion to the focal length as well. A very useful rule of thumb to remember is: *The Airy disk diameter in the visible part of the spectrum is approximately equal to the $f/\#$ expressed in microns.*

This is easy to see if you consider a wavelength of $0.5 \mu\text{m}$, which would be approximately the center of the visible spectrum. In this case, the physical diameter of the Airy disk

$$D = 2.44 \times 0.5 f/\# = 1.22 f/\#$$

Figure 3.7
A Clarifying Illustration of "Diffraction-Limited" Imagery



which is approximately equal to the $f/\#$ number itself expressed in microns!

We now show three separate lenses of diameter D , $2D$ and $3D$ all of identically the same $f/\#$. What this means is that the Airy disk or diffraction blur will be identical in all three cases. You can see quite clearly that for each lens the focal length increases in proportion to the increase in diameter (since the $f/\#$ is identical). What this means is that the angular subtense of the Airy disk also decreases in proportion to the diameter or the increase in focal length. The resulting relationship becomes

$$\text{Angular diameter of the Airy disk} = \frac{2.44 \lambda}{\text{clear aperture diameter}}$$

The angular diameter is expressed in radians if the wavelength and the clear aperture diameter are in the same units.

Note that in all of the preceding discussion, the diameter of the Airy disk is assumed to be the diameter of the first dark ring in the diffraction pattern.

Derivation of System Specifications

There is a broad term “systems analysis” which generally refers to the task of deriving the basic optical system parameters based on the functional system performance requirements. We can apply what we learned earlier to perform a simple, yet noteworthy systems analysis example.

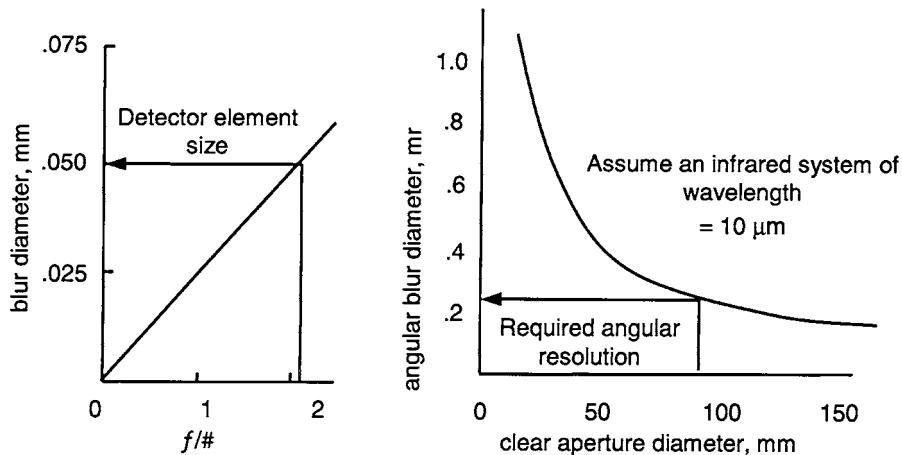
Consider, for example, an optical system used in the long-wave infrared (IWIR) which operates in the 8- to 12- μm spectral band. Our task is to derive the system $f/\#$ and clear aperture diameter. Let us assume that the detector is mercury cadmium telluride or (HgCdTe) with a 50- μm pixel pitch and further assume that we need to resolve 0.25 mrad in object space. These values are typical for an IWIR system such as a forward looking infrared (FLIR).

Earlier in this chapter, we discussed that as a rule of thumb the smallest resolvable image blur should be matched to the pixel size of the detector (sensor), that is, smallest element size. Thus, we would require that the diffraction blur or Airy disk should be approximately the same diameter as our 50- μm pixel. Recall that the diameter of the Airy disk, $D = 2.44 \lambda f/\#$, and we can solve for the $f/\#$ to produce a 50- μm Airy disk diameter, and the result is $f/2.2$ at $\lambda = 10 \mu\text{m}$. Before we continue, it is interesting to note that for diffraction-limited optics, an $f/2.2$ system that is 6 mm in diameter will have exactly the same diffraction blur diameter as an $f/2.2$ system that is 3 m in diameter, and that is 50 μm !

For a given diffraction blur diameter, as the focal length increases, the angular subtense of the Airy disk decreases proportionally. We can use the relationship that the angular diameter of the Airy disk = $2.44 \lambda / (\text{clear aperture diameter})$ to solve for the clear aperture diameter required so that the 50- μm Airy disk subtends 0.25 mrad in object space, and the result is a 100-mm-diameter clear aperture.

Figure 3.8 shows parametrically how the $f/\#$ and clear aperture diameters relate to the diffraction-limited image blur or Airy disk diameter and the angular subtense of the diffraction blur. This illustrates how we can quickly and easily take the most basic system functional requirements and derive the system $f/\#$ and clear aperture diameters. Do keep

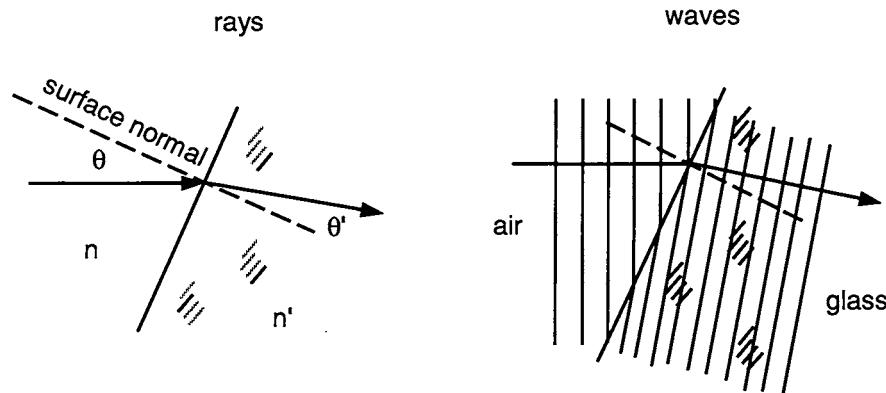
Figure 3.8
Example of Systems Analysis



in mind that this assumes diffraction-limited optics. Further, it is based on the criteria that the Airy disk is matched to the pixel pitch. These are generally good assumptions to work with, and as your system needs become better understood, you may need to revise the results.

As we begin to learn more about image formation, it is important to understand just how light bends or refracts when passing through an air-glass or glass-air interface. As shown in Fig. 3.9, the *refractive index of a material* = n , where n is the ratio of the velocity of light in a vacuum to the velocity of light in the denser material such as glass. Since the light or electromagnetic radiation slows down in the denser material, the refractive index is always greater than unity. For optical glass the refractive index ranges from about $n = 1.5$ to $n = 1.85$.

Figure 3.9
Bending of Light at an Optical Surface



Diffraction, Aberrations, and Image Quality

According to the Snell's law, $n \sin \theta = n' \sin \theta'$. With air on the input side of the interface, the equation reduces to $\sin \theta = n' \sin \theta'$. For small angles the equation further reduces to $\theta = n'\theta'$. We will use this result later!

We discussed earlier how light could be represented by either rays or wavefronts, where the two are orthogonal to one another. We will be using both representations throughout this book and we hope that you will become "bilingual" or fluent with both representations. To help understand these concepts, we show in Fig. 3.9 how a light ray, as well as a series of wavefronts, is incident on an air-glass interface and how the light bends or refracts. From ray optics, we can simply use Snell's law to determine the angle of refraction. Consider how we represent the same thing using wave optics. The wavefronts are traveling from left to right, with their peaks separated by the wavelength of light. As the wavefront enters the denser medium such as glass, its velocity is reduced by $1/n$, with the result being that the wavefronts are closer together. There is a fundamental law of physics, which says that the wavefronts must be continuous at the interface between the media. Considering the velocity reduction along with the wavefront continuity requirement, we can see how the entire wavefront is rotated around in a clockwise direction as it proceeds into the denser medium. Interestingly, you can use this construction to rederive Snell's law!

This page intentionally left blank

The Concept of Optical Path Difference

Optical Path Difference (OPD) and the Rayleigh Criteria

OPD is an extremely useful measure of the performance of an imaging optical system. If the wavefronts proceeding to a given point image are spherical, concentric, and centered at the point image for a given field of view, then the imagery will be geometrically perfect, or *diffraction limited*. As shown earlier, the image will then be a perfect Airy disk. This is, in effect, the reverse of our earlier example where we threw a rock into a pool of water to illustrate the wave nature of light and diffraction. If we think of the water waves traveling in reverse to where the rock entered the water, we will emulate light imaging to a point image. By definition, the wavefronts will be perfectly spherical, concentric, and centered where the rock entered the water. Recall also that rays are perpendicular to the wavefronts. It is thus clear that if the wavefronts are spherical, concentric, and centered at a point in the image, then the rays will all come to that same point as defined by the center of curvature of the wavefronts. As we learned earlier, diffraction at the limiting edge of the pupil will create an Airy disk, which is the reason why we do not have a perfect point image.

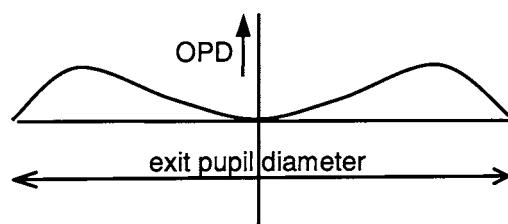
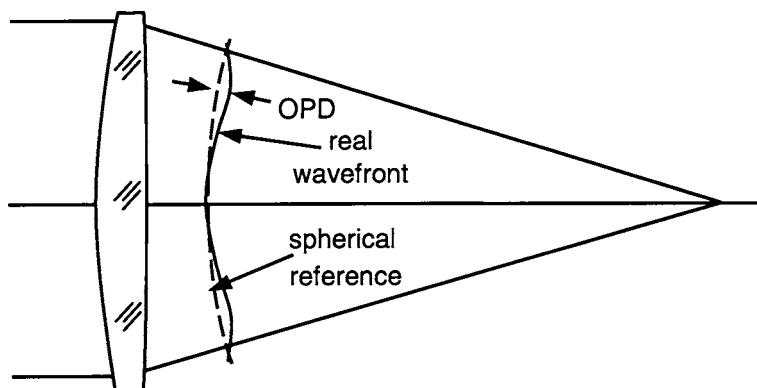
Consider Fig. 4.1 where we show a hypothetical lens with a perfectly spherical reference wavefront and a real wavefront. The real wavefront departs from sphericity due to aberrations induced by the lens. The *optical path difference* is the difference between the real wavefront and a spherical reference wavefront, which is usually selected to be a near best fit to the aberrated wavefront.

One of the reasons the OPD is so valuable a parameter is evident from the Rayleigh criteria. Lord Rayleigh (real name William Strutt, a Nobel Prize winner for discovering the gas argon) showed that:

An optical instrument would not fall seriously short of the performance possible with an absolutely perfect system if the distance between the longest and shortest paths leading to a selected focus did not exceed one-quarter of a wavelength.

What the Rayleigh criteria says is that if the OPD is less than or equal to one-quarter of a wave (one-quarter of the wavelength of the light),

Figure 4.1
Optical Path Difference (OPD)



The Concept of Optical Path Difference

then the performance will be almost indistinguishable from perfect. If this is the case, then the imagery of a point object will be very nearly a perfect Airy disk. This is a *very* useful tool, and as will become evident, its validity is quite broadly applicable. It is important to note, however, that it is not 100% infallible, and should only be used as a guide or relative measure of the level of optical performance.

Figure 4.2 shows the appearance of the image of a point source, which is known as a *point-spread function* (PSF), for optical path differences of 0 waves (a perfect Airy disk), 0.25 wave, 0.5 wave, and 1.0 wave. Note that

Figure 4.2
Image of a Point Source with Different Amounts of Peak-to-Valley Optical Path Difference Due to Coma

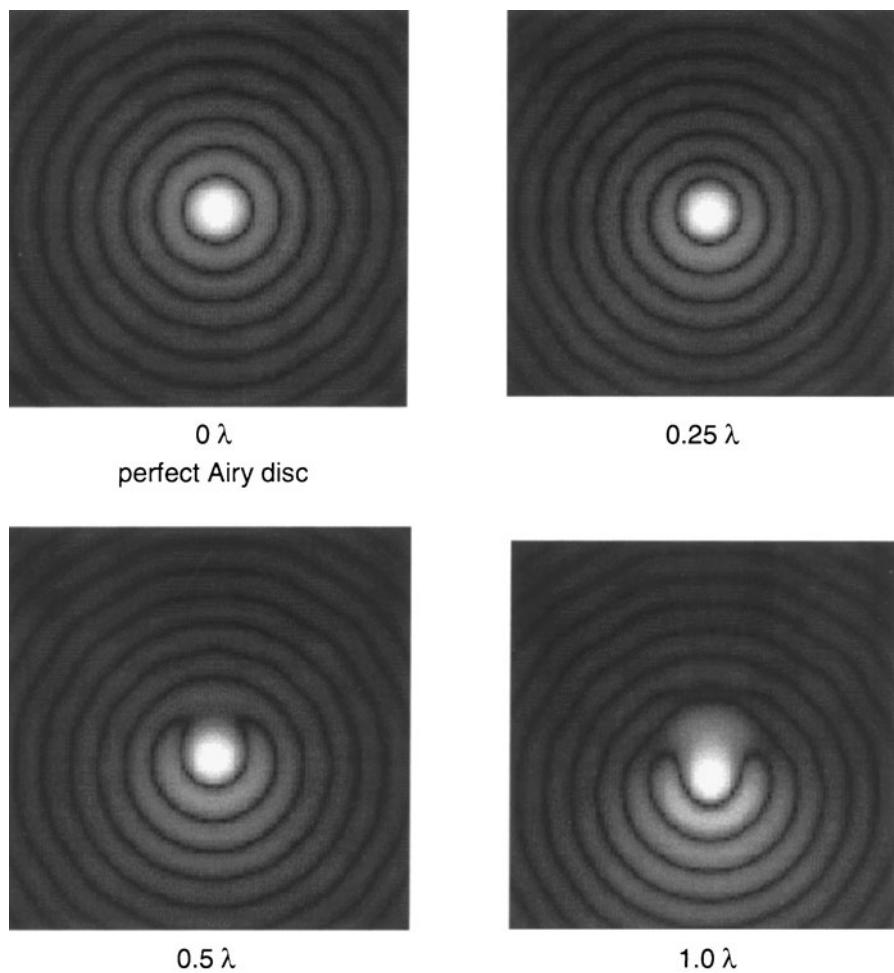
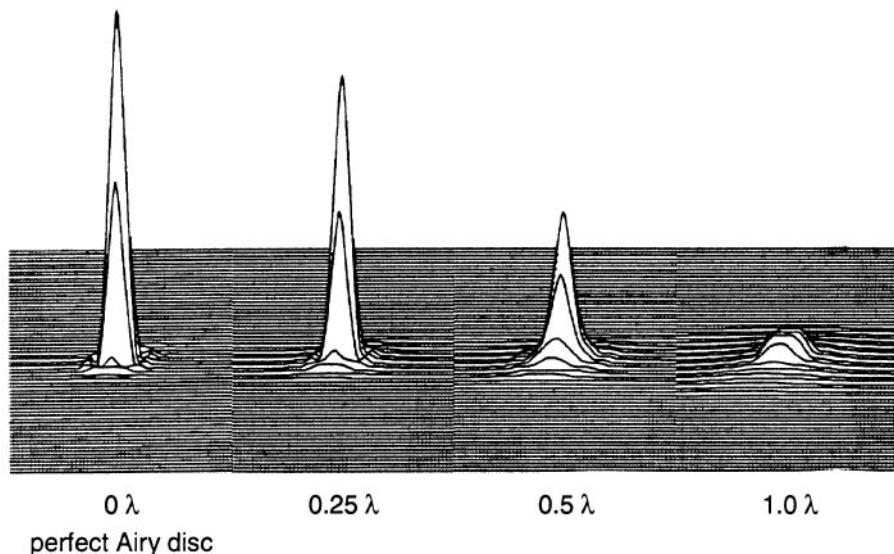


Figure 4.3
Image of a Point
Source with Different
Amounts of Peak-to-
Valley Optical Path
Difference Due to
Spherical Aberration

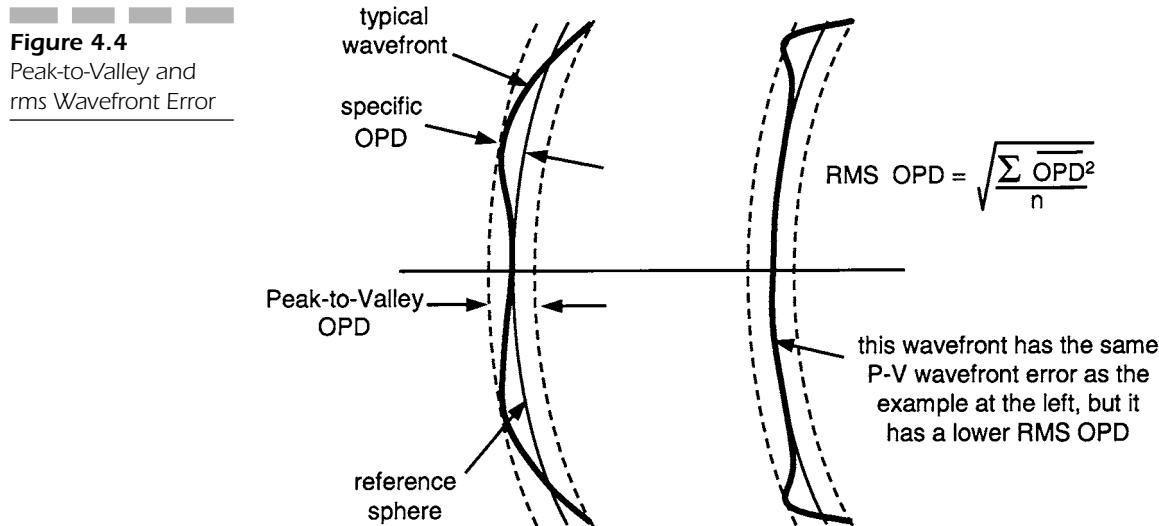


the 0.25 wave imagery is qualitatively nearly indistinguishable from the perfect Airy disk. The character of the central maximum is maintained, and the first bright ring is fully intact. As soon as we go to 0.5 wave and above, the imagery is clearly degraded from perfect. Figure 4.3 shows perspective views of the point spread function for the same values of the OPD; however, these data are for spherical aberration rather than coma. Here, too, we can conclude that the 0.25 wave imagery is nearly indistinguishable from perfect. We do see a drop in peak intensity; however, the overall character of the pattern is very similar to the perfect Airy disk.

Peak-to-Valley and RMS Wavefront Error

The OPD as shown here is known as *peak-to-valley* (P-V) optical path difference. Peak-to-valley is the total difference between the portion of the wavefront closest to the image (leading, or ahead of the reference wavefront) and the farthest lagging portion of the wavefront (lagging, or behind the reference wavefront). Figure 4.4 shows this as the separation between the two dashed reference spheres.

The Concept of Optical Path Difference



There is another term, and that is *rms wavefront error*. The definition of rms wavefront error is shown in Fig. 4.4 as the square root of the sum of the squares of the OPDs as measured from a best-fit reference spherical wavefront over the total wavefront area. The rms wavefront error represents more of an averaging over the wavefront than the P-V wavefront error. The example shown on the right-hand side of Fig. 4.4 has the same P-V OPD as the left-hand side; however, the rms would be lower. This is because most of the wavefront error, or wavefront distortion, is at the outer periphery of the aperture, and over most of the area of the wavefront, the wavefront is nearly perfect.

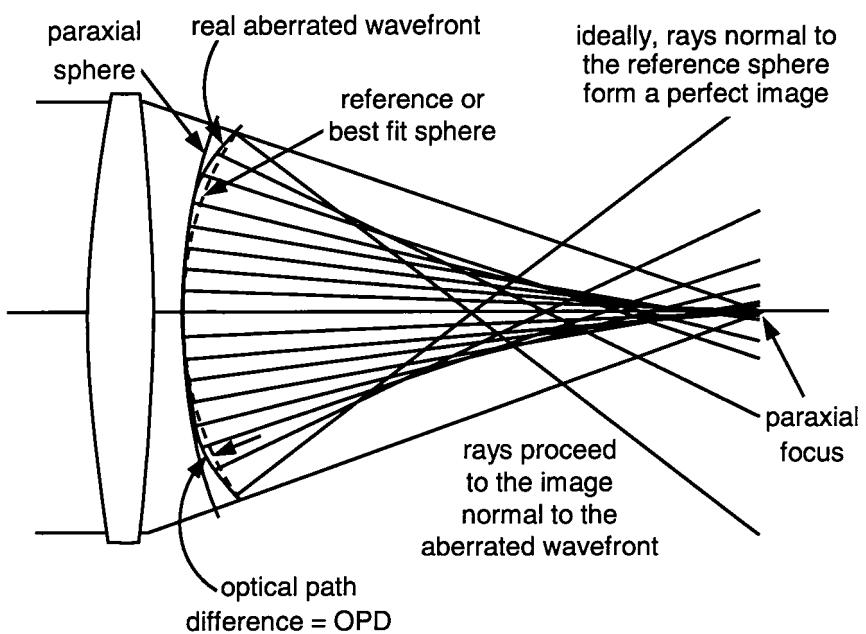
Consider, for example, a large telescope mirror 3 m in diameter. In order to assure near-diffraction-limited performance, let us assume that the P-V wavefront error is specified as 0.125 wave on the surface. This is the P-V departure from the ideal or perfect surface profile. In reflection, the wavefront departure will be double this value, or 0.25 wave which just meets the Rayleigh criteria. Now let's further assume that the optical shop produces a mirror which has a P-V surface departure from the nominal of 0.02 wave, with the exception of a small depression the size of a pencil eraser 0.5 wave deep. The mirror is clearly out of spec as the reflected wavefront will have a P-V error of 1.0 wave, which is four times the Rayleigh criteria. However, the area of this small depression in the surface would be 0.0004% of the total mirror area, an almost negligible amount. This will have virtually zero effect on the optical performance

of our telescope, and if the scattering from such an error were of concern, we could simply paint the 6-mm-diameter depression with a flat black paint. While there will still be some scattering from the mirror/paint interface, this, too, will be extremely negligible in all but the most demanding applications (such as with space telescopes).

The rms wavefront error typically ranges from approximately one-fifth to one-third of the P-V error. This ratio is highly dependent on the correlation of the wavefront, where the correlation is the inverse number of bumps over the surface. For a given number of bumps, a lower correlation has greater surface slope errors and conversely if we assume a ratio of $5\times$ between P-V and rms, the Rayleigh criteria of 0.25 wave P-V equates to 0.05 wave rms.

Figure 4.5 shows an exaggerated illustration of just how an aberrated wavefront proceeds to an image. This figure reminds us of several key points such as the fact that rays are perpendicular to the wavefront. The peak-to-valley OPD is the maximum deviation from the real wavefront and a spherical reference wavefront, which best fits the real wavefront. While the figure is quite exaggerated, it is drawn to scale and the various factors we have learned about all apply.

Figure 4.5
OPD Showing Wavefronts and Ray Paths



The Wave Aberration Polynomial

The optical path difference, or the wave aberration function, can be mathematically expressed in the form of a polynomial for rotationally symmetric optical systems.

A single ray proceeding from a given point in the object through an optical system is defined by the coordinates in the object plane and its coordinates in the pupil of the system. The wave aberration function can be expressed as a Taylor expansion polynomial in field and pupil coordinates. The wave aberration polynomial can be simplified by using the symmetrical properties of the optical system. In its final form the wave aberration polynomial has two quadratic terms which are not the intrinsic aberrations—they present a focal shift, five terms of the fourth power of the field and pupil coordinates which are primary aberrations, and sixth, eighth, and tenth, etc., power terms which are higher-order aberrations.

In order to obtain the coefficients in the wave aberration polynomial, it is sufficient to trace a small number of rays and then fit the data to the polynomial. To obtain the higher-order aberrations, it is necessary to do finite ray tracing, but the primary ray aberrations can be calculated by a paraxial ray trace. In optical systems with moderate to small apertures and fields, primary aberrations dominate. The wave aberration polynomial, W or OPD, is of the form

$$\begin{aligned} W = & W_{020} r^2 + W_{111} h r \cos \theta + W_{040} r^4 + W_{131} h r^3 \cos \theta + W_{222} h r^2 (\cos \theta)^2 \\ & + W_{220} h^2 r^2 + W_{311} h^3 r \cos \theta + \dots \text{(higher-order terms)} \end{aligned}$$

where h is the height of the object and r and θ are polar ray coordinates in the pupil (see Fig. 4.6).

It can be shown that the ray coordinates in the image plane relative to the perfect image coordinates are proportional to the partial derivatives of the wave aberration polynomial, that is,

$$\frac{\partial y'}{\partial y} \sim - \frac{\partial W}{\partial y} \quad \frac{\partial x'}{\partial x} \sim - \frac{\partial W}{\partial x}$$

This means that if the OPD or the wave aberration polynomial is known, the ray intersections in the image plane or spot diagrams can be easily calculated. The exponent of the pupil radius term is higher by one in the wave aberration polynomial than in the ray-intercept equations. Thus, for example, third-order spherical aberration affects the

Figure 4.6
Nomenclature for
Wave Aberration
Polynomial

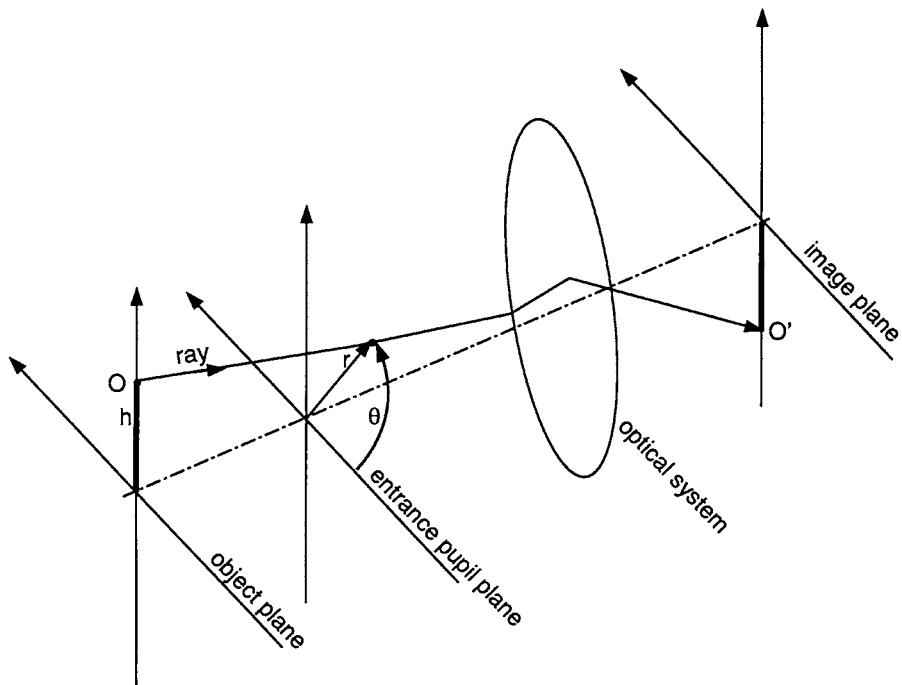


image blur diameter in proportion to the cube of the radius of the pupil, whereas the optical path difference is proportional to the fourth power of the pupil radius.

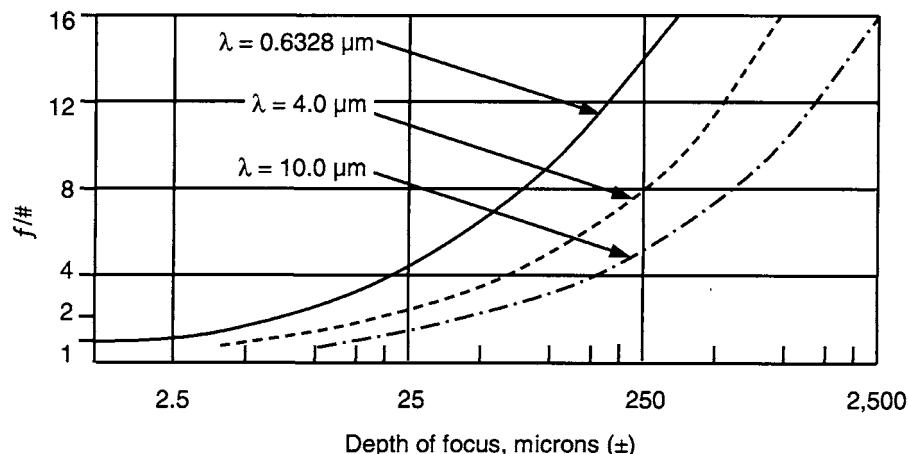
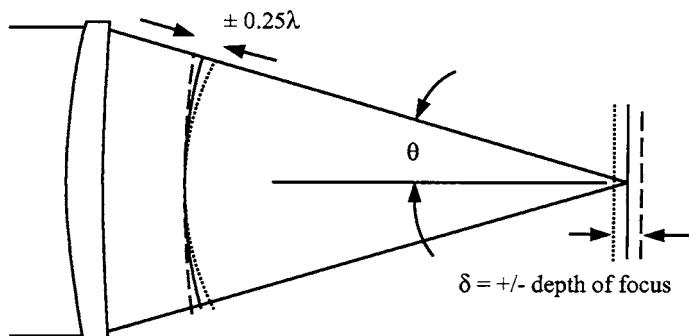
Depth of Focus

As we now know, if the optical path difference is less than or equal to $1/4\lambda$, our system meets the Rayleigh criteria and the system imagery is nearly indistinguishable from perfect. This result can be effectively used to determine just how much defocus is tolerable to maintain diffraction-limited performance.

Consider an otherwise perfect optical system, as shown in Fig. 4.7. The solid line in the upper part of Fig. 4.7 represents the nominally perfect spherical wavefront proceeding to the nominal image plane. If we now locate a compass point displaced fore and aft of the nominal image plane and draw two circles which touch the nominal wavefront on the

The Concept of Optical Path Difference

Figure 4.7
Depth of Focus



optical axis, these circles will depart from the nominal wavefront along the limiting marginal ray by an amount which is, in effect, the optical path difference. We now adjust the compass point until this displacement from the nominal wavefront is ± 0.25 wave. This yields the image plane locations which correspond to one-quarter wave of defocus. The depth of focus which corresponds to an OPD of $\pm 1/4 \lambda$ is

$$\delta = \pm \lambda / (2 n \sin^2 \theta) = \pm 2 \lambda (f/\#)^2$$

An extremely useful rule of thumb is that the depth of focus in the visible is approximately $\pm (f/\#)^2$, in micrometers. Thus, for an $f/4$ lens in the visible the depth of focus is approximately $\pm 16 \mu\text{m}$. For an $f/2$ system

the depth of focus is approximately $\pm 4 \mu\text{m}$, and so on. In the lower portion of Fig. 4.7, we show the depth of focus for systems in the visible, the medium-wave infrared (3 to $5 \mu\text{m}$), and the long-wave infrared (8 to $12 \mu\text{m}$), respectively. This is shown as a function of $f/\#$. It will become very apparent that as the $f/\#$ and wavelength increase, the depth of focus increases as well. This increase is linear with wavelength and quadratic with $f/\#$.

Do keep in mind that this assumes an otherwise perfect system. If your lens system has some inherent aberrations and/or wavefront errors due to design or manufacturing errors, then it will not be nominally perfect to begin with and you may not be able to allow a full one-quarter wave of depth of focus in image location or defocus before you degrade the performance by the quarter-wave limit.

There is another term, “depth of field,” which is often confused with depth of focus. Both terms are defined here, which should dispel any further confusion.

Depth of focus. The amount of image defocus which corresponds to being out of focus by one-quarter wave. This means that the optical path difference between the real wavefront leaving the exit pupil at its outer periphery and a reference wavefront centered at the nominal image plane is one-quarter of the wavelength of light.

Depth of field. This is a term used mostly in photography. What it means is that if you focus a camera at a given distance or range, how much further from the camera and closer to the camera than this distance will objects be in acceptable focus. This is analogous to depth of focus; however, it is not as stringent, and it is directly related to how acceptable the image looks to the eye.

CHAPTER **5**

Review of Specific Geometrical Aberrations and How to Get Rid of Them

As discussed in Chap. 3, aberrations are the failure of the optical system to produce a perfect or point image from a point object. The geometry of focusing light using spherical surfaces is simply not perfect, and spherical surfaces are used primarily due to their inherent ease of manufacturing. Many lenses can be ground and polished at the same time, as was shown in Fig. 3.2. Lenses are blocked together on the rotating part of the machine called a “block.” The top part, which is called a tool, has the desired radius of curvature, and it moves back and forth as the block rotates, forming spherical surfaces of the same radius on all lenses.

As was discussed, paraxial optics applies Snell’s law using the small angle approximation where the sine of the angles of incidence on surfaces is equal to the angle, in radians. In paraxial optics, there are no aberrations whatsoever, and by definition, the image of a point object is a perfect point image. Aberrations occur because in a real system the angles of incidence are nearly always so large that the paraxial

approximation is invalid and this causes the rays not to converge to a single point image.

As will be discussed in Chap. 7, the use of nonspherical, or aspheric, surfaces can often help significantly in minimizing, if not eliminating, aberrations. It is important to note that the use of aspherics does not automatically guarantee that the aberrations will be zero; in fact, for the most part, this will not happen. Their use is yet another technique for minimizing and balancing aberrations. There are techniques for manufacturing aspheric surfaces or aspheric lenses such as injection molding of plastic lenses, compression molding of glass, or diamond-turning aspheric surfaces in plastic, some crystals, or metal. Aspheric surfaces are used for additional aberration correction, but for the most part, spherical surfaces are used in optical systems. Aspheric optical components are often expensive, such as diamond-turned surfaces and glass-molded lenses, and not sufficiently accurate or unstable with a change of temperature such as plastic lenses.

There is a class of small lenses used in optical storage applications such as CD read lenses where aspherics are mandatory. These lenses are about the diameter of an aspirin tablet and are compression molded or manufactured by other techniques. In addition, many of today's digital cameras contain very small lenses (less than 6 to 8 mm in diameter) and glass aspherics are becoming more common in these application areas.

The index of refraction of the glass and other transmitting materials is used for making lenses changes with the wavelength of light, a phenomena called *dispersion*. The result is aberrations which change as a function of the wavelength. These aberrations are called *chromatic aberrations*. The image of a point is a superposition of the images for the entire wavelength band or spectral range, each of them blurred with the presence of monochromatic aberrations.

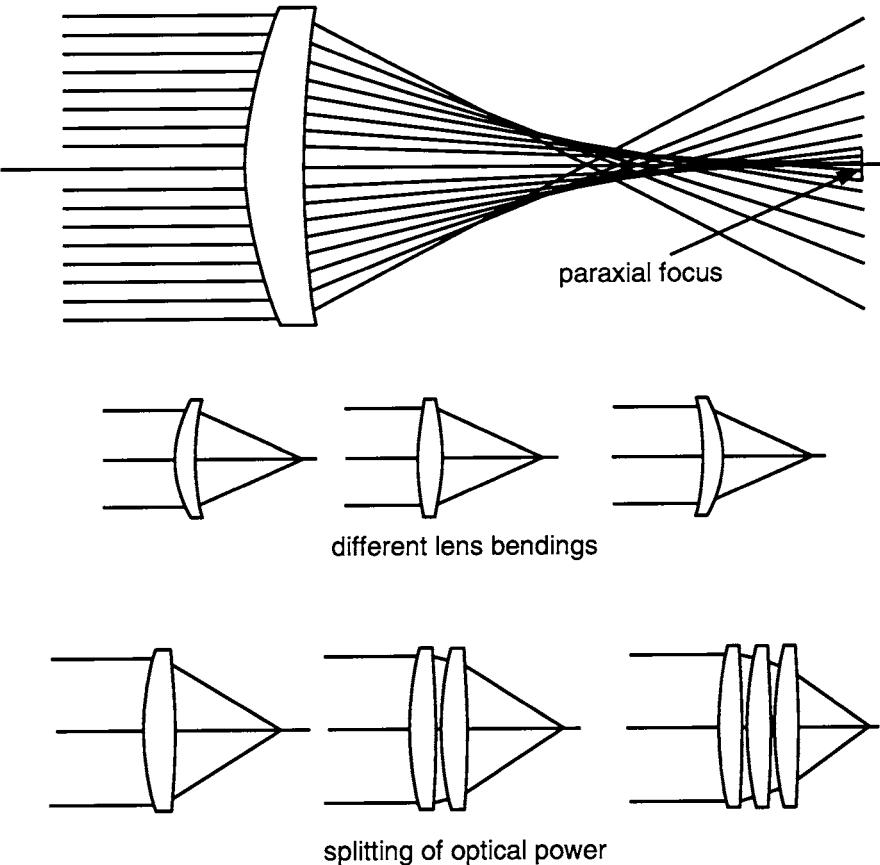
With a well-chosen combination of optical parameters such as lens shapes, number of optical elements, and different optical materials, aberrations in real optical systems with large ray angles can be reduced to a minimum or may be able to be eliminated to the level of the diffraction limit.

Spherical Aberration

If light is incident on the single lens shown in Fig. 5.1, rays that are infinitely close to the optical axis will come to focus at the paraxial image position. As the ray height above the optical axis at the lens

Review of Specific Geometrical Aberrations

Figure 5.1
Spherical Aberration



increases, the rays in image space cross the axis or focus closer and closer to the lens. This variation of focus position with aperture is called *spherical aberration*.

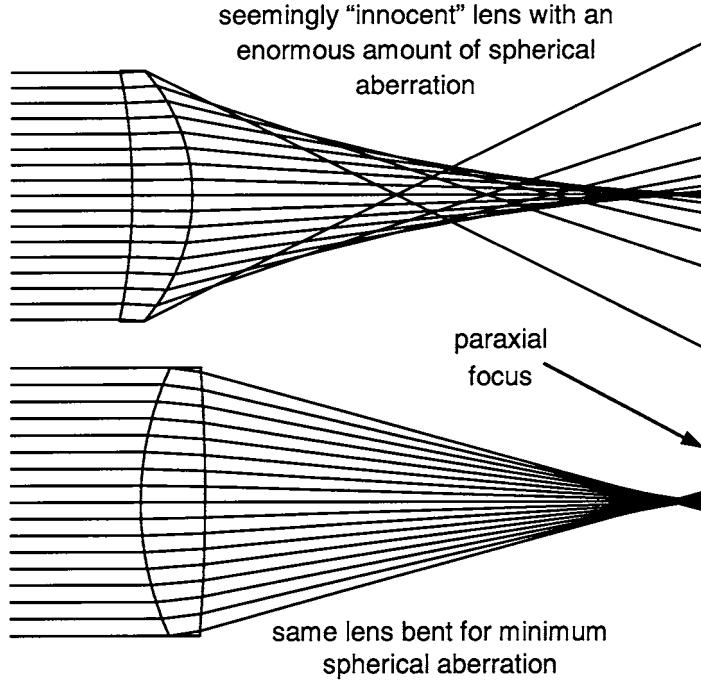
The magnitude of this spherical aberration depends on the height of the ray in the entrance pupil. The amount of spherical aberration is proportional to the cube of ray height incident onto the lens. If the spherical aberration is measured along the optical axis, it is called *longitudinal spherical aberration*. More often, it is measured as a lateral or transverse aberration, and it represents the image blur radius. For a given focal length lens, a lens with twice the diameter will have eight times larger image blur. For a given focal length and aperture of a single lens, spherical aberration is a function of the object distance and bending (shape) of the lens.

Also shown in Fig. 5.1 are lenses of different bending. The meaning of the term “lens bending” is that the focal length and hence the power of the lens is maintained while changing the radii of both surfaces. This would be the same as physically bending a lens made of flexible plastic. Spherical aberration is highly dependent on the relative lens bending, as will be discussed later.

Another powerful method of controlling spherical aberration is by splitting the optical power into more elements, as shown in the lower portion of Fig. 5.1. By splitting the optical power among several elements, the angles of incidence on each surface can be reduced, resulting in reduced aberrations. As we learned earlier in Chap. 3, reducing the angle of incidence results in a smaller deviation between paraxial rays and real rays, and hence reduced aberrations.

Consider Fig. 5.2 where we show a single $f/2$ lens element with an enormous amount of spherical aberration. The lower part of Fig. 5.2 also shows an $f/2$ lens; however, in this case the lens is bent for minimum spherical aberration.

Figure 5.2
Spherical Aberration
as a Function of Lens
Bending



Review of Specific Geometrical Aberrations

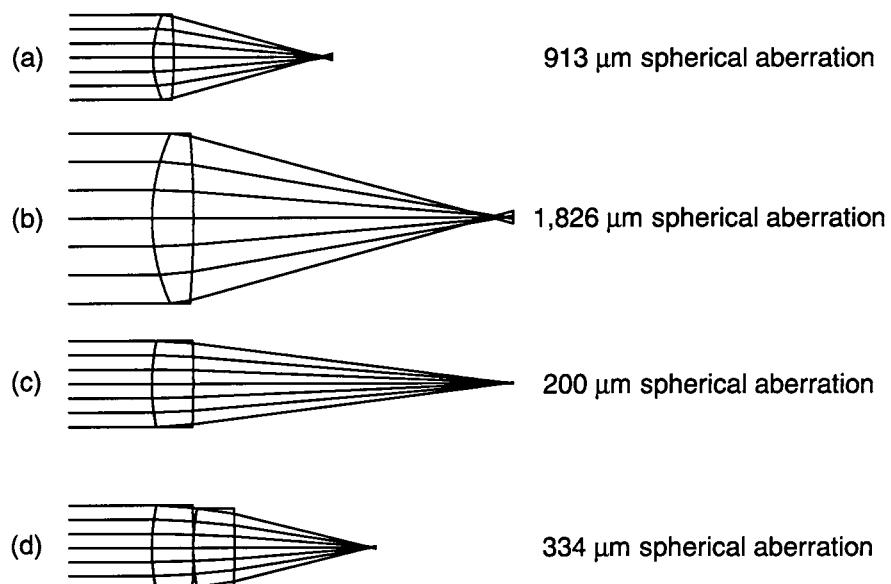
When the object point is at a finite distance, the shape of the lens changes for minimum spherical aberration. In a symmetrical case, when the distance of the object point from the lens is the same as the distance of the image, an equiconvex lens is the bending which produces minimum spherical aberration.

Let's look further into reducing the spherical aberration by splitting a lens into several elements. The resulting lens will perform the same function, keeping the total optical power of the elements the same as of the original lens.

We will demonstrate in a simple way how spherical aberration can be reduced by a factor of 2, if the lens is split into two lenses. We will do this in several logical steps:

1. The first step is to start with a lens bent for minimum spherical aberration, and this is shown in Fig. 5.3a for a 25-mm-diameter $f/2$ lens of BK7 glass. The residual spherical aberration is 913 μm from real ray tracing.
2. We now scale the lens up by a factor of 2, as shown in Fig. 5.3b. The focal length of our new lens is twice as large, the diameter is twice as large, and the spherical aberration is also twice as large. Note that when we scale a lens, all parameters with units of length scale by the same factor such as the radii and thickness.

Figure 5.3
Splitting Optical Power to Reduce Spherical Aberration

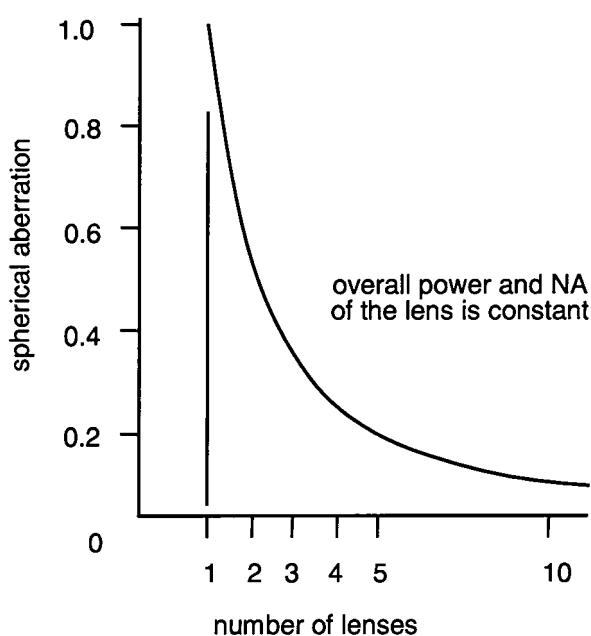


The refractive index is unitless, and thus remains unchanged. The spherical aberration is now doubled to 1826 μm .

3. We now reduce the aperture of this scaled-up lens by a factor of 2, as shown in Fig. 5.3c. The spherical aberration reduces by the cube of the aperture, which means by a factor of 8, which is 228 μm . Real ray tracing gives 200 μm , which is quite close. Now we have a lens with approximately four times less spherical aberration than the starting point. This new lens has the same aperture as the starting lens, but its focal length is twice as large.
4. Now we add one more identical lens of the same power (same focal length), as in Fig. 5.3d. The spherical aberration is doubled (approximately), but it is twice as small as the aberration of the starting lens. The real ray tracing shows that our final solution has 334 μm of spherical aberration, which is 36% of the starting value with a single element of the same focal length.

The new configuration consists of two lenses performing the same function as the single starting lens, but having one-half of the spherical aberration. The theoretical result of splitting a single lens into multiple lenses is shown in Fig. 5.4. This result shows that if we split an element

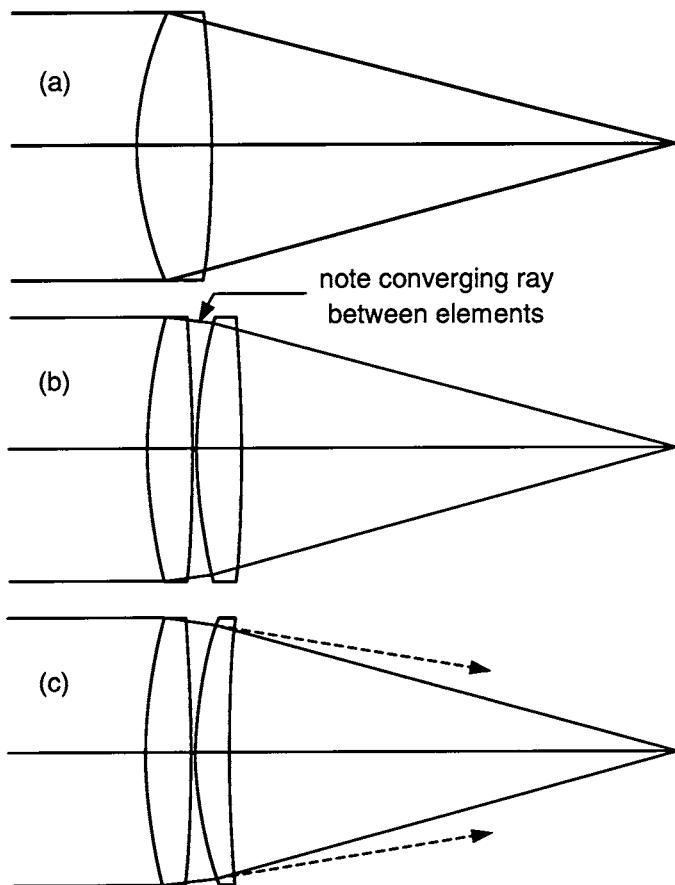
Figure 5.4
Spherical Aberration
as a Function of
Number of Elements



Review of Specific Geometrical Aberrations

into four to five elements, the spherical aberration will reduce to about 10 to 15% of the single-element starting point. As it turns out, when we split power in a real lens, the results are significantly better. This is because the light exiting the first element will be converging, and if the second element is now bent for minimum spherical aberration based on converging incident light, the resulting spherical aberration is reduced even further. This way, by introducing even more optical elements, the spherical aberration can be reduced significantly. Figure 5.5 shows the situation. Figure 5.5a represents a single element bent for minimum spherical aberration. Figure 5.5b shows two identical elements as derived previously. Note that the light between the two elements is converging as it enters the second element. Figure 5.5c shows how the second element

Figure 5.5
Illustration of How to
Achieve a Further
Reduction of Spherical
Aberration

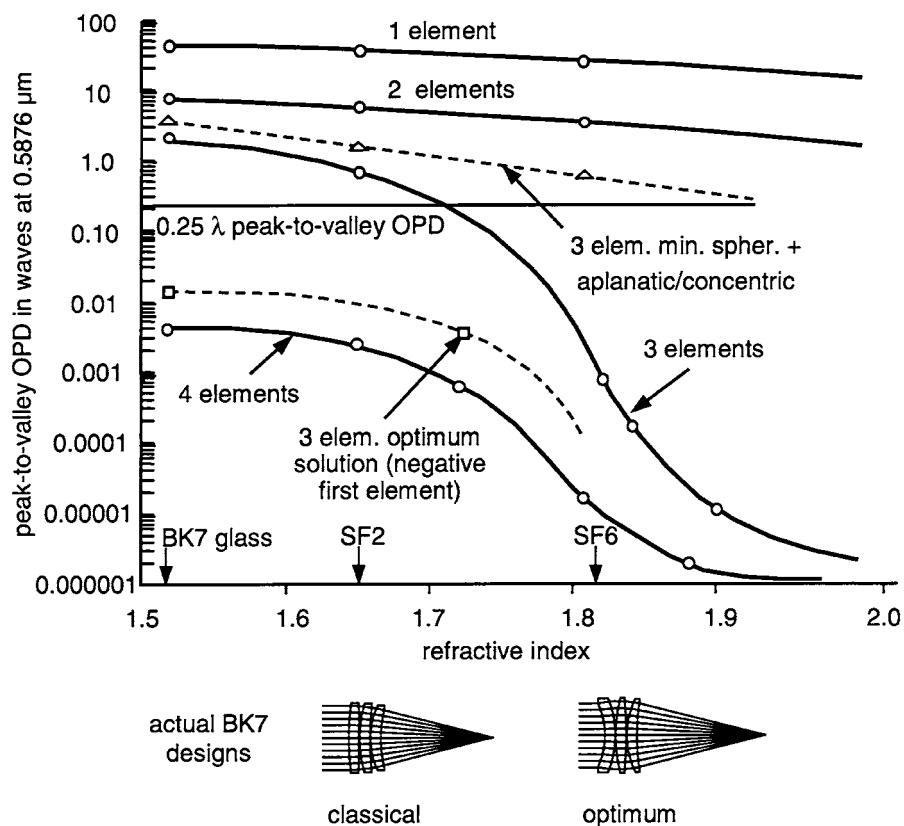


can more optimally “curl” or bend more strongly so as to minimize the angles of incidence onto its surfaces, thereby reducing the spherical aberration from the design in Fig. 5.5b with the identical elements.

It is instructive to consider the design of an $f/2$, 100-mm focal length lens for minimum spherical aberration with one, two, three, and four components, and we will do this for glasses with refractive indices ranging from 1.52 to 1.95. We will now plot the peak-to-valley optical path difference for all of these cases in Fig. 5.6.

The results are quite dramatic. Note that for a single $f/2$ element with a 100-mm focal length of BK7 glass (refractive index 1.517) the spherical aberration is approximately 40 to 50 waves P-V. Splitting the element into two elements reduces the OPD to about 6 to 8 waves, and splitting it into three elements further reduces it to about 2 waves. And four elements results in about 0.004 wave, a significant reduction. There

Figure 5.6
Spherical Aberration
as a Function of
Number of Elements
and Refractive Index



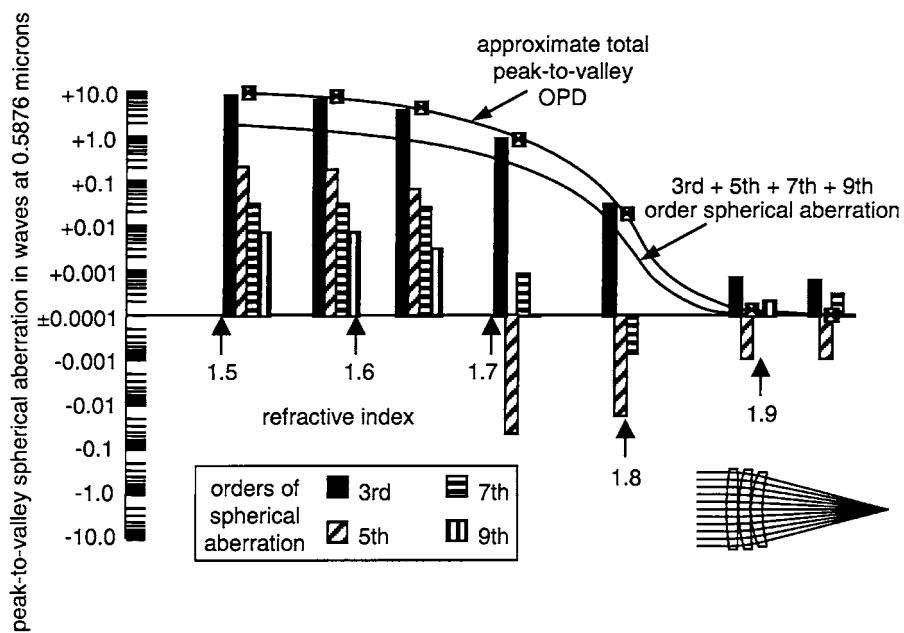
is a further reduction in OPD as the refractive index is increased, especially for three and four elements where nearly six orders of magnitude reduction in OPD is achieved by simply increasing the refractive index from 1.5 to 1.9!

This is for what we have termed the “classical” solution. This is where each element is bent somewhat more than its predecessor in order to minimize the angles of incidence and thus the overall spherical aberration. As we will see later, there is a configuration which yields an even better solution, and we call this the “optimum” configuration. It is characterized by a negatively powered meniscus first element, and it yields two orders of magnitude less OPD than the classical solution, even at the lower refractive index region.

Note that the preceding parametric analysis is based on monochromatic light and was computed only on axis. While this is somewhat of an idealized situation, the insight we have gained into aberration reduction is of major significance and it further enhances our understanding of aberrations and where they come from. However, from the analysis thus far we really do not know just why such a dramatic reduction in the aberration is achieved.

Spherical aberration terms in the wave aberration polynomial are the fourth, sixth, eighth, etc., order in terms of the pupil radius. The exponent of the pupil radius term is larger by one in the wave aberration polynomial than in the ray-intercept equations. When we talk about spherical aberration image blur size in the image plane, we talk about third, fifth, seventh, ninth, etc., order in terms of the pupil radius. Let us again look at three component lenses optimized for the smallest spherical aberration, and compare the lenses from low-index glass $n = 1.5$, and then increase the index up to $n = 2$. The spherical aberration as a function of the index of refraction is shown in Fig. 5.7. The contribution to the third-, fifth-, seventh-, and ninth-order spherical aberration is shown for refractive indices ranging from 1.5 to 2.0. Generally, lower orders of aberration have higher values, and they are predominant in the polynomial. As the index increases to somewhere around $n = 1.7$, the fifth-order spherical aberration changes sign and starts to balance the third-order aberration, so that the overall spherical aberration has a significant drop. Although the spherical aberration changes a lot with the change of the index of the components, there is only an imperceptible change in the shape of the lenses. The surfaces become a little shallower, but the overall shape of the lens remains the same.

Figure 5.7
Third-, Fifth-, Seventh-,
and Ninth-Order
Spherical Aberration
Versus Refractive
Index for Three
Elements



As a final illustration of what is happening, consider Fig. 5.8 where we show the classical solutions for refractive indices from approximately 1.5 to 1.7. The graphical data is the deviation of the wavefront from perfect as a function of the normalized pupil radius. Note that the OPD decreases from about 2 waves P-V to about 0.25 wave. Figure 5.9 shows the data for refractive indices ranging from 1.8 to 1.95, and we see that the OPD reduces from 0.002 wave to several ten-thousandths of a wave.

In all of these examples the relative shapes of the elements has remained nearly constant.

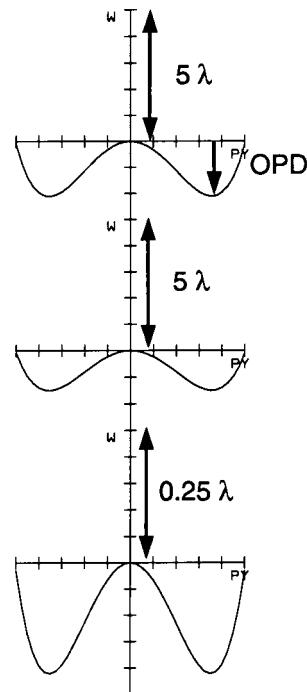
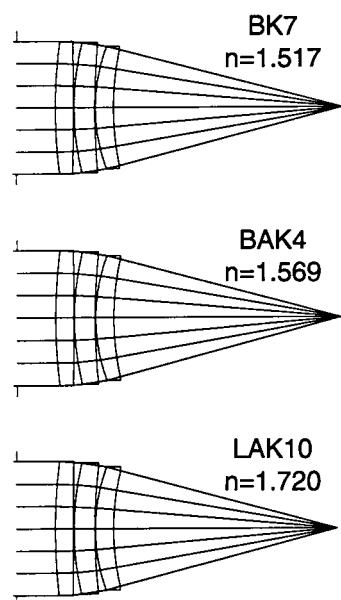
We noted earlier that there is a more optimum solution, and we show the three-element “classical” and optimum solutions along with the plot of optical path difference in Fig. 5.10. We are able to reduce the P-V OPD from 2 waves to less than 0.007 wave by changing the configuration. Both of these configurations use BK7 glass. The reduction in aberration is due to balancing of the fifth-order spherical aberration against the third-order, as described previously.

We have shown in the above material the different orders of spherical aberration, and how we can effectively and efficiently reduce the aberration by splitting the optical power into multiple elements, with the reduction of the angles of incidence serving as the primary means for

Review of Specific Geometrical Aberrations

Figure 5.8

Lens Configuration and Plot of Optical Path Difference for Optimized Lenses of Refractive Index 1.517 to 1.720



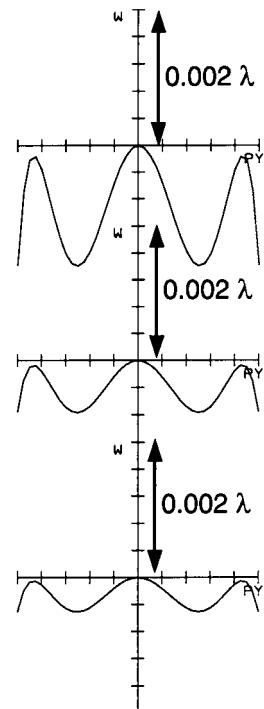
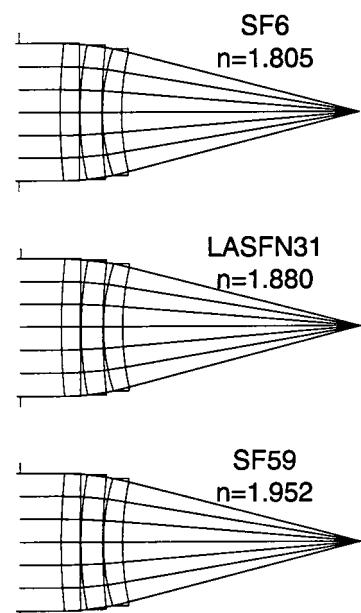
lower aberration. The lower the angle of incidence on a surface, the closer we are to the paraxial region where $n \times \theta = n' \times \theta'$ and the aberrations approach zero. There is a very interesting and useful tool called a "Pagel diagram" as shown in Fig. 5.11. The Pagel diagram is a plot, typically in the form of a bar chart showing the contribution of each surface within a lens system to the third order spherical aberration (other primary aberrations for a system with a finite field of view). Note how in this case (the classical solution) the contribution from the six surfaces are approximately +0.007, +0.007, +0.004, +0.004, -0.004, and +0.001, respectively. The sum of these contributions is +0.019, as shown.

It is shown in Fig. 5.12 how we can achieve a significantly higher level of performance by changing the first element to a negatively powered meniscus shape. Indeed, the negative contribution of the first surface balances nearly perfectly the positive contributions of the other five surfaces, for a net spherical aberration of nearly zero.

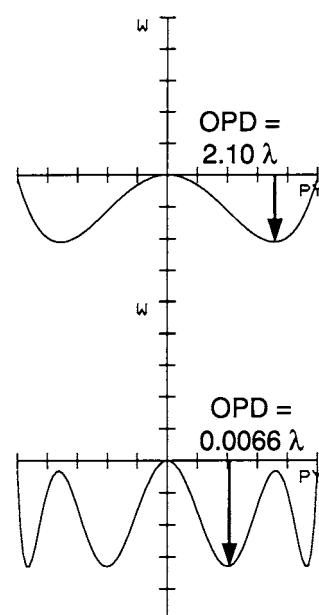
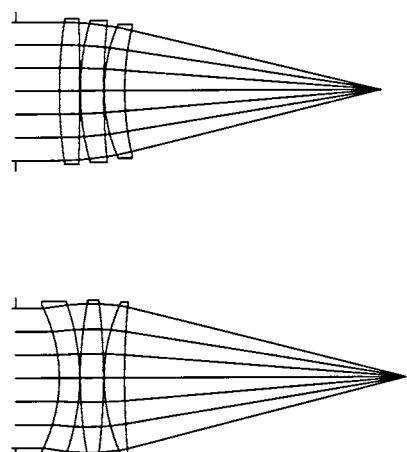
We have discussed orders of aberration in several contexts thus far. By carefully evaluating the plot of optical path difference, you can actually see visually the different orders. Consider Fig. 5.13 where we show again the

Figure 5.9

Lens Configuration and Plot of Optical Path Difference for Optimized Lenses of Refractive Index 1.805 to 1.952

**Figure 5.10**

"Classical" and Optimum Solutions with Three Elements for BK7 Glass



Review of Specific Geometrical Aberrations

Figure 5.11
Pagel Diagram of
3 Element "Classical"
Solution Showing
Surface Contributions
to Spherical Aberration

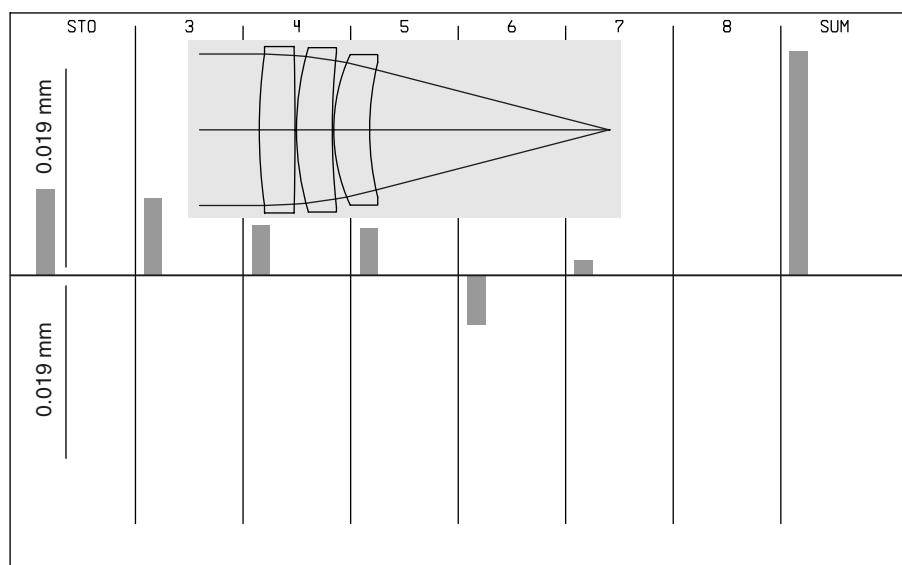


Figure 5.12
Pagel Diagram of
Optimum Solution

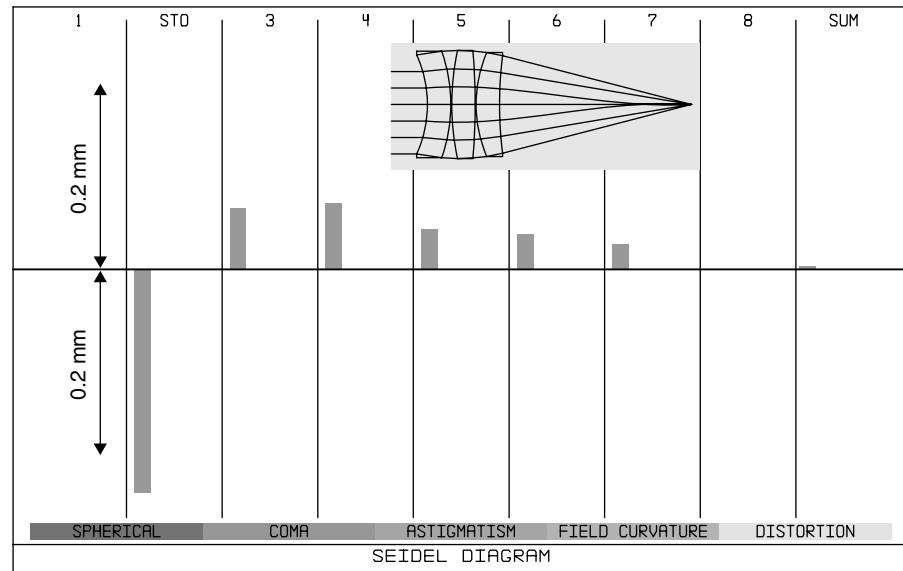
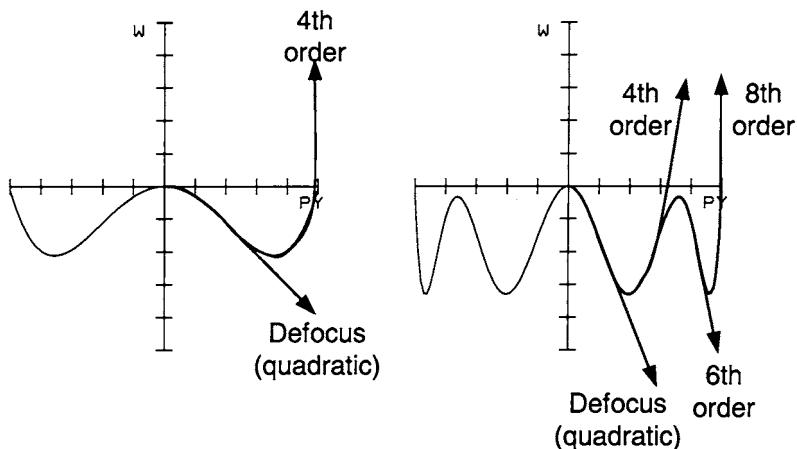


Figure 5.13
Illustration of the
Orders of Wavefront
Aberration



plots of OPD for the “classical” and optimum designs just discussed. However, here we show each of the orders of spherical aberration. Recall that the exponent is one higher than the transverse ray aberration polynomial, so the third-order spherical aberration is proportional to the fourth-order in OPD and so on. Of course, focus shift is thus quadratic in OPD as we would expect. Note that each time we see an inflection in the data this is equivalent to another order being added. There will almost, by definition, be higher orders than those shown; however, these data show clearly the presence of the different orders and how they tend to balance each other.

Coma

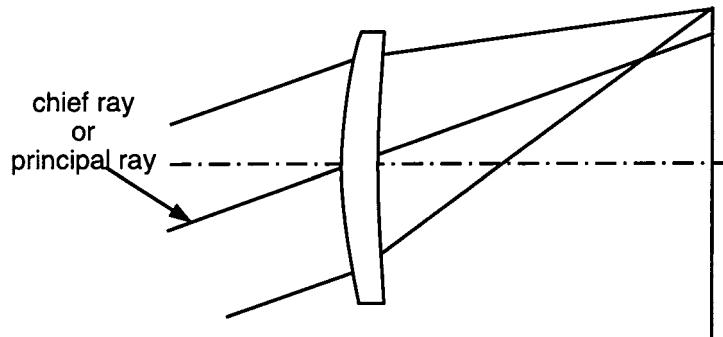
When we move away from the optical axis in field of view, the image of a point becomes nonrotationally symmetric. In Fig. 5.14 parallel rays come from an infinitely distant point which does not lie on the optical axis of the lens. They enter the lens at an angle, and they are focused by the lens to a certain height from the optical axis, defined by the field angle and the focal length of the lens. If the lens itself limits the bundles of rays from different points in the field, we say that the aperture stop is located on the lens. Rays that go through the center of the aperture stop are called *chief rays*. There is only one chief ray for each point in the object.

Rays that go through the aperture stop and lie in the plane of the drawing are called *meridional rays*. Rays which do not lie in the meridional

Review of Specific Geometrical Aberrations

Figure 5.14

Coma



plane are called *skew rays*. A plane perpendicular to the drawing in which lies the chief ray is called the *sagittal plane*. The meridional and sagittal planes have one common ray, the chief ray.

In an optical system *coma* is defined as the variation of magnification with aperture. Rays that transmit through the lens through different portions of the aperture stop cross the image plane at different heights from the optical axis. In the case of the single positive lens shown in Fig. 5.14, a ray passing through the top and bottom edge portions of the lens converges to a point in the image plane which is further from the optical axis than the point of convergence of other skew rays.

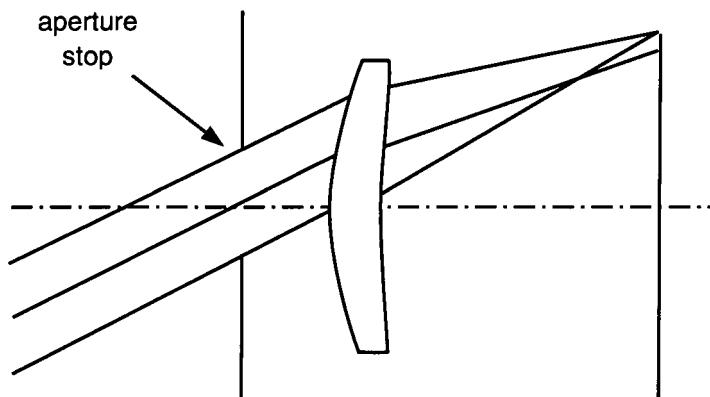
The shape of the image of a point as formed by a system with coma has the shape of a comet. The height of the image is usually defined by the position of the chief ray on the image plane. In the presence of coma, most of the light energy is concentrated in the vicinity of the chief ray. Coma is linearly proportional to the field of view and proportional to the square of the aperture.

When the aperture stop is not on the lens, moving the position of the stop can control the coma. Having greater ray symmetry on the way through the lens about the aperture stop reduces the coma. Figures 5.15 and 5.16 show the aperture stop to the left and the right of the lens, respectively, and it is clear that the off-axis ray bundles have a higher degree of symmetry and hence significantly reduced coma when the stop is to the right of the lens. This is due to the greater symmetry on the first surface, which results in reduced angles of incidence and hence reduced aberration.

We can best understand coma by looking at Fig. 5.17, which shows the cause of coma. In the top part of the Fig. 5.17, we show a collimated bundle of light incident obliquely onto a convex surface (we will only consider here the first surface of the lens). Note that the entire bundle is displaced

Figure 5.15

Coma with Stop in Front of Lens

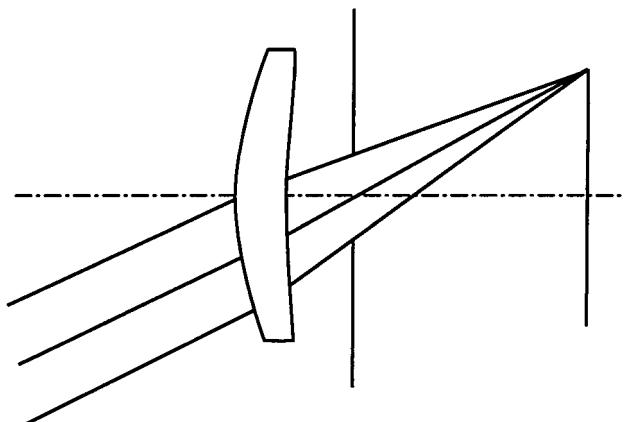


from the normal to the surface, which is shown as a dashed line. Note also that the upper ray is incident onto the lens surface at a very high angle of incidence with respect to the surface normal, and as we know, this results in a significant ray bending or angle of refraction. The angles of incidence decrease rapidly as we transition to the lower portion of the ray bundle. The coma formation in this situation is quite evident.

Consider now the lower portion of Fig. 5.17, where we show a light bundle whose central or chief ray is normal to the surface. Now we have greatly reduced angles of incidence and, furthermore, the upper and lower limiting rays are symmetrical with each other. The net effect is that there is no coma contribution from this surface whatsoever, and the residual aberration is the same as spherical aberration.

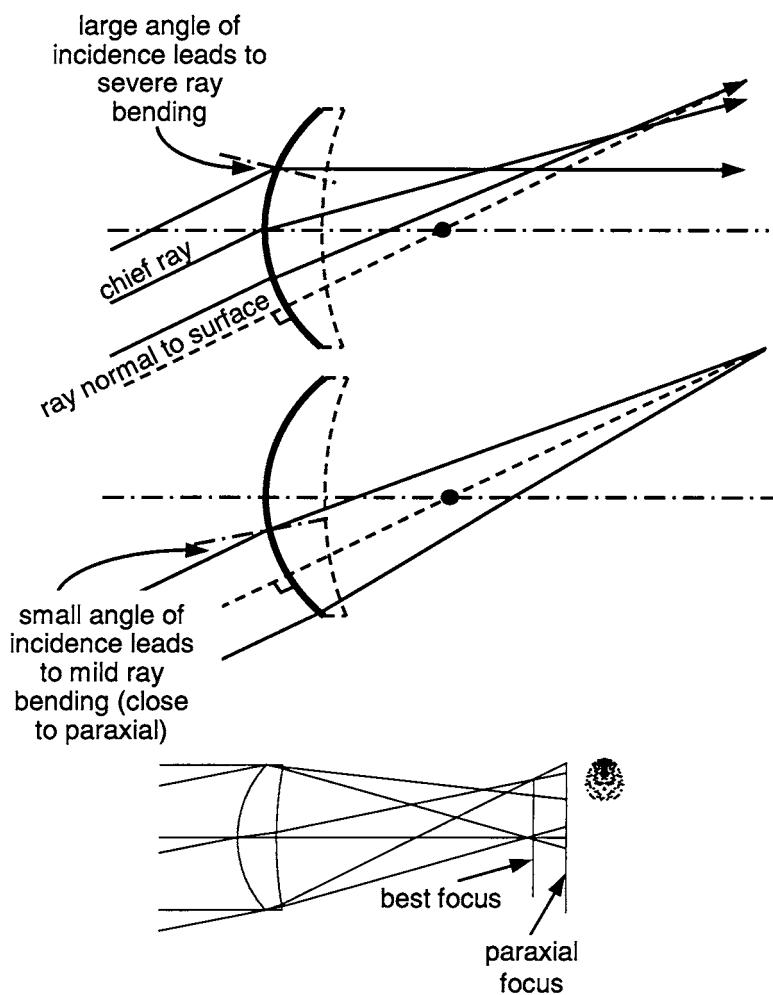
Figure 5.16

Reduced Coma with Stop Aft of Lens



Review of Specific Geometrical Aberrations

Figure 5.17
Where Does Coma Come From?

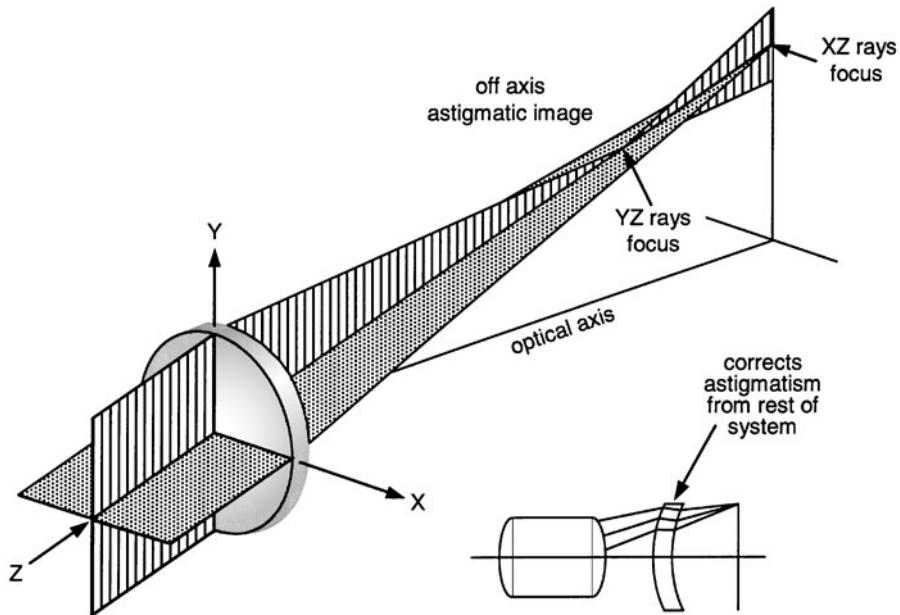


Astigmatism

In the presence of astigmatism, rays in the meridional and sagittal planes are not focused at the same distance from the lens. An astigmatic image formed by a positive lens is shown in Fig. 5.18.

Rays in the meridional plane focus along the line that is perpendicular to the meridional plane. Rays in the sagittal plane are focused further away from the lens, along the line perpendicular to the sagittal plane. Between the astigmatic foci, the image of a point is blurred. It takes the

Figure 5.18
Astigmatism



shape of an ellipse or circle. The smallest size of the image blur is half-way between two astigmatic foci when it is circular. Astigmatism is linearly proportional to the lens aperture and to the square of the field angle.

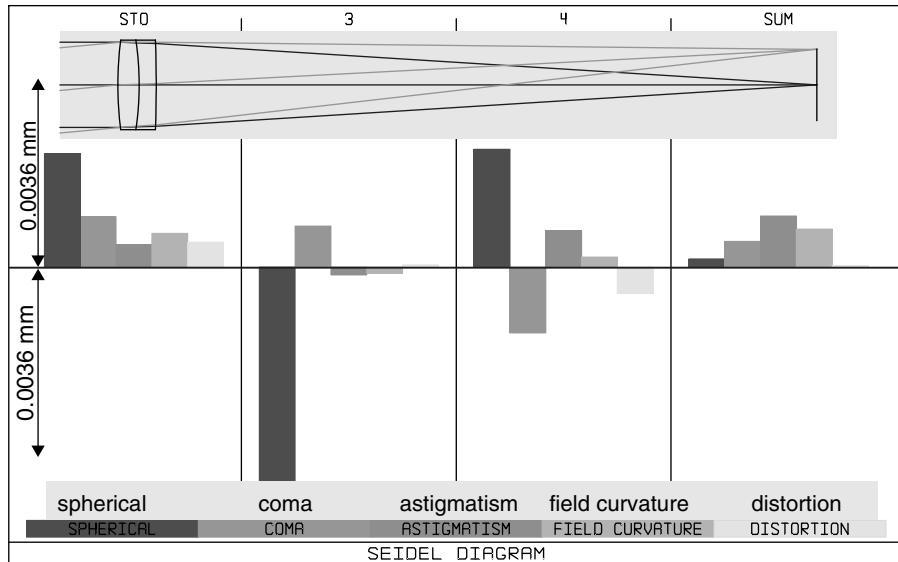
Astigmatism can be controlled by changing the shape of the lens and its distance from the aperture stop, which limits the size and position of the bundle of rays passing through the lens.

A tilted plate in a converging cone of light introduces astigmatism. A weak meniscus lens close to the image plane acts similar to a tilted plate with a tilt angle which changes from zero on axis to a certain angle at the edge of the field. This way, astigmatism created by the meniscus can partly or completely cancel the astigmatism of the rest of the optical system. Also shown in Fig. 5.18 is an example showing just how this works.

To show clearly how this works, Fig. 5.19 shows a Pagel diagram of an achromatic doublet with a small finite field of view. The astigmatism is the largest residual aberration. We now will add a weak meniscus element and reoptimize the design, with the result shown in Fig. 5.20. It is clear that the astigmatism at the image is virtually zero with the balancing of the aberration handled mostly by the meniscus element. Remember that this element form is analogous to a varying tilted plate in a converging cone as described previously.

Review of Specific Geometrical Aberrations

Figure 5.19
Pagel Diagram of
Achromatic Doublet



Where does astigmatism come from? An oblique cone of light incident on a lens is shown in Fig. 5.21. Assume that this element is immersed somewhere in the middle of an optical system. The area or footprint on the surface of the light cone shown extends over more of the surface in the y or tangential direction than in the x or sagittal

Figure 5.20
Pagel Diagram of
Field Corrected
Achromatic Doublet

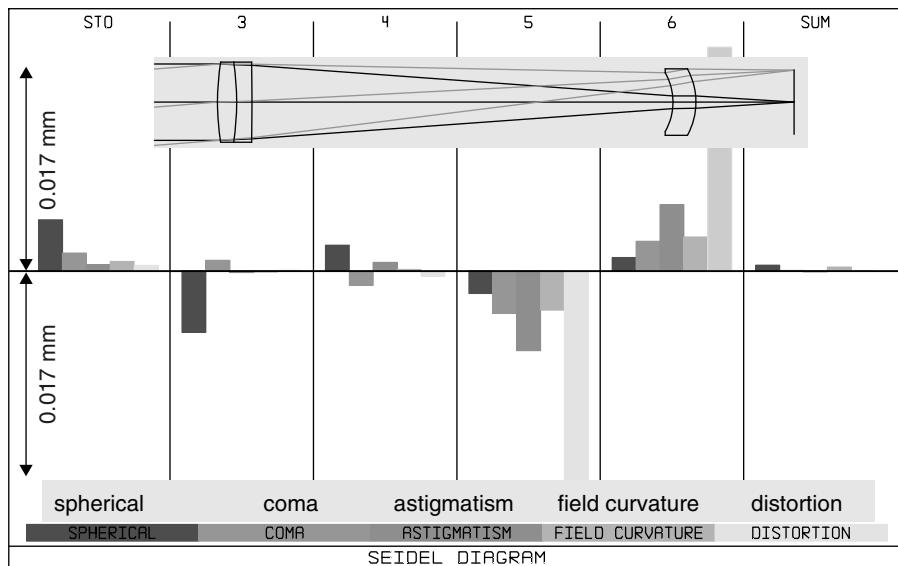
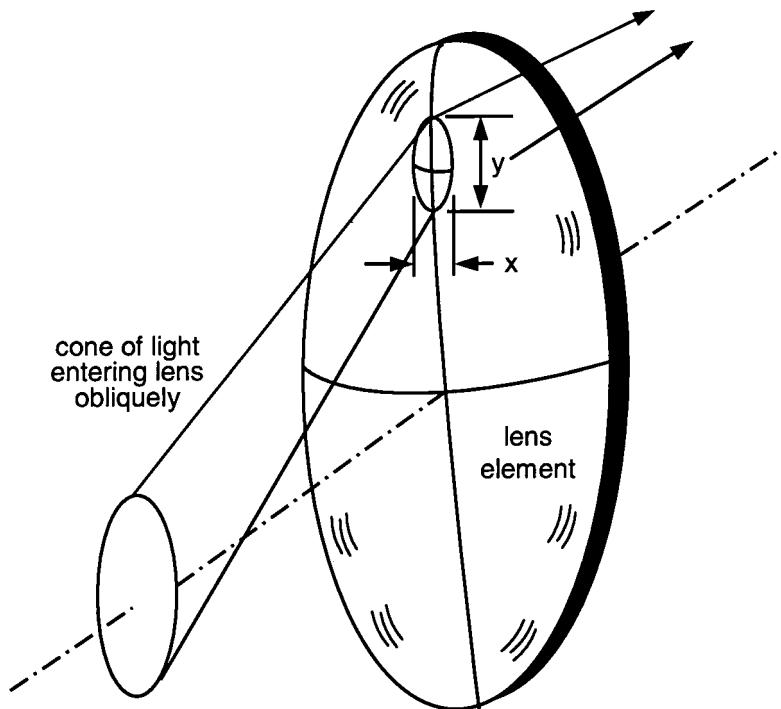


Figure 5.21
Where Does Astigmatism Come From?



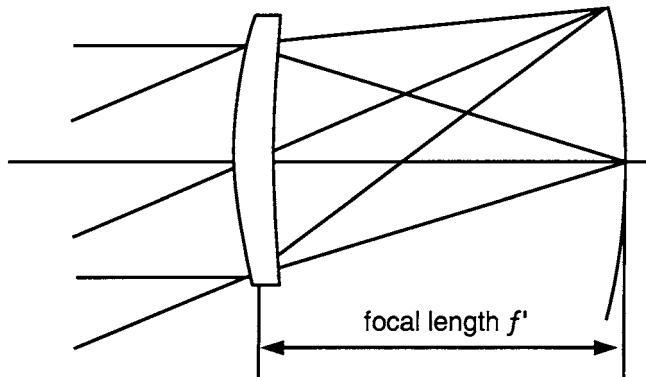
direction. Recall that the rate of change of slope of a sphere is constant everywhere on the sphere. Thus, the extreme tangential rays see a greater slope change on the surface than the extreme sagittal rays and are hence refracted at greater angles. This causes the tangential ray fan to focus closer to the lens than the sagittal ray fan, and this is astigmatism. As the surface is spherical, we will also find in many cases an off-axis form of spherical aberration called "tangential oblique spherical aberration" which is introduced in the tangential direction.

Field Curvature and the Role of Field Lenses

A positive lens forms an image on a curved surface, as shown in Fig. 5.22. In the absence of astigmatism, a surface on which the image is formed is called the Petzval surface. If a lens has no astigmatism, the sagittal and

Review of Specific Geometrical Aberrations

Figure 5.22
Field Curvature



tangential images coincide with each other, and they coincide with the Petzval surface.

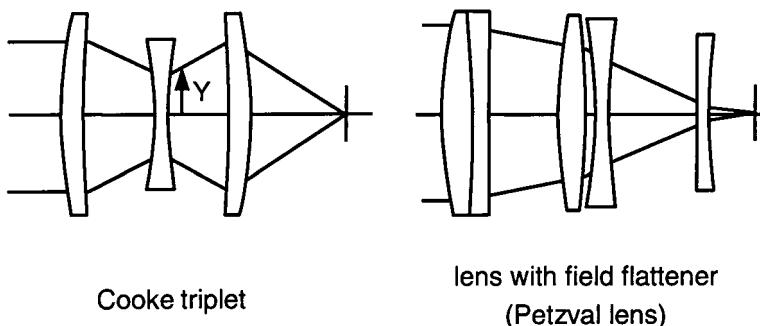
In the presence of astigmatism, both sagittal and tangential image planes are closer to the lens than the Petzval surface, and the tangential image is three times further from the Petzval surface than the sagittal image. The curvature of the Petzval image is inversely proportional to the product of the index of refraction of the lens and its focal length. If there are many components in the optical system, the resulting Petzval curvature is a sum of Petzval curvature contributions from all lenses.

We know that with a 35-mm camera, we can take nice sharp photographs using a flat film. Which method is then used in designing a lens to get a flat image plane? Since the contribution a lens element makes to the Petzval sum is proportional to its optical power, simply splitting of the elements will not change the field curvature. However, positively and negatively powered components can be combined to reduce the field curvature to zero. When negatively powered lenses are added to the system, the resulting power is also reduced.

Fortunately, there is a solution to this problem. The contribution a lens element makes to the system power is proportional to the product of its power and the height of the marginal ray which is the ray going through the edge of the aperture stop. This way, if the position of a negative lens in an optical system is suitably chosen so that its power is substantial, but the height of the marginal ray on the lens is relatively low, its contribution to the overall optical power is relatively low while still having a significant field curvature.

Two examples where a negative component is effectively used to reduce a field curvature are shown in Fig. 5.23. The first example is the

Figure 5.23
Negatively Powered
Elements with Small
Value of Y to Flatten
Field



Cooke triplet. A negative component is located in the middle between two positive lenses. The marginal ray height on both positive lenses is higher than on the negative lens. However, the power of the negative lens significantly reduces the field curvature created by the two positive lenses.

The second example is a Petzval lens, where the negative component is located very close to the image plane. Its contribution to the power of the whole lens is very small, since the height of the marginal ray is extremely small when the lens is close to the image. If the lens is placed at the image plane, it does not change the overall power of the system.

Is there any reason to put the lens at the image plane? Yes, indeed both positive- and negative-powered lenses are often located either in the image plane or very close to it. They are called *field lenses*.

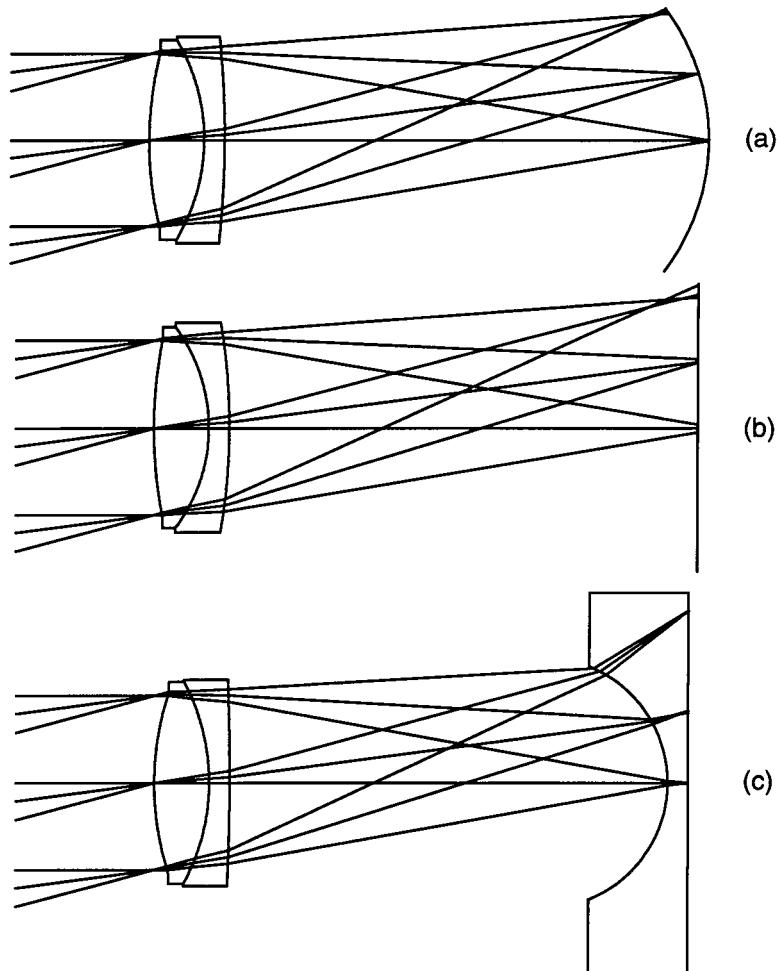
The first case when a lens is located at the image plane or just in front of the image plane is when the negative lens is used as a *field flattener* to correct the field curvature and flatten the field. This is common in complex wide field of view, fast (low $f/ \#$) lenses.

How does a negative lens flatten the field? Let us imagine the case of a simple positive lens that forms an image a certain distance from the lens. If we add a block of glass between the lens and the image plane, the image will move away from the lens. The image shift is proportional to the thickness of the glass block. In the case of the negative lens in front of the image plane, ray bundles that focus close to the edge of the field of view pass through the part of the lens where the glass thickness is larger than in the center of the lens. This way image points that are closer to the periphery of the image are shifted away from the focusing lens more than the ones in the center of the field. This results in a flatter image plane.

Review of Specific Geometrical Aberrations

A doublet lens forms a sharp image on a spherical surface shown in Fig. 5.24a. If an achromatic doublet is an objective lens of a telescope, which focuses the image on a reticle, or on a CCD detector, it is desirable to correct the field curvature of the lens. The reticle is usually engraved on a flat piece of glass, and in the case of a CCD detector, the sensitive area is always flat. When the CCD detector is adjusted for best focus, both the center and the edge of the field are slightly blurred, as in Fig. 5.24b, and the sharpest image is obtained for the intermediate field. Figure 5.24c shows how a field lens in front of the image reduces the field curvature and the image blur at the edge of the field.

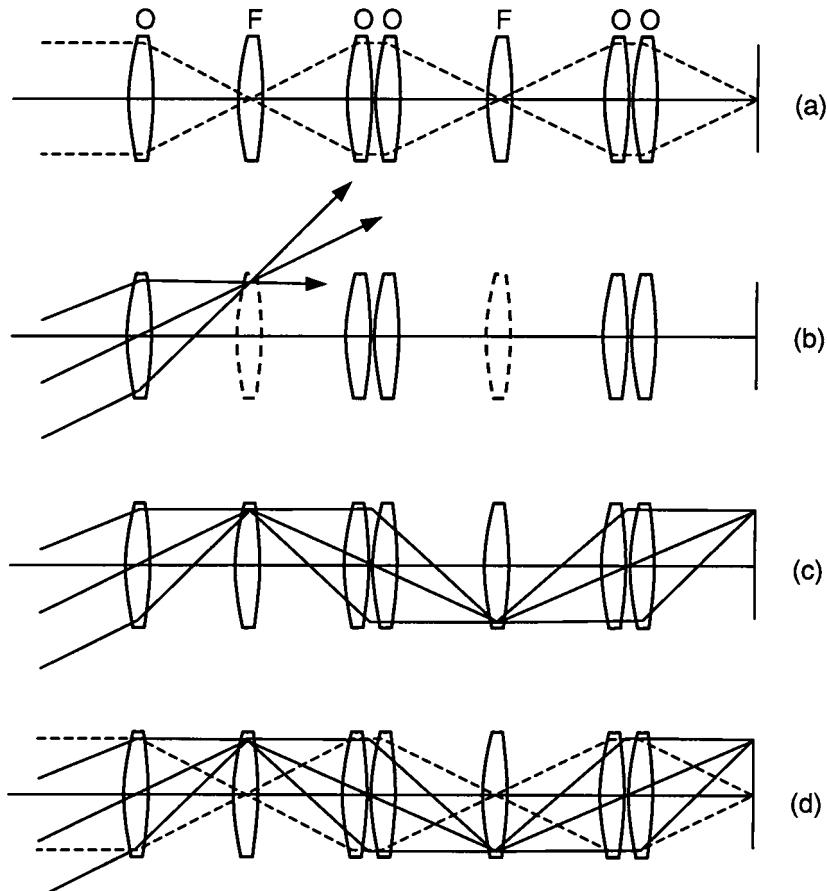
Figure 5.24
Field Flattener Example: How a Negative Lens Flattens the Field



The other types of field lenses are positive-powered lenses used in the systems with one or more intermediate images and relay optics. A submarine periscope is an example of such a system. The optics is inside a tube, which is 10 m long or even longer, but the diameter of the optics does not exceed 250 to 300 mm. Another example is an endoscope, which is on a much smaller scale, but with a similar ratio of the system length and diameter. Schematically, these systems are shown in Fig. 5.25a, where lenses labeled O are objective lenses and lenses labeled F are field lenses.

What is the function of these positive field lenses? They cannot correct the field curvature; they actually *increase* the field curvature already introduced by the other positive lenses. If we look at the axial beam shown in the schematic drawing, and assume from the beginning that

Figure 5.25
Field Lens Example:
How a Positive Lens
Reduces Vignetting

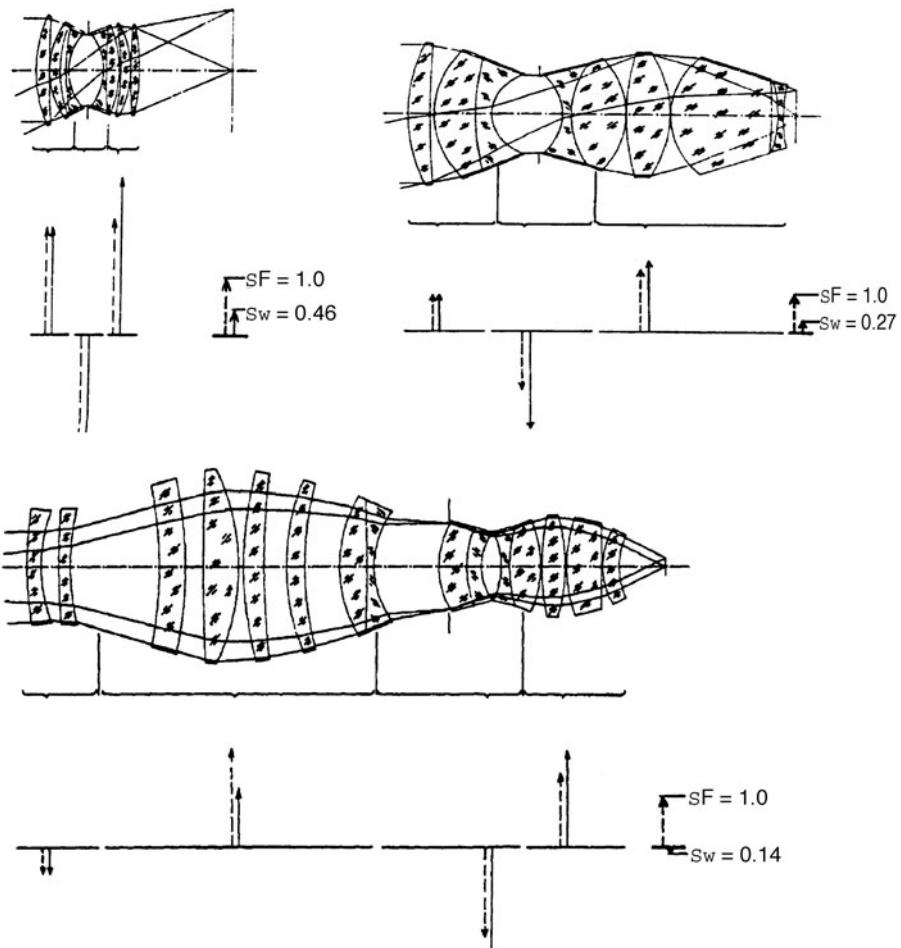


there are no field lenses (only the O lenses are present), the axial beam will be focused at the first intermediate image plane, then relayed with two lenses to the second intermediate plane, and finally relayed with another two lenses to the final image plane. There is no problem with the axial ray bundle. Now consider the beam entering the optical system at an angle. It will be focused at a certain height from the optical axis in the first intermediate image plane. However, the cone of light is so tilted that almost the whole ray bundle is going to miss the two relay lenses, and it will hit the housing. This is called *vignetting* in the optical system and it reduces the amount of light in the image periphery. If we now add a positive lens in the image plane, it will not do anything to the axial beam, but it will redirect the cone of light coming from the edge of the field into the relay lenses. There will be no vignetting and almost no change in the position and the size of the image. The image brightness is going to be uniform across the field, but the system will have a significant amount of field curvature. The primary purpose of these positive field lenses is to reduce or eliminate the vignetting in the system.

There was a paper in 1980 given by Erhard Glatzel of Carl Zeiss in Germany, where he talked about designing lenses in microlithography that imaged a mask onto a 50-mm-diameter silicon wafer at a 5:1 reduction ratio. Microlithography lenses are the most sophisticated lenses in our industry, since they have to resolve submicrometer structures in the flat image, as shown in Fig. 5.26. Glatzel starts from the basic relation for the Petzval sum. Suppose that only one type of glass is used in the entire lens. In order to correct the field curvature, the sum of the powers of all components in the system has to be zero or very close to zero. At the same time, total power of the system has to be positive. The total power is proportional to the sum of weighted powers of the components, where the weighting factor is the normalized height of the marginal ray on the component. Glatzel first analyzes a so-called planar camera lens similar to a double Gauss-type lens. The positive-powered components are the first two and the last two components. The negatively powered components are located in the center of the lens. This power distribution is very similar to a Cooke triplet lens. Below the lens layout, Glatzel plots the contribution to the power sum as well as the weighted power sum from corresponding lens groups above. Petzval curvature in this first lens is not reduced to zero but has a residual of 0.46.

In order to reduce the Petzval sum or field curvature, Glatzel forces the central elements to have a more negative optical power. The Petzval sum is lower than in the first example but it is still present. Unfortunately, the

Figure 5.26
Reduced Field Curvature Lenses as
Described by Glatzel
of Carl Zeiss



radii of the central elements have become quite strong or severe, thus increasing the angles of incidence on their surfaces (especially the short-radius inner concave surfaces), therefore introducing their own aberrations.

The final step in the design of the flat-field lithography lens is the addition of a negative group in the front of the whole lens. This group has a very important role. Because of its negative power, it slightly reduces field curvature, but it expands the beam and makes it possible for the following positive group to make a high contribution to the total power of the system. The final lens design has -0.14 of field curvature.

While this still has a noticeable result, the net aberration of the system is extremely well corrected. This form of lens still represents the state of the art in lithography optics.

While we are discussing this design, we should take a few minutes to look at the design form. After the initial two negative elements in the front, Glatzel uses four larger single elements to take the diverging light and converge it into the next group. He uses four elements in order to split the power and minimize angles of incidence so as to reduce aberrations. However, note that the second element seems to have most of the positive power, also the first element of this group seems to be quite concentric about the diverging light cone, and it seems to have little or no optical power. We point this out in order to suggest that this element may indeed be able to be removed from the design. This is something that you should always be looking for during the optimization of your design in order to, where possible, simplify the design. It is difficult, if not impossible, to make a definitive judgment on this without working with the specific design data and optimization. It is certainly possible that this element may be canceling some higher-order aberration residual.

Figure 5.27 shows two photos of a 17-element lithographic lens produced by Corning/Tropel. It is clear how the individual lens elements are bent in order to minimize the ray angles of incidence on many of the surfaces which, in turn, will tend to reduce the residual aberrations. This is especially clear in the 30-element lens in the lower figure. These lenses are typically extremely difficult to manufacture, assemble, and align due to the large number of surfaces.

Distortion

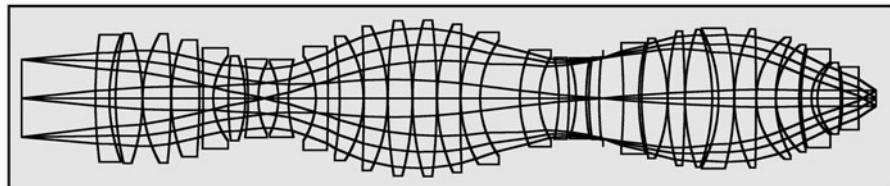
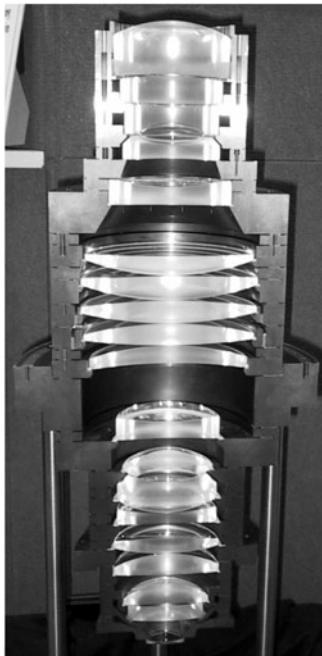
The only aberration that does not result in image blur is distortion. If all other aberrations in the system, except distortion, are corrected, an object point is imaged onto a perfect image point, which is displaced from its paraxial position. The amount of distortion can be expressed either as a lateral displacement in length units or as a percentage of the paraxial image height. Distortion is defined as

$$\text{Distortion} = \frac{y - y_p}{y_p}$$

where y is the height in the image plane and y_p is the paraxial height.

Figure 5.27

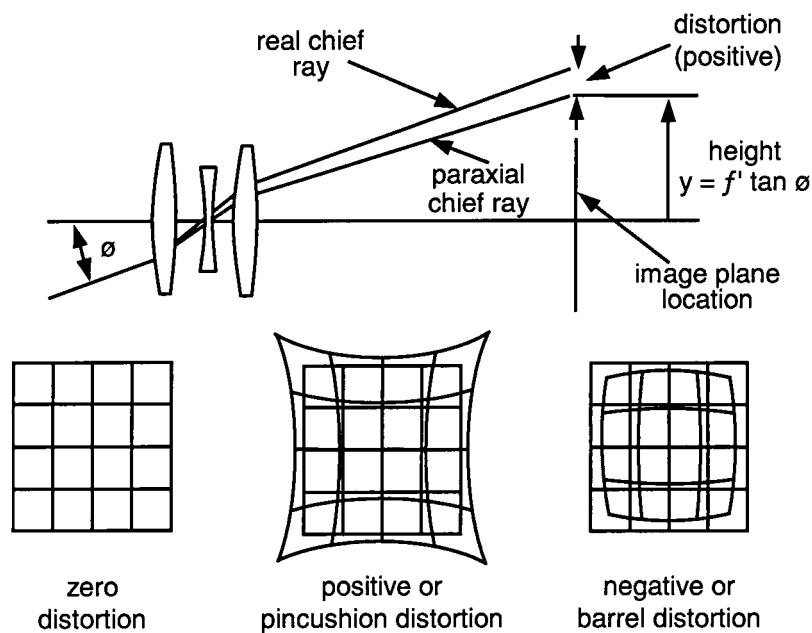
A 30-Element Lithography Lens from the Patent Literature



Third-order distortion increases with the cube of the field of view. A distorted image of a rectilinear object is shown in Fig. 5.28. Distortion can be *positive or pincushion distortion* or alternately *negative or barrel distortion*. For a thin lens with the aperture stop on the lens, distortion is equal to zero. The thickness of a lens and its position relative to the aperture stop determines its contribution to the system distortion. An example of a system where a correction of the distortion is a big challenge is a wide field-of-view eyepiece. Its field of view may be as high as 70° , and its aperture stop, which is the pupil of the eye, can be in the order of 20 mm away from the system. If the object and the image are interchanged, the lens that has a barrel distortion in one direction has a pincushion distortion in the opposite direction.

Review of Specific Geometrical Aberrations

Figure 5.28
Distortion



This is a very interesting subtlety, and we urge the reader to spend due time thinking about it prior to predicting what a distorted image will look like to make sure your sign of distortion is modeling the real world properly.

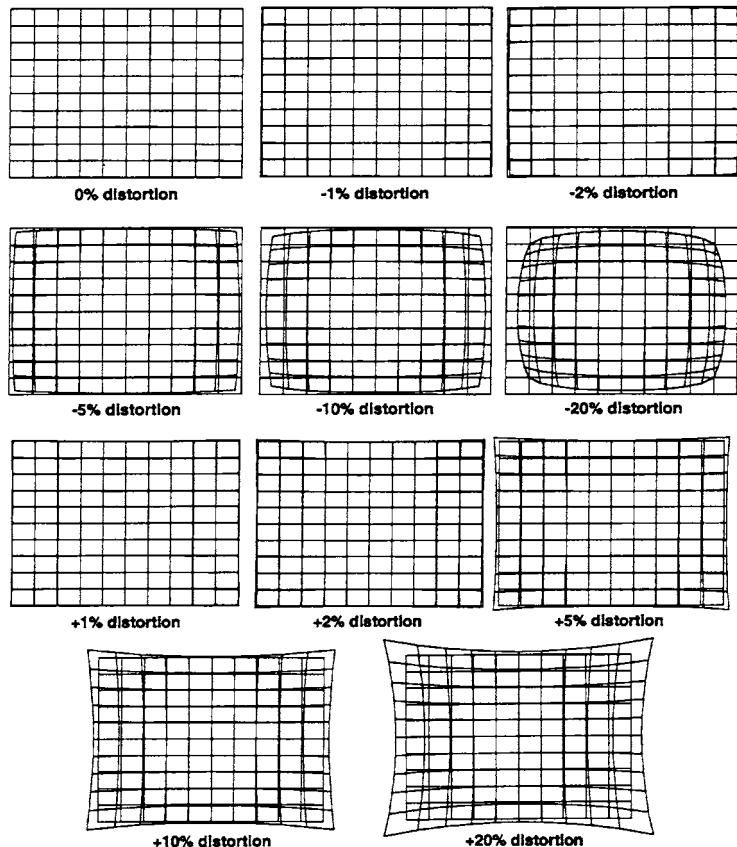
Figure 5.29 shows different amounts of negative and positive distortion. Since distortion is a cosmetic-type aberration not affecting resolution, its appearance is very important, especially in visual systems. Generally, distortion in the order of 2 to 3% is acceptable visually.

Figure 5.30 shows where distortion comes from. In this situation, the aperture stop is located to the left of the lens, and the angle of incidence on the lens by the ray bundle is large enough so that there is a reasonable difference between the paraxial angle of refraction and the real ray angle of refraction. As with spherical aberration, the real rays are refracted more severely than the paraxial rays. In this case, this causes the real image to be pulled inward from the paraxial image thus causing negative or barrel distortion.

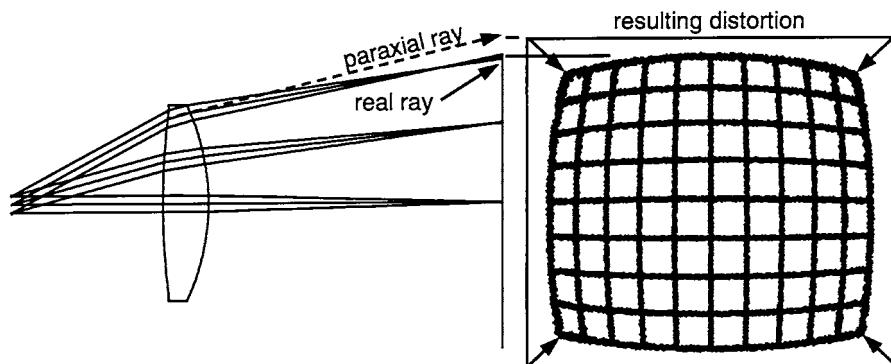
Table 5.1 summarizes the aperture and field dependence of the primary aberrations.

Figure 5.29

Illustration of Different Amounts of Negative and Positive Distortion

**Figure 5.30**

Where Does Distortion Come From?



Review of Specific Geometrical Aberrations

TABLE 5.1

Summary of Third-Order Monochromatic Aberration Dependence on Aperture and Field Angle

Aberration	Aperture Dependence	Field Dependence
Spherical	Cubic	—
Coma	Quadratic	Linear
Astigmatism	Linear	Quadratic
Field curvature	Linear	Quadratic
Distortion	—	Cubic

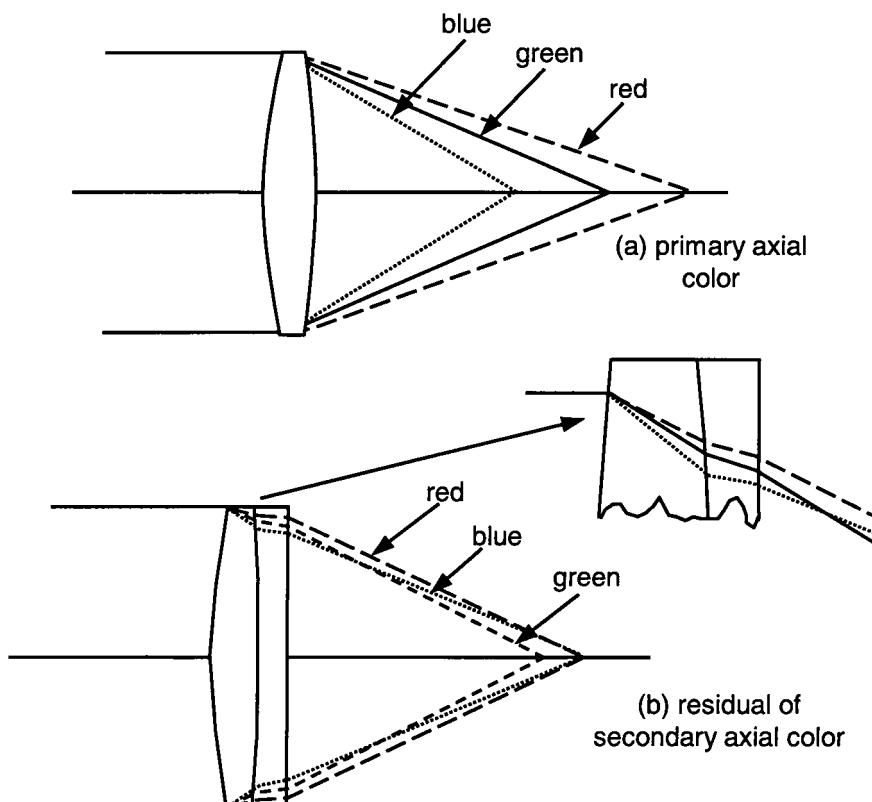
Axial Color

If white light is incident onto a glass wedge or a prism, it is decomposed into a rainbow. This is called *dispersion*. Blue light is refracted more severely than red light, since the index of refraction is higher for shorter wavelengths than for the longer wavelengths. The properties of lenses also vary with wavelength. White light coming from an axial infinitely distant object, which is incident upon a convergent lens, is shown in Fig. 5.31a. The edge of the lens acts like a wedge, refracting or bending the blue light more than the red light. This causes the blue light to focus closer to the lens than the red light. This longitudinal variation of focus with wavelength is called the *axial chromatic aberration* or *axial color*. In the absence of spherical aberration, a system with uncorrected chromatic aberration forms a bright spot surrounded with a purple halo coming from the blue and red light.

Is there a way to correct the axial color? A lens that focuses an infinitely distant object is shown as an example in Fig. 5.31b. In order to bring the blue and the red to focus together, a positive lens must be split into two lenses made of glasses with different dispersions. The first is a positive lens with low dispersion glass. This type of glass is called a *crown* glass. The second lens has a lower optical power than the first one, so that the total power of the doublet is positive. However, the second lens is made of a high-dispersion glass called a *flint* glass, which means that it spreads light more with color, and it cancels most of the axial chromatic aberration created by the first lens because of its negative power. This doublet is called an *achromatic doublet*.

Figure 5.31

Axial Chromatic Aberration



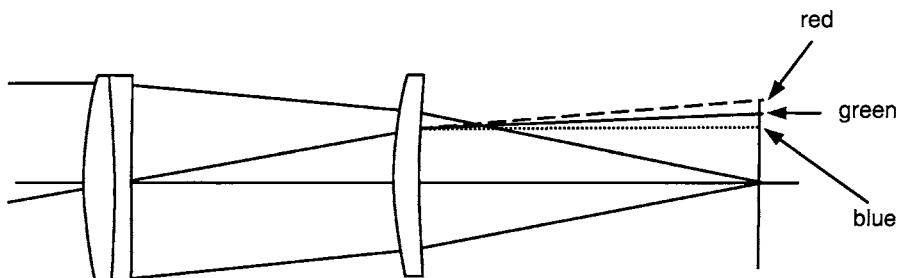
Lateral Color

When a lens forms an image of an off-axis point at different heights for different wavelengths, the lens has lateral chromatic aberration or chromatic difference of magnification. This aberration is quite common in wide field-of-view systems. A very descriptive name for lateral color is *color fringing* since this is what is seen when looking at an image formed by a lens with lateral color.

Lenses that are further from the aperture stop in a system contribute more to lateral color than the lenses with smaller chief ray heights. Lateral color created by a lens is shown in Fig. 5.32. The chief ray is going through the single-lens element close to its outer periphery. Shorter wavelengths are bent or refracted more severely than the longer wavelengths. The blue image is formed closer to the optical axis than the red image. In

Review of Specific Geometrical Aberrations

Figure 5.32
Off-Axis Lateral Color
(Color Fringing)



wide field-of-view systems, lateral color is often the aberration that is the most difficult to correct. Its correction may require the use of anomalous dispersion glasses, which are often expensive, or diffractive elements.

Visual systems often have lateral color due primarily to the eyepiece, which inherently has a lot of lateral color. If you look, for example, through a pair of binoculars at a sharp bright/dark edge close to and tangent to the edge of the field of view, you will likely see severe lateral color. This is often not a problem, as the user will most often place the object of interest at the center of the field of view. For this reason visual optical systems tend to be somewhat more forgiving than other systems.

Parametric Analysis of Aberrations Introduced by Plane Parallel Plates

Let us assume that we have a diffraction-limited $f/1$ lens. What is the spherical aberration introduced by a plane parallel plate inserted in the converging cone between the lens and the image plane? The spherical aberration introduced by the plate increases as a function of the plate thickness. Figure 5.33 shows how the blur diameter due to the spherical aberration changes as a function of the $f/ \#$ of the lens and the plate thickness. The shaded area is the region where the optical system is diffraction limited. For our $f/1$ lens, the Airy disk diameter is $1.5 \mu\text{m}$. If a plane parallel plate of glass 50 mm thick is inserted in the $f/1$ cone of light, the spherical aberration in the focused image increases to 1.25 mm. The image blur is much larger than the Airy disk, which means that the glass plate significantly degrades the performance. We would have to stop down the aperture of the lens to about $f/5$ to reduce the size of the

Figure 5.33

Spherical Aberration
of a Plane Glass Plate,
Refractive Index =
1.517 (BK7)

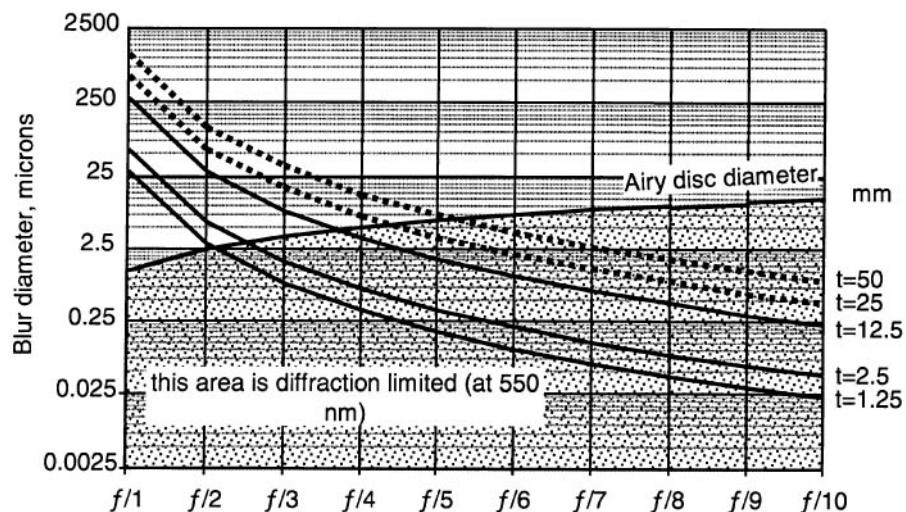


image blur to the Airy disk. At the point where the diffraction blur and the aberrations of the glass plate are equal, the glass plate will not have a detrimental effect on the image quality.

Third-order spherical aberration, coma, astigmatism, and distortion of a plane parallel plate depend on the index of refraction of the plate, n , and the thickness, d , and they are proportional to

$$\frac{(n^2 - 1) d}{n^3}$$

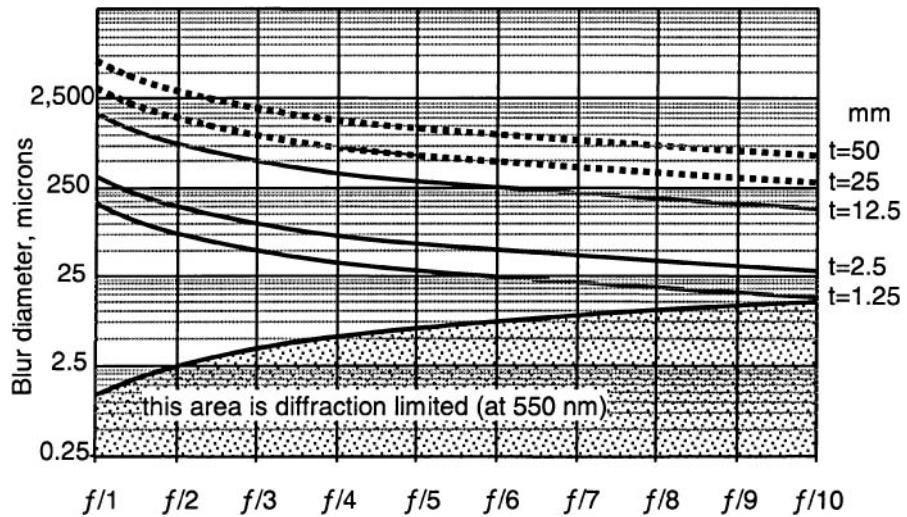
What happens if a plane parallel plate is inserted in a converging cone of light at 45° ? If the tilted plate is inserted in a rotationally symmetric optical system in a convergent beam, the optical system is no longer rotationally symmetrical. A result of this is the presence of field aberrations such as coma, astigmatism, and lateral color in the on-axis field position.

Let us again assume that we have a diffraction-limited, very fast $f/1$ lens and we need to split the beam into two beams, one reflected at 90° and the other beam transmitted through the tilted plate beamsplitter. The reflected beam is unchanged after reflection off the plate, but the transmitted beam has on-axis astigmatism introduced by the tilted plate. The thicker the plate, the more astigmatism is present. The blur diameter associated with third-order astigmatism as a function of the thickness of the tilted plate and the $f/ \#$ of the lens is shown in Fig. 5.34.

Is there a way to correct this on-axis astigmatism? There are a few different viable methods of correction. Astigmatism is proportional to the

Review of Specific Geometrical Aberrations

Figure 5.34
Third-Order Astigmatism
Blur Diameter, in
Micron, as a Function
of the Thickness of a
45° Tilted BK7 Plate
and the f/# of the
Lens



square of the tilt angle, so it is therefore not dependent on the sign of the plate tilt. Thus, astigmatism cannot be compensated with a second plate tilted in the opposite direction from the first plate and in the same plane of tilt. This is shown in Fig. 5.35. However, if the second plate is tilted in a plane which is orthogonal to the plane of tilt of the first plate, the astigmatism can be corrected for the most part. The second method of correction involves the use of a weak spherical surface on the tilted plate or a weak wedge instead of the plane parallel plate. Although it is more difficult for fabrication, a good correction of the astigmatism can be achieved with a decentered cylindrical surface on the tilted plate. You can also think of this component as a wedged cylinder. In this way, the astigmatism, as well as a smaller residual of coma, are reasonably well corrected.

The most severe aberration introduced in an optical system by a tilted plate is astigmatism. However, coma and lateral color are also significant in the case of a fast lens even when the plate thickness is less than 1 mm. The

Figure 5.35
Correction of Astigmatism
from a Tilted Plate in a Converging Beam

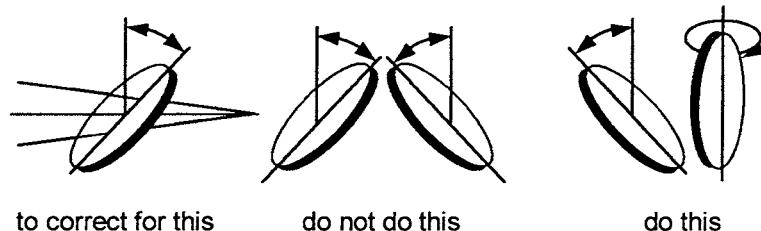
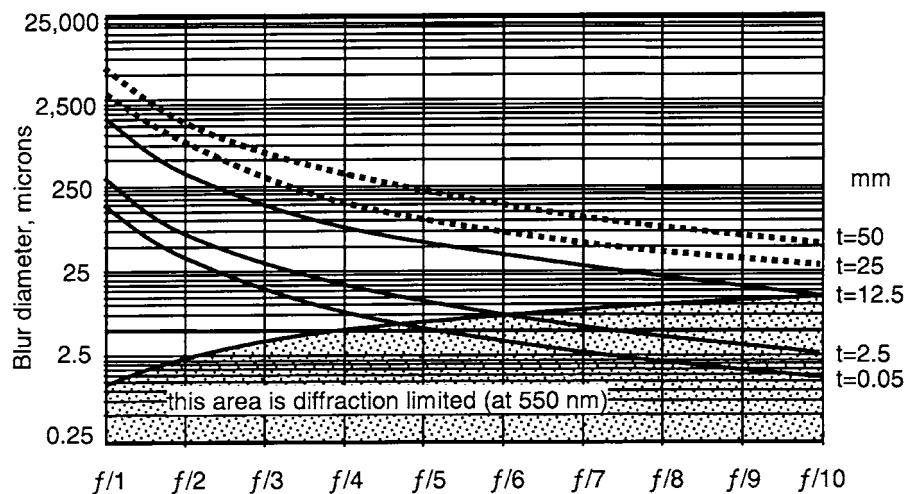


Figure 5.36

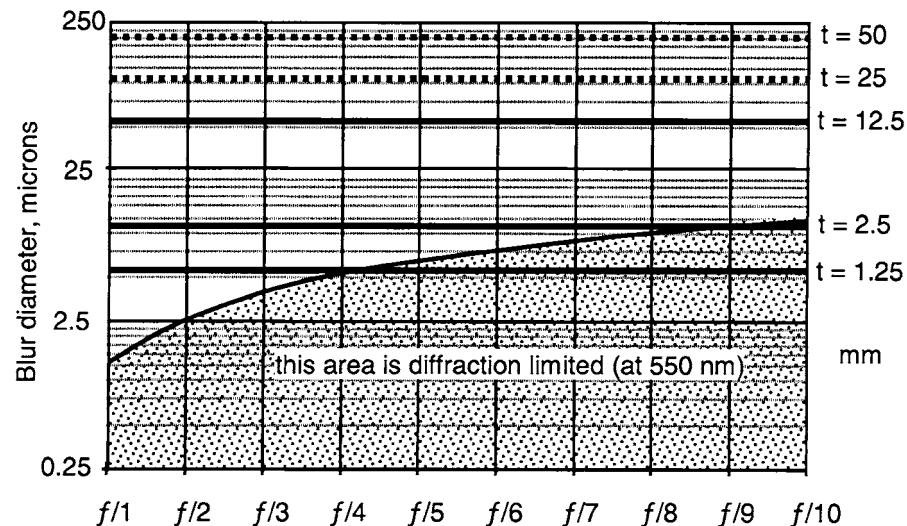
Tangential Coma Blur Diameter As a Function of the Thickness of a 45° Tilted Plate of Index 1.5 and the $f/ \#$ of the Lens



residual tangential coma blur in a system with a 45° tilted plate is shown as a function of plate thickness and the $f/ \#$ of the lens in Fig. 5.36. Lateral color blur in a system with a 45° tilted plate is shown as a function of plate thickness and the $f/ \#$ of the lens in Fig. 5.37. Note that the lateral color is independent of the $f/ \#$. This is because the lateral color is only dependent on the height of the chief rays in the different wavelengths.

Figure 5.37

Lateral Color Blur Diameter in Microns as a Function of the Thickness of a 45° Tilted BK7 Plate and the $f/ \#$ of the Lens



CHAPTER

6

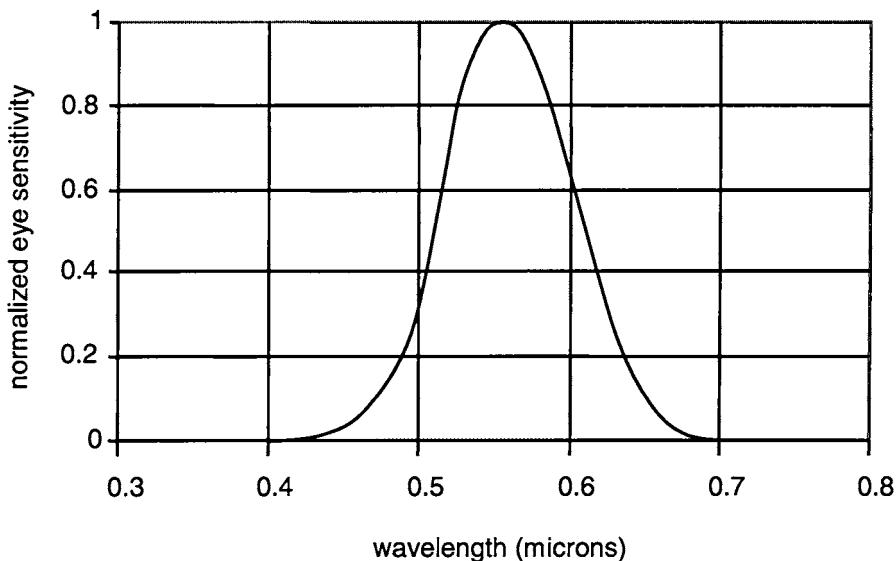
Glass Selection (Including Plastics)

Material Properties Overview

Every optical system works in its own particular wavelength region determined by the spectral characteristics of the light source, the spectral sensitivity of the sensor, as well as any other factors or components which alter the net sensitivity of the system. If an optical system is a visual system, the optical materials must be transmissive between approximately 425 and 675 nm, as determined by the photopic spectral response curve of the human eye. The photopic eye sensitivity is shown in Fig. 6.1. Optical glasses are the most commonly used materials in optical systems. However, there are some optical plastics with good transmission in the visible spectrum that can be injection molded. In high-volume production, this technology is significantly cheaper than classical glass manufacturing methods. Operating temperature range is very important when choosing optical materials. Optical materials change their index of refraction with temperature, and they also expand differently, changing the lens shape and optical power. Optical plastics have approximately one order of magnitude higher coefficient of thermal expansion than glasses.

If the temperature in an optical system rises to a few hundred degrees Celsius, plastic materials cannot be used because the plastic will melt. Most optical glasses can withstand temperatures of a few hundred degrees Celsius without changing their shape. In illumination systems close to the light source, the temperature can go up to 900°C. In this

Figure 6.1
Photopic Spectral Eye
Sensitivity Curve



case, glass optics will melt too. Fused quartz or fused silica is often used in these systems because it can operate at temperatures close to 1000°C.

Manufacturers of optical components usually include in their catalogues information about the standard optical materials they use. An example of the general information on optical materials for the visible, near ultraviolet (UV), and near infrared (IR) spectral regions as provided by Melles-Griot is shown in Fig. 6.2.

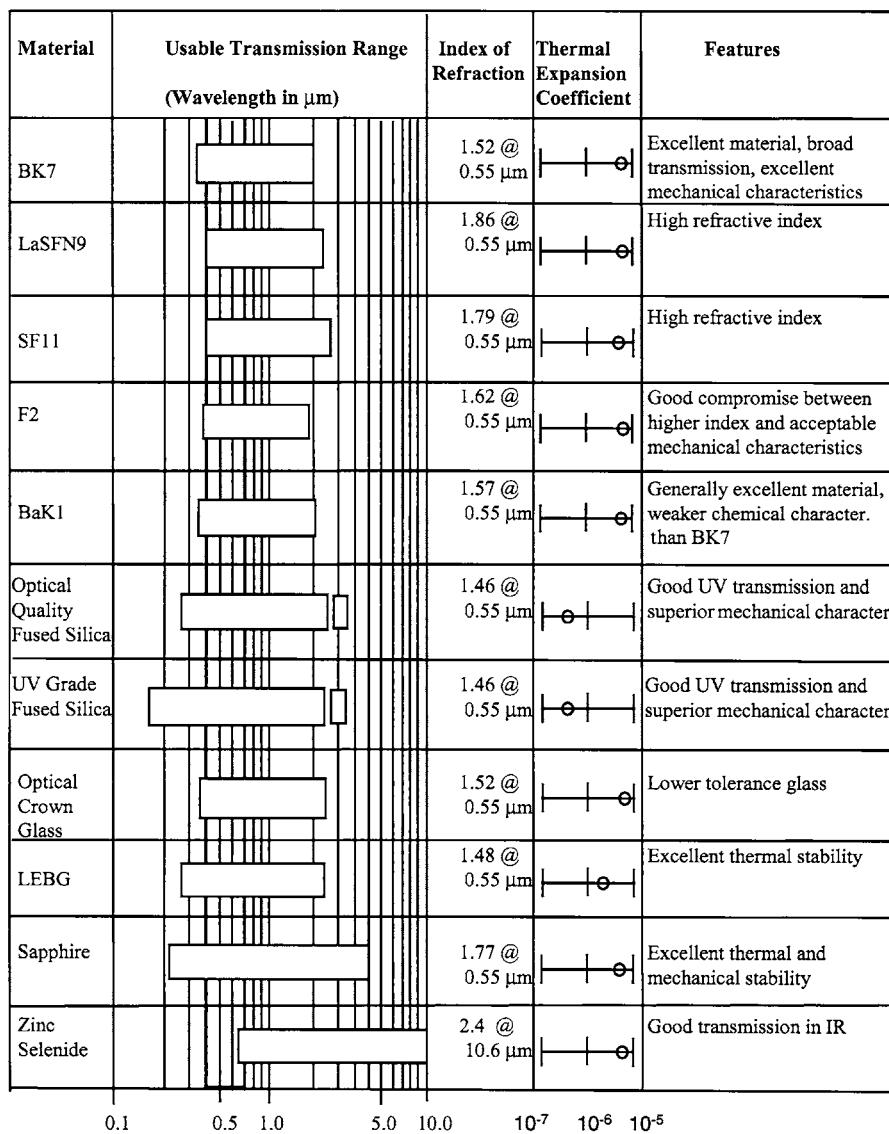
The Glass Map and Partial Dispersion

The refractive index of all optical materials changes as a function of wavelength. The refractive index increases as the wavelength decreases. This means that optical systems with refractive components have chromatic aberrations. In fact, the performance of an optical system is often limited by chromatic aberrations rather than monochromatic aberrations.

In the time of Sir Isaac Newton, it was believed that it was not possible to correct chromatic aberrations by combining different types of glasses. Newton thought that the chromatic aberrations of all lenses were

Glass Selection (Including Plastics)

Figure 6.2
Material Properties
Overview from
Melles-Griot Catalog
(Simplified)



proportional to their powers, with the same constant of proportionality for all glasses. This is the reason why Newton built a reflecting telescope. In the eighteenth century, however, it was found that, with the proper choice of glasses and powers, it was possible to design an *achromatic doublet*, which was chromatically corrected for two wavelengths.

Let us consider two thin lenses made from two different glass types and cemented together as shown in Fig. 6.3. We want to find the condition for this doublet to be an achromatic doublet, chromatically corrected for the red C line wavelength 656.27 nm and for the blue F line 486.13 nm. Generally, the *crown* materials are less dispersive than the flints, and in an achromatic doublet we combine the less dispersive crown as the positive element and the more highly dispersive *flint* as the negative element. The central wavelength is usually chosen as the *d* line, which is 587.56 nm. If the power of the first lens is P_1 and the second lens P_2 , then the total power of the doublet is

$$P = P_1 + P_2$$

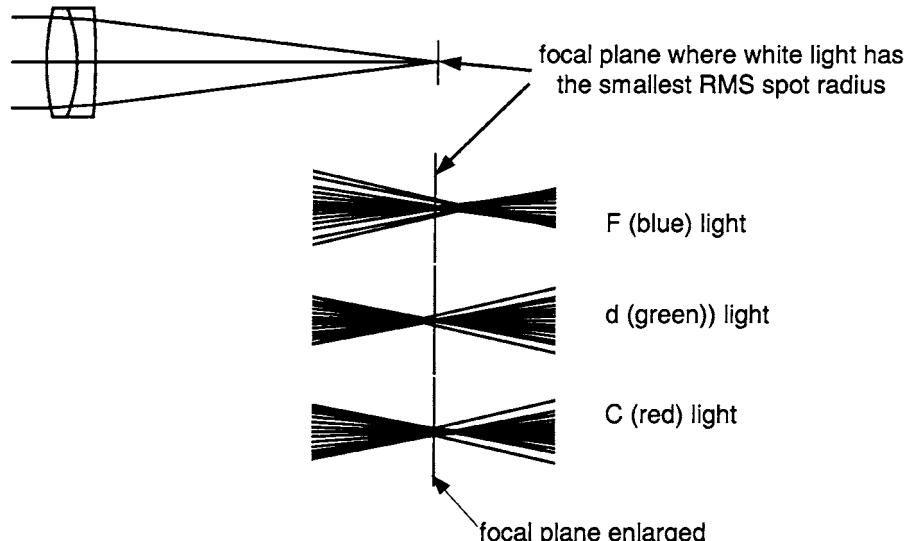
An achromatic doublet will have the same power for the *C* line wavelength and the *F* line wavelength if

$$(P_1 + P_2)_C = (P_1 + P_2)_F$$

or

$$P_1 \frac{n_{1C} - n_{1F}}{n_{1d} - 1} = -P_2 \frac{n_{2C} - n_{2F}}{n_{2d} - 1}$$

Figure 6.3
Focusing of White Light with an Achromatic Doublet from BK7 and SF2 Glasses



Glass Selection (Including Plastics)

where P_1 and P_2 are the powers of two thin lenses at 587.56 nm. The value $(n_F - n_C)$ is called the *main dispersion*. The ratio

$$V = \frac{n_d - 1}{n_F - n_C}$$

is called the V number or the Abbe number. The condition for a doublet to be an achromatic doublet becomes

$$\frac{P_1}{V_1} = -\frac{P_2}{V_2}$$

From this relation we can obtain the focal lengths of two components of an achromatic doublet as

$$f_1 = f \frac{V_1 - V_2}{V_1}$$

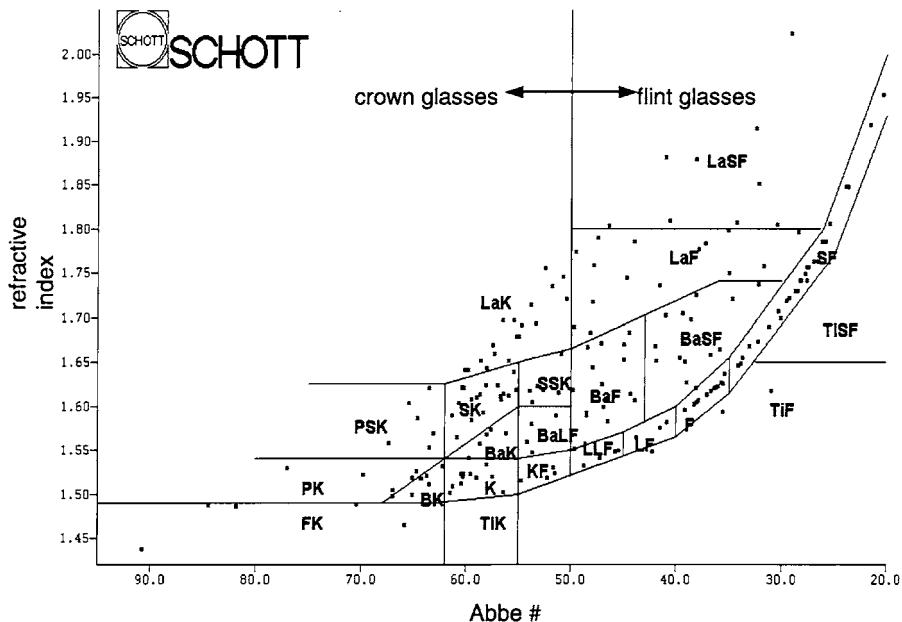
$$f_2 = -f \frac{V_1 - V_2}{V_2}$$

where f is the focal length of the doublet. The net result of this is to derive the powers of the less dispersive crown element and the more dispersive flint element, so that the combined doublet focal length in the red and blue wavelengths are the same. When this condition is reached, the central wavelength (green) is defocused slightly toward the lens.

In the second half of the eighteenth century, Ernst Abbe worked closely with Otto Schott on testing different types of optical glass and this encouraged the development of new glass types. It was found that the most suitable way to characterize optical glass was the specification of the index of refraction for the d line, n_d , and the Abbe number, V , which determines the glass dispersion. Manufacturers of optical glasses provide a glass map or an n_d/V_d diagram in which the Abbe number is plotted as the abscissa and the index, n_d , as the ordinate. The glass map from the Schott glass catalogue is shown in Fig. 6.4. Schott is the largest manufacturer of glass in the world, but there are other manufacturers, including Hoya, Ohara, Pilkington, Corning, and Sovirel.

The n_d/V_d diagram subdivides the various types of glasses into groups, each having a specific designation such as BK with BK7, BK1, and others. These designations are generally related to the fundamental materials used in the manufacture of the specific group such as LAFN31 which is a lanthanum flint glass. There is also a more general division of glasses into "crown" and "flint" glasses. The crown glasses are

Figure 6.4
Schott Glass Map
(Abbe Diagram)



the ones with $n_d > 1.60$, $V_d > 50$ or $n_d < 1.60$, $V_d > 55$; the other glasses are flints. The available refractive indices range from 1.45 to 2 and the V number from 80 to 20.

Mathematically, the dependence of refractive index on the wavelength of light can be expressed in a few different ways, but none of the expressions is highly accurate over the entire glass transmission range. These relationships are empirically derived from measured data. The Sellmeier dispersion formula is

$$n^2 - 1 = \sum_i \frac{1}{\lambda^2 - \lambda_i^2} c_i \lambda^2$$

and the formula from the 1967 Schott catalogue is

$$n^2 = A_0 + A_1 \lambda^2 + \frac{A_2}{\lambda^2} + \frac{A_3}{\lambda^4} + \frac{A_4}{\lambda^6} + \frac{A_5}{\lambda^8}$$

These are the two most commonly used formulas. Besides these two formulas, the Hartmann and Conrady formulas are also offered in some of the lens design programs. The six constants that characterize glass dispersion vary considerably between glasses, and thus the general shapes of all dispersion curves are different.

Glass Selection (Including Plastics)

In addition to the main dispersion ($n_F - n_C$), which is the difference between the index of refraction for the blue and for the red line, the “partial dispersion” is also commonly used. Partial dispersion in the blue is the difference in index of refraction between 435.83 and 486.13 nm, and the red partial dispersion is the difference in index of refraction between 653.27 and 852.11 nm.

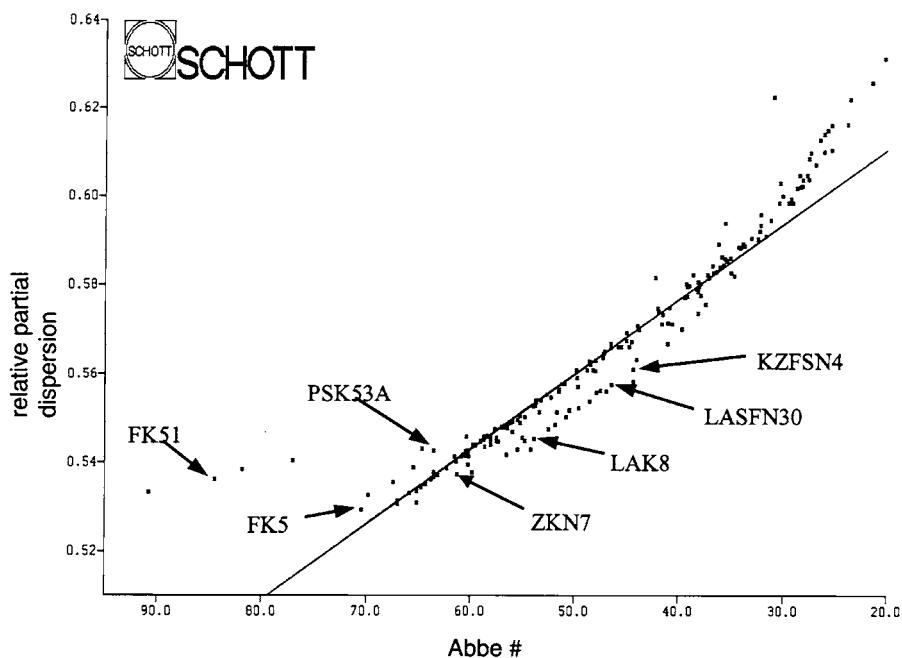
Perhaps even more important is the “relative partial dispersion,” which is the ratio of the partial dispersion and the main dispersion. Generally, the relative partial dispersion is

$$P_{x,y} = \frac{n_x - n_y}{n_F - n_C}$$

A glass map with the relative partial dispersion as a function of the Abbe number from the Schott catalogue is given in Fig. 6.5.

The derived formulas for the design of an achromatic doublet provide a chromatic correction for *F* and *C* wavelengths. However, dependent on the choice of glasses, there will be a residual mismatch of dispersions, resulting in a larger or smaller “secondary spectrum.” This secondary spectrum is the difference in image position between the central wavelength (green

Figure 6.5
Relative Partial
Dispersion



or yellow) and the now common blue and red image position. In order to eliminate the secondary spectrum, we should find a pair of glasses with different V values but the same relative partial dispersion. Abbe showed that the majority of glasses, the so-called normal glasses, exhibit an approximately linear relationship between the relative partial dispersion and the Abbe number. Thus

$$P_{x,y} \approx a_{x,y} + b_{x,y} V_d = (P_{x,y})_{\text{normal}}$$

which can be clearly seen in the relative partial dispersion map in Fig. 6.5. The reduction of the secondary spectrum requires the use of at least one glass type which does not lie on the $(P_{x,y})_{\text{normal}}$ line. The glasses that lie away from the line of normal glasses are often expensive and may be difficult to manufacture. Some of them are shown in the relative dispersion chart in Fig. 6.5. KZFSN4 is seven times more expensive than BK7, LaK8 is eight times, PSK53A is 11 times, and LaSFN30 24 times.

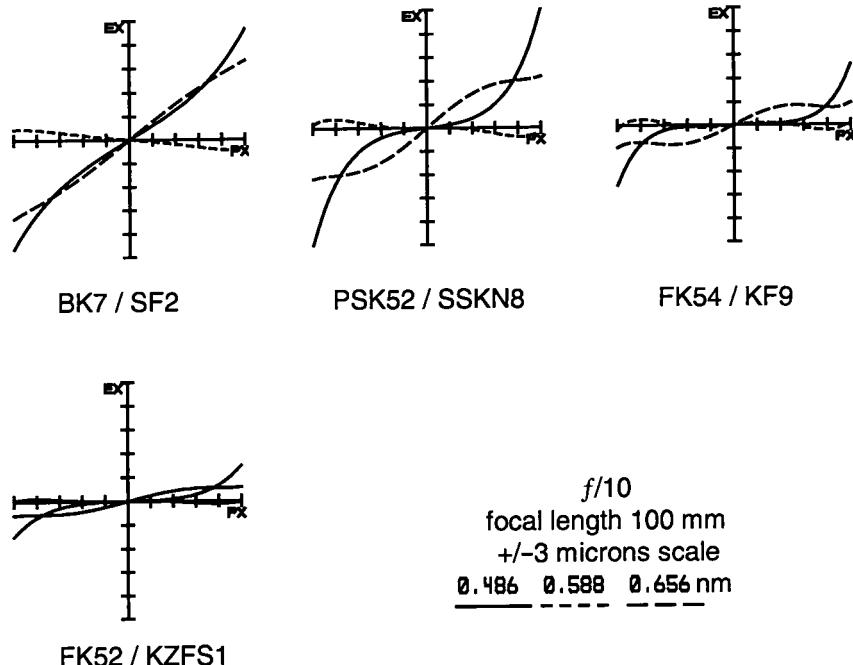
Parametric Examples of Glass Selection

In this section we will show how secondary spectrum can be corrected and to which level, with the right choice of glasses. It will also be shown how the spherical aberration and the secondary spectrum of a doublet are dependent on the $f/\#$. These data are shown as parametric analyses.

The first parametric study is shown for the case of an $f/10$ achromatic doublet using different glass combinations. Four doublets will be compared, the first using two normal glasses, and then using anomalous dispersion glasses for one or both of the elements. The first doublet is designed with two normal glasses: BK7 and SF2. It is an $f/10$ lens with a 100-mm focal length. The ray aberration curves are shown in Fig. 6.6. An explanation of these curves will be given in Chap. 10. The difference in the Abbe numbers between the two glasses should be sufficiently large, so that the shape or the power of each individual component is reasonable. Note that as the Abbe number difference between the two glasses of an achromatic doublet decreases, the relative powers of the positive and negative elements get stronger. This yields greater spherical aberration, as will be seen. For the $f/10$ lens the spherical aberration for the central wavelength is very small, the rms spot diameter is less than 1 μm . However,

Glass Selection (Including Plastics)

Figure 6.6
Secondary Spectrum
Correction As a
Function of Glass
Selection



both the red and the blue foci are away from the green focus, which means that the secondary spectrum aberration is not corrected.

The second achromatic doublet is designed with PSK52 and SSKN8 glasses. SSKN8 is a normal glass, but PSK52 has anomalous dispersion. The green spot diameter is the same as in the case of the first doublet, but the polychromatic spot diameter is slightly smaller because the secondary spectrum is lower. Note here that we have an increase of spherical aberration which changes with wavelength. The blue has positive spherical aberration and the red has negative spherical aberration. This change in spherical aberration with wavelength is called *spherochromatism*.

The third case is a doublet with FK54 and KF9 glasses. Although KF9 is a normal glass, FK54 has an extremely high anomalous dispersion, resulting in a much better secondary spectrum correction than in the previous case. FK54 is over 30 times the cost of standard BK7 glass!

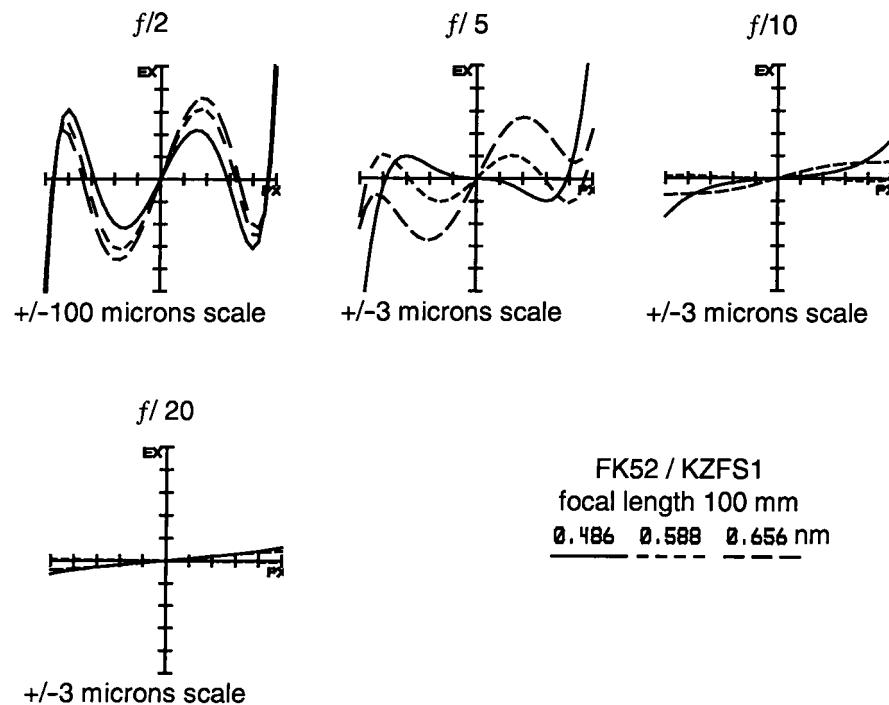
The fourth case is a doublet with FK52 and KZFSN4. Both glasses have anomalous dispersions. The polychromatic spot diameter is less than 1 μm , and the secondary spectrum is completely corrected. Furthermore, the spherochromatism is extremely well corrected as well.

The second parametric study is the analysis of spherical aberration and secondary spectrum correction as a function of the lens $f/ \#$ for a chosen set of glasses.

The first pair of glasses are FK52 and KZFS1, as shown in Fig. 6.7. As demonstrated for an $f/10$ lens, the secondary spectrum can be very well corrected, since both glasses have anomalous dispersion. The Abbe numbers for these two glasses are very different; however, the relative partial dispersion is not. In the case of the $f/2$ lens, the spherical aberration is dominant, with the spot diameter close to 200 μm . As the $f/ \#$ increases to $f/5$, the spot diameter decreases dramatically to about 3 μm . Chromatic aberration is more pronounced in this case. As the $f/ \#$ increases to $f/20$, both the spherical and the chromatic aberrations are extremely well corrected, and the spot diameter is less than 1 μm .

The second pair of glasses to be considered is LASFN31 and SFL6, as shown in Fig. 6.8. Although these glasses are high-index glasses and have anomalous dispersion, their relative partial dispersion is quite different and the secondary spectrum correction is poor. The difference in the

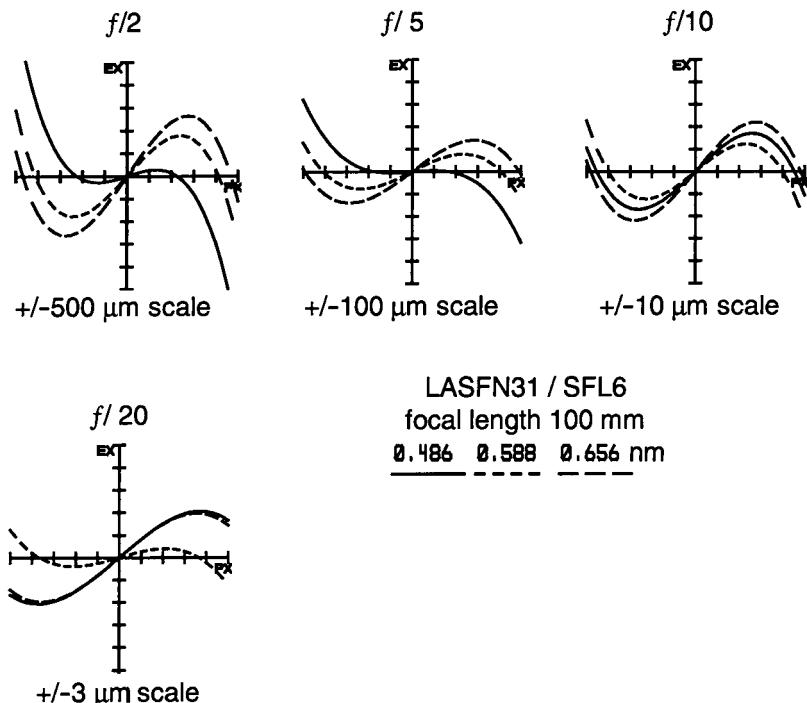
Figure 6.7
Spherical Aberration
and Secondary Spec-
trum Correction As a
Function of $f/ \#$



Glass Selection (Including Plastics)

Figure 6.8

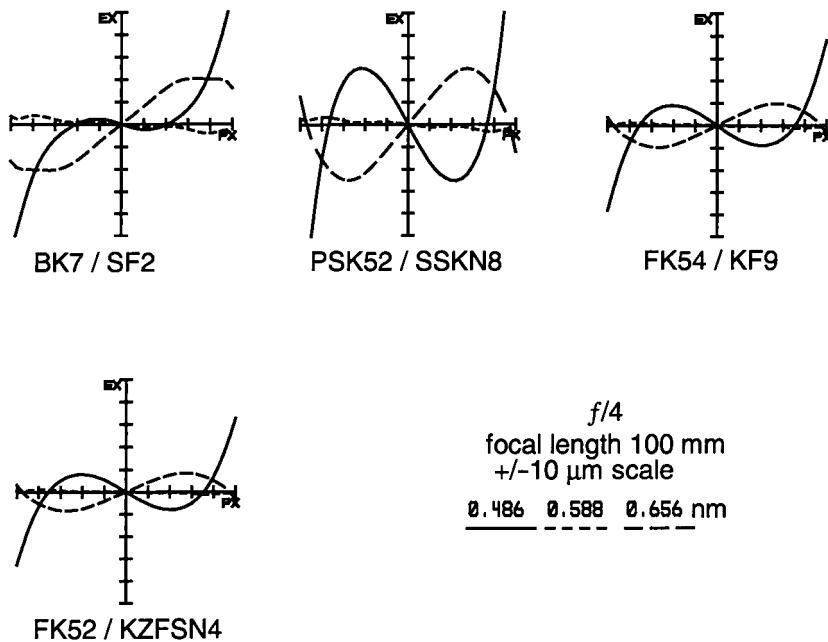
Spherical Aberration and Secondary Spectrum Correction As a Function of $f/\#$



Abbe numbers is not large. This makes it difficult to correct for the spherical aberration too, since we need to correct both aberrations simultaneously, trying to find the optimum balance between the two. In the case of the $f/2$ lens, the spot diameter is around $400 \mu\text{m}$. As the $f/\#$ increases, the spot diameter decreases to about $2 \mu\text{m}$ at $f/20$, largely because of the uncorrected secondary spectrum.

The third parametric study is done for the case of an $f/4$ achromatic doublet using different glass combinations but allowing one surface to be aspheric for nearly complete correction of spherical aberration. This allows us to better see the change in residual spherical aberration with wavelength or spherochromatism. Four doublets are compared, with the same glass combination as the first parametric study. Ray aberration curves are shown in Fig. 6.9. In all four cases, the spherical aberration for the central wavelength is almost perfectly corrected with the asphere on the front surface of the doublet. The spot diameter is determined by the chromatic blur of the blue and the red wavelength, and the difference between four cases of glass combination is small. Note that the shape of the ray aberration curves for the blue and the red color are similar in all cases.

Figure 6.9
Secondary Spectrum Correction As a Function of Glass Selection with One Aspheric Surface



How to Select Glass

Let us imagine that we have to design a wide field-of-view objective that operates in the visible spectrum. We will most likely start with five to seven elements and select a starting configuration based on our prior work, a patent, or we may elect to derive the design from basic principles. In most cases, it is sufficient to optimize a lens for only three wavelengths if the system is a visual system. However, in the case of a large field of view, it is better to work with five wavelengths properly weighted and a larger number of field points because of the potential problems with lateral color correction. At some point in the design, we will start to change the glass types either manually or by varying the glass characteristics automatically in the optical design program, and allowing them to move across the glass map until they settle in the locations which provide the lowest merit function. We will notice that the glasses tend to go from FK, PSK, across SK, LaK, LaSF to SF. Very rarely a chosen glass will be KE, LLE, LE or F. Even the glasses, such as BaLE, BaF, or BaSF are not so often chosen in the optimization process. However,

there are some types of glasses that are in the central region of the glass map such as KzFS and TiF that are often chosen in the glass optimization process. The reason for this is that, unlike BaF or BaLF glasses, which are the normal glasses, these glasses have anomalous relative partial dispersions and color correction, particularly secondary spectrum, is much easier with these glasses.

Generally, glasses in the central region of the glass map (BaLF, BaF, BaSF glasses) are not frequently used because they are normal glasses and secondary spectrum is not easily corrected with them. The other reason is that their V number has a medium value, which means that primary color is not easily corrected either. The exception is KzFS glasses, which have anomalous dispersion. Unfortunately, as will be discussed later, KzFS glasses are not preferred Schott glasses.

Now comes a very important step in the optical design—we have to check many parameters for each glass, including its availability, price, transmission properties, thermal properties, staining, etc., and make sure that the glass choice is the optimum one. Here the most important parameters an optical designer has to consider in the process of glass selection will be described. This will also be covered in Chap. 18 from a manufacturer's perspective.

AVAILABILITY Glasses are divided into three groups: preferred, standard, and inquiry glasses. Preferred glasses are always available. Note that just because a glass is "preferred" does not mean that it is of good optical characteristics or low in cost, nor does it mean that the glass is easy to work in the shop; it only refers to availability. Quick delivery of standard glasses is generally possible as these glasses are generally in inventory. Inquiry glasses are available only on request, and they are normally not in stock. An optical designer should make every effort to design the system with preferred glasses. The optical design software program Zemax contains an option to use only preferred glasses from the Schott catalogue when a system is optimized with the "hammer" optimization and "substitute" glasses are used. This will be illustrated in the case studies in Chap. 22.

TRANSMITTANCE Most optical glasses transmit light well in the visible and the near IR wavelength spectrum. However, in the near UV, the light is more or less absorbed by most glasses. If an optical system has to transmit UV light, the most commonly used materials are fused silica and fused quartz. Some optical glasses, such as a few SF glasses, have a

reduced transmittance in the deep blue wavelength spectrum, and they have a yellowish appearance. Glass absorption as a function of wavelength is given in glass catalogues for 5- and 25-mm-thick glass plates.

INFLUENCE OF STRESS ON THE REFRACTIVE INDEX Optically isotropic glasses become anisotropic through mechanically and thermally induced stress. This means that the *s* and *p* polarization components of light undergo refraction with different indices of refraction. High index alkali lead silicate glasses (dense flints) display a relatively large absolute refractive index change with a small stress birefringence. On the other hand, borosilicate glasses (crowns) exhibit a small absolute change in refractive index with a relatively large stress birefringence. If the optical system has to transmit polarized light and has to maintain the state of polarization throughout the system or part of it, the choice of materials is very important. For example, when there is a prism in such a system with a relatively large mass, and there is a source of heat in its proximity, there can be a gradient of temperature inside the prism. It will introduce stress birefringence, and the polarization axis will be rotated inside the prism. In this case a better choice for the prism material should be one of the SF glasses rather than crown glasses.

CHEMICAL PROPERTIES Optical glasses acquire their properties through their chemical composition, melting process, and finishing methods. In order to achieve the desired optical properties, optical glasses often exhibit reduced resistance to environmental and chemical influence. There is no single test method sufficient to describe the chemical behavior of all optical glasses. Four characteristics of glass resistance to environmental and chemical influence are given for each type of glass. In the Schott catalogue, glasses are sorted in four groups depending on their climatic resistance, which is the resistance to the influence of water vapor in the air. Water vapor in the air, especially under high relative humidity and high temperature, can cause a change in the glass surface in the form of a cloudy film that generally cannot be wiped off. Glasses are sorted in six groups depending on their stain resistance, which is a resistance to the influence of lightly acidic water without vaporization and possible changes in the glass surface. When the glass is in contact with an acidic aqueous medium, not only can stains appear on the glass surface, but the glass can also be decomposed. Optical glasses are divided into eight groups according to their resistance to acids. The last division of glasses is in four groups according to their resistance to alkalis.

THERMAL PROPERTIES Optical glasses have a positive coefficient of thermal expansion, which means that glasses expand with an increase in temperature. The expansion coefficient, α , lies between 4 e-6 and 16 e-6/K for optical glasses. There are a few things that one should consider when designing an optical system to work in the given temperature range:

- Thermal expansion or contraction of glass should not be in conflict with the expansion or contraction of the lens housing.
- The optical system may have to be athermalized, which means that the optical characteristics of the system are unchanged with a change in the lens shape and index of refraction with temperature change.
- Change in temperature can cause temperature gradient in glass, and this can result in temperature-induced stress birefringence.

Most optical design programs have the capability of system optimization simultaneously at several different temperatures. The programs take into account both the expansion of glass elements and changing of their shape, expansion of the housing and the spacers between lenses, as well as the change of index of refraction of the glass materials.

Plastic Optical Materials

In a high-volume production environment, optical components or optical systems require low-cost materials and low-cost fabrication techniques. Plastic optics are used frequently today primarily for this reason. Plastic optical materials also have lighter weight, higher impact resistance, and offer more configuration possibilities than glass materials. Configuration flexibility is one of the greatest advantages of plastic optics. Aspheric lenses and elaborate shapes can be molded, for example, lenses with integral mounting brackets, spacers, and mounting features for easy alignment.

There are some issues, however, that must be considered when using plastic as an optical material. The principal disadvantage of plastic is its relatively low heat tolerance. Plastic melts at a much lower temperature than glass. It is less resistant to surface abrasion and chemicals. Adhesion of coatings on plastic is generally lower than on glass because of the limitation on the temperature at which the coatings are

deposited, due to a low melting temperature of plastic. Further, the durability of coatings on plastic lenses is less robust than on glass. In addition, coatings on plastic often craze over time. The use of ion-assisted deposition of plastic coatings offers harder and more durable coatings on plastic.

The choice of optical plastic materials is very limited, which means that there is not a lot of freedom in the optical design process. A very important limitation is the high thermal coefficient of expansion and a relatively large change in refractive index with temperature. The refractive index of plastic materials *decreases* with temperature (it increases in glasses), and the change is roughly 50 times greater than in glass. The thermal expansion coefficient of plastic is approximately 10 times higher than that of glass. High-quality optical systems can be designed with a combination of glass and plastic lenses. In a combination with glass components in the system, plastic lenses can reduce the price and complexity of the optical system tremendously. When the optical power is mainly distributed over the glass components in the system, with one or two weak-powered plastic aspheric correctors, optical aberrations, especially distortion in wide field-of-view systems, can be very efficiently removed. Weak-powered plastic elements are used to minimize the effect on focus with temperature change.

Plastic optics can be injection molded, compression molded, or fabricated from cast plastic blocks. Fabrication of plastic elements by machining and polishing from cast plastic blocks is economical in the case of large optical elements, where the molding process has severe limitations. Compression molding offers a high degree of accuracy and control of optical parameters. However, injection molding is the most economical process. It offers moderate optical performance, which is acceptable in a lot of applications. Manufacturing of molds is an expensive process, but it pays off in high-volume production. During the system development phase, plastic optics can be very successfully diamond turned for prototyping, since the cost of diamond turning is lower than the cost of the manufacturing of molds. With today's high-quality diamond turning, the scattering effect from the turning grooves is most often under control, and if you have a good vendor, this should not be of concern for visible applications. Sometimes "postpolishing" is required to remove the turning mark residuals.

During the design of systems with plastic elements, the optical designer has to control the shape of the lenses more carefully than for glass elements. The shape (or bending) of the lens should be optimized

Glass Selection (Including Plastics)

for a good flow of the plastic material inside the molds. The thickness of the lens should be quite small, and the parting line, which is the line of contact of two molds, should go through the lens material. It is also important to eliminate inflection points on the lens surfaces in the case of compression molding. This limits the available lens shapes, and requires more parameters to be controlled in the optimization process. Additionally, the lens shape and the refractive index change with temperature have to be monitored, or the system has to be optimized for a given temperature range.

A few of the most commonly used plastic materials are acrylic (poly-methyl methacrylate), polystyrene, polycarbonate, and COC (cycloolefin copolymer):

Acrylic. The most common and important optical plastic material. It has a good clarity and a very good transmission in the visible spectrum, a high Abbe number (55.3), and very good mechanical stability. Acrylic is easy to machine and polish, and it is a good material for injection molding.

Polystyrene. Also a good plastic, cheaper than acrylic, but it has a slightly higher absorption in the deep blue spectrum. Its index of refraction (1.59) is higher than that of acrylic but the Abbe number is lower (30.9). It has a lower resistance to ultraviolet radiation and scratches than acrylic. Acrylic and polystyrene make a viable achromatic pair.

Polycarbonate. More expensive than acrylic, but it has very high impact strength and a very good performance over a broad temperature range. Polycarbonate is often used for plastic eyeglasses. A common form of polycarbonate in eyeglasses is CR39.

COC. A relatively new material in the optics industry, it has many characteristics similar to acrylic. However, its water absorption is much lower and it has a higher heat distortion temperature. COC is also brittle. A new brand name for COC is Zeonex. Comparative properties of optical plastics are shown in Table 6.1.

A Visual Aid to Glass Selection

Figure 6.10 shows what will be referred to as “a visual aid to glass selection.” The glass map is divided into six general regions. The following short summary represents how one selects glasses within these six

TABLE 6.1

Optical and Physical Properties of Optical Plastics

Property	Acrylic	Polystyrene	Polycarbonate	COC
Index @588 nm	1.49	1.59	1.586	1.533
Abbe#	55.3	30.87	29.9	56.2
$dn/dT \times 10^{-5}/^{\circ}\text{C}$	-8.5	-12	-10	-9
Linear expansion coefficient/ $^{\circ}\text{C}$	6.5×10^{-5}	6.3×10^{-5}	6.8×10^{-5}	6.5×10^{-5}
Transmission (%)	92	88	90	91
Birefringence	Low	High/low	High/low	Low
Tensile strength (lb/in ²)	10,000	6000	9000	8700
HDT at 264 lb/in ² ($^{\circ}\text{C}$)	92	82	142	120–180
Impact strength (ft-lb/in)	0.3	0.4	>5	0.45
Density (g/cm ³)	1.2	1.05	1.2	1.02
Water absorption (%)	0.3	0.02	0.15	0.01
Advantages	High stiffness, hardness, chemical resistance, and low cost	High index and low cost	Excellent impact resistance and high HDT	High stiffness, high HDT, low water absorption
Disadvantages	Brittle and heat resistant	UV absorption, birefringence, and low-impact strength	High birefringence, low Abbe #, and poor scratch resistance	Brittle

regions. While this is extremely helpful, do note that this does not include glasses with anomalous dispersion characteristics such as the so-called “short flints” of which KZFSN4 is an example (see note below for region 6):

1. Crown (+ elements) low-average dispersion, very common glass
2. Flint (– elements), higher dispersion

Glass Selection (Including Plastics)

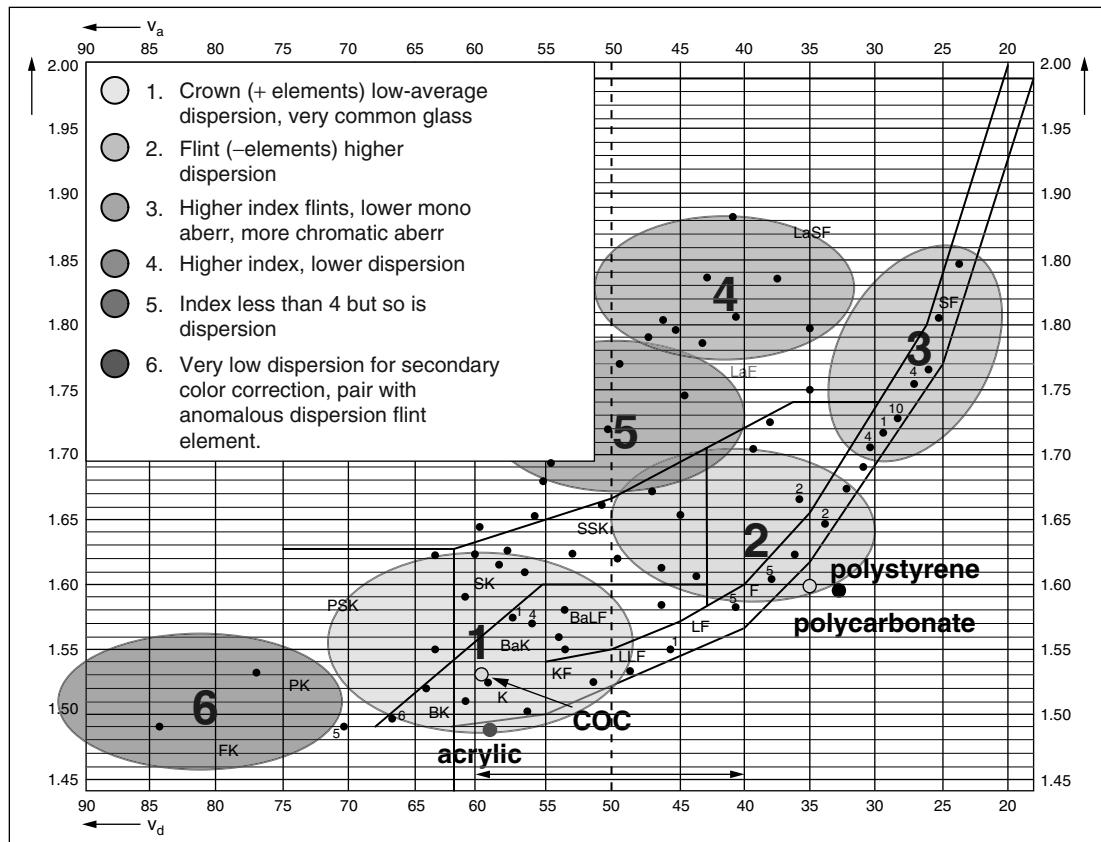


Figure 6.10

A Visual Aid on How to Select Glasses

3. Higher index flints, lower mono aberration, more chromatic aberration
4. Higher index, lower dispersion
5. Index less than 4 but so is dispersion
6. Very low dispersion for secondary color correction, pair with anomalous dispersion flint element

This page intentionally left blank

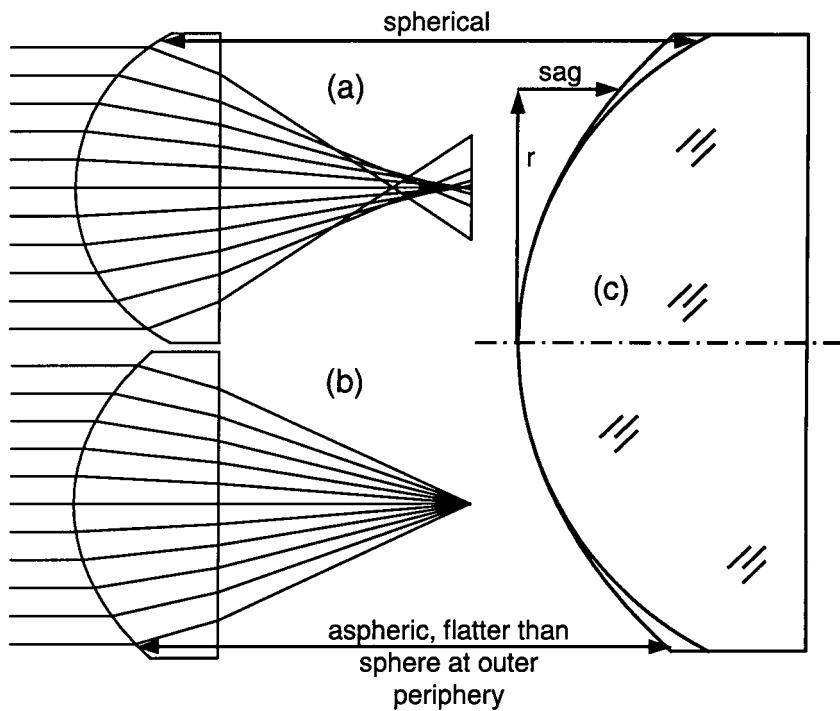
Spherical and Aspheric Surfaces

Definition of an Aspheric Surface

A *spherical* surface is defined by only one parameter, the radius or curvature of the surface. If the surface is refractive, with different indices of refraction before and after the surface, then the power of the surface is defined by the surface radius and the indices of refraction of the two media. *Radius* and *curvature* are reciprocal to one another.

Figure 7.1a shows a plano convex lens element with a spherical radius, imaging an axial point from infinity. The spherical aberration is quite evident. The high angle of incidence of the upper limiting ray of approximately 45° to the surface normal causes this ray to refract very strongly and ultimately to cross the axis significantly closer to the lens than rays closer to the optical axis. A spherical surface has the property that the rate of change of the surface slope is exactly the same everywhere on the surface, and thus the aberration is inevitable. Let us consider reducing the slope of the surface toward the outer periphery of the surface in order to flatten the shape in the region surrounding the outer rays. If we make the surface shape gradually flatter as we proceed outward from the optical axis, we can differentially reduce the refracting ray angle so that the net effect is to bring all of the rays to a common focus position, as shown in Fig. 7.1 b. Figure 7.1c compares the spherical surface, which is steeper at its edge, with the aspheric surface, which is flatter at its edge. While correction of spherical aberration is not the only application of aspheric surfaces, it is one of the major application areas.

Figure 7.1
Comparison of a Spherical and an Aspheric Lens



Aspheric surfaces cannot be defined with only one curvature over the entire surface because its localized curvature changes across the surface. An aspheric surface is usually defined by an analytical formula, but sometimes it is given in the form of a sag table for coordinate points across the surface. The sag of a surface is shown in Fig. 7.1. The most common form of an aspheric surface is a rotationally symmetric surface with the sag defined as

$$z = \frac{cr^2}{1 + \sqrt{1 - (1+k)c^2 r^2}} + \sum a_i r^{2i}$$

where c is the base curvature at the vertex, k is a conic constant, r is the radial coordinate measured perpendicularly from the optical axis, and $a_i r^{2i}$ are the higher-order aspheric terms.

When an aspheric surface is not rotationally symmetric, it is given either as a *biconic* surface with two basic curvatures and two conic constants in two orthogonal directions or as an *anamorphic asphere*, which has additional higher-order terms in two orthogonal directions.

Spherical and Aspheric Surfaces

Another form of aspheric surface is a *toroid* or *toric*. A toroid has, in effect, the shape of a doughnut. If a doughnut were sitting on the table, we all would agree that it had a basic outer diameter. If we now cut the doughnut vertically into two halves and we look at the cut, we see a circle whose diameter is less than the diameter by perhaps a factor of 5 or thereabouts. These two radii define a toroid, the overall outer radius of the doughnut and the smaller cross-sectional radius. If you work with torics in lens design, it is extremely important that you understand fully and completely the definition used by the computer program you are using, so take the time to study the manual in depth in this regard. In addition, if need be, set up a sample to assure that your understanding is correct. While a doughnut is one form of toroid, a football shape is another, and it is imperative to understand which one your equation is representing, especially if it is to be manufactured.

Conic Surfaces

In the case where the higher-order aspheric terms are zero, the aspheric surface takes the form of a rotationally symmetric conic cross section with the sag defined as

$$z = \frac{c r^2}{1 + \sqrt{1 - (1+k)c^2 r^2}}$$

where c is the base curvature at the vertex, k is a conic constant, and r is the radial coordinate of the point on the surface. In Table 7.1, it is shown how a conic surface takes on the following surface types as a function of the conic constant, k , in the sag equation.

TABLE 7.1

Conic Section Types

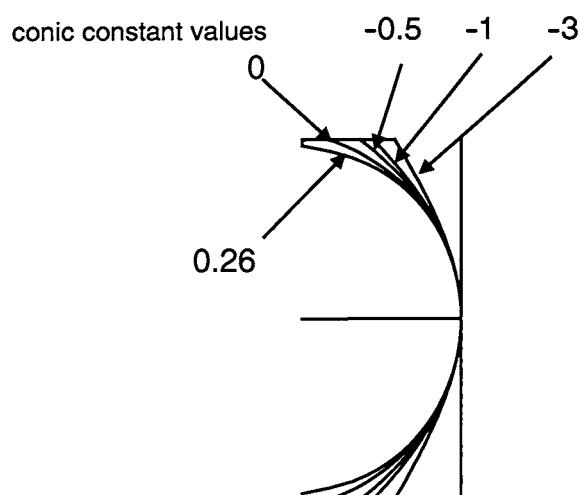
Conic Constant k	Surface Type
0	Sphere
$k < -1$	Hyperboloid
$k = -1$	Paraboloid
$-1 < k < 0$	Ellipsoid
$k > 0$	Oblate ellipsoid

Figure 7.2 shows five surfaces having different conic constants but the same curvature. Most of us are generally familiar with the surface shapes described. One surface we do not come across often is the oblate ellipsoid, sometimes called the oblate spheroid. This can be thought of as the shape of the Earth as it rotates about its axis. Due to centrifugal force, the diameter is greater at the equator than in the polar direction. The oblate ellipsoid has its foci orthogonal to the optical axis.

Conic surfaces, either reflective or refractive, are free of spherical aberration for one particular set of conjugate points. Let us look into a set of different conic surfaces. A spherical surface forms an aberration-free image if the object is at the center of curvature of the surface. An ellipsoid forms an aberration-free image for a pair of real image conjugates on the same side of the surface and a hyperboloid for conjugates on two different sides of the surface. A parabolic mirror forms a perfect image of a point for an axial object at infinity. This is the reason why parabolic mirrors (sometimes combined with hyperbolic mirrors) are widely used in astronomical optics.

When the object point is moved axially from the position of the aberration-free conjugate, a certain amount of spherical aberration is introduced. If the point is moved laterally, other aberrations, such as coma, astigmatism, and field curvature, contribute to image blurring.

Figure 7.2
Conic Surfaces with
the Same Curvature
and Different Conic
Constants



Application of Aspheric Surfaces in Reflective and Refractive Systems

Aspheric surfaces are widely used and often essential in reflective systems due to the small number of surfaces and typically large apertures. While a complex lens may consist of 18 spherical radii in order to minimize the aberrations, a reflective system can only have two surfaces in most cases. A simple spherical reflecting telescope suffers from spherical aberration and coma. A spherical mirror is shown in Fig. 7.3a. A point object at infinity is focused by the spherical mirror at a distance from the mirror equal to the one-half of the mirror radius, and this distance is the focal length. Third-order spherical aberration results, and the wavefront error or OPD is proportional to the $(\text{aperture})^4$, as described in Chap. 5.

A parabolic reflecting telescope is shown in Fig. 7.3b. This is a classic example of how spherical aberration can be corrected. An infinitely distant axial point is imaged to a perfect aberration-free image point. Unfortunately, the image quality degrades quickly when the object is moved off axis. Coma is the aberration that restricts the field of view of the parabolic telescope to a very small field.

A very common form of reflecting telescope is the two-mirror Cassegrain telescope, with both mirrors being conic surfaces. The *classical Cassegrain telescope* has a paraboloidal primary mirror and a hyperboloidal secondary mirror, as shown in Fig. 7.4. f_1 is the location of the

Figure 7.3
Conic Surfaces for
Reflecting Systems

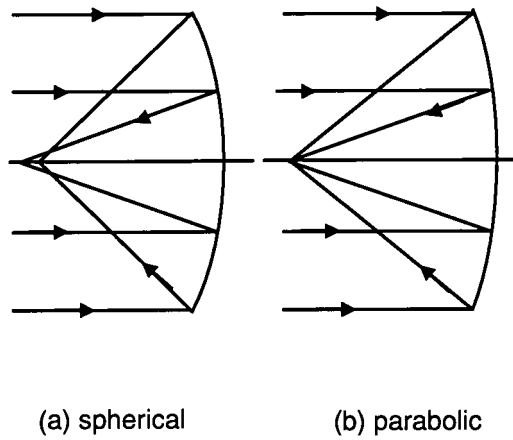


Figure 7.4
Cassegrain Telescope

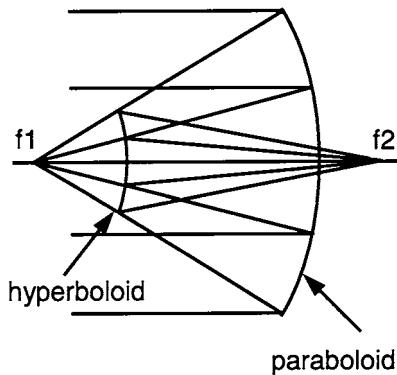
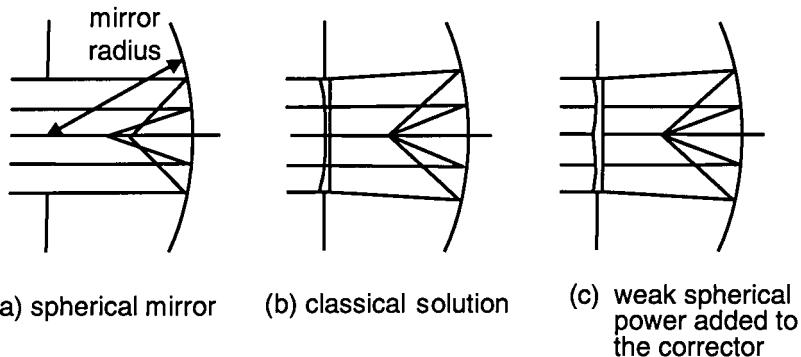


image which would be formed by the large primary mirror, and f_2 is the location of the image of the entire system. Note that the secondary mirror reimages f_1 to f_2 . A similar configuration, called a *Ritchey-Chrétien* Cassegrain telescope, has both primary and the secondary hyperbolic mirrors. The classical Cassegrain performance is limited by off-axis coma while the *Ritchey-Chrétien* Cassegrain is, in effect, a coma-free Cassegrain. Its limiting aberration is astigmatism.

Another well-known type of telescope is a Schmidt telescope, which is shown in Fig. 7.5. It consists of a spherical mirror with an aspheric corrector plate located at the center of curvature of the mirror. Third-order spherical aberration results in a wavefront aberration function proportional to the (aperture)⁴. If the aperture stop is located at the center of curvature of the mirror, there is symmetry for all field positions. Apart from the aperture stop being obliquely viewed by the

Figure 7.5
Schmidt Telescope



oblique bundle of off-axis rays, the oblique rays are focused in the same manner as the axial bundle. The chief ray is normally incident onto the mirror everywhere in the field of view. The image is formed on a spherical surface, with the image radius equal to one-half the mirror radius.

Without any aberration correction, the rays that are closer to the aperture edge are focused closer to the mirror than the paraxial rays. The wavefront distortion, which is proportional to the fourth power of the aperture radius, can be corrected with a wavefront distortion of the opposite sign introduced by the aspheric corrector plate placed in the aperture stop as in Fig. 7.5b to provide effective “parabolization” of the spherical mirror. A fourth-order aspheric deviation from the flat base surface of the glass corrector introduces a negative fourth-order wavefront distortion. The aspheric refractive corrector reduces the spherical aberration. However, some chromatic aberration is introduced by the wedged shape at the outer periphery of the corrector, close to the aperture edge. In order to minimize this chromatic aberration, a very weak positive power is added to the corrector, such that the corrector has zero power at 0.7 of its aperture, as shown in Fig. 7.5c. This shape of the corrector is not only the optimum shape for the correction of spherocromatism (the variation of spherical aberration with wavelength), it is also best suitable for manufacturing.

The majority of optical systems are based on the use of spherical components because they are easier to manufacture. However, there are cases where aspheric optical components have a significant advantage over spherical ones. In astronomical optics, reflective aspheric components are widely used. Today, with the significant development of new plastic materials, low-cost molded aspheric refractive optics are finding their place in the large consumer market. Precision diamond grinding and compression molding of glass aspheric lenses is also becoming more common.

Refractive aspheric lenses are widely used in many kinds of illumination systems, from the condensers in projection systems and microscopes, to street lamps and searchlights. Since in many cases these are not imaging systems, the manufacturing tolerances on these components are somewhat forgiving. In the case where the optical components are not exposed to the heat from the light source, aspheric optical elements can be injection molded. In projection systems, aspheric condensers are often molded glass lenses, mostly from B270 glass. B270 is a very common low-cost glass similar to BK7, which is used extensively in “float glass” for low-cost windows and mirrors. Glasses that are moldable have a

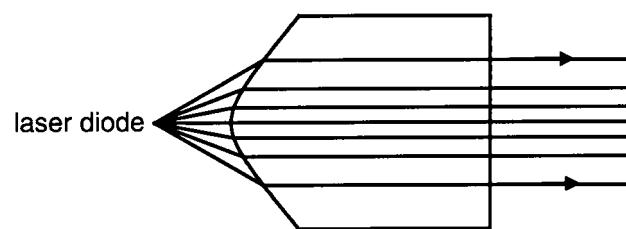
lower temperature at which they become soft than the standard optical glasses. Heat generated by the light source, or by the absorption of light by the optical components themselves, is often a severe problem in optical systems and requires the use of heat-resistant materials such as fused silica or fused quartz.

Another field where aspherics find their place is in systems for focusing of laser beams, for example, in CD players or data storage systems, when coupling light from a laser diode into a fiber, or when collimating light from laser diode arrays. These applications require high-precision optical components, as well as optically stable components in a given temperature range. The small size of these components makes them easier to mold. Optical plastics are used wherever they are acceptable, because of the much lower manufacturing costs. However, accurate lens shape and the very good temperature stability of glass aspheric lenses make them a better solution for applications where high precision is required. Glass aspheric lenses are manufacturable in diameters smaller than 25 mm. At diameters greater than 25 mm, aspheric glass lenses become too expensive for high-volume manufacturing. Single-glass bi-aspheric lenses are used for focusing or collimation of $NA = 0.5$ laser beams with diffraction-limited performance.

An infinitely distant object imaged through a planohyperbolic lens is focused to an aberration-free spot. This feature is used in the case of a planohyperbolic fiber lens (the cross section of the fiber has a planohyperbolic shape) to collimate the fast axis of the laser diode arrays, as shown in Fig. 7.6. The distance of the laser diode from the lens is determined by the index of refraction of the lens.

Improvements in the quality of injection-molded plastic optics make the use of them possible in camera lenses and projection lenses. The optical design of these kinds of lenses is difficult because many parameters have to be considered such as

Figure 7.6
A Planohyperbolic
Collimator



Spherical and Aspheric Surfaces

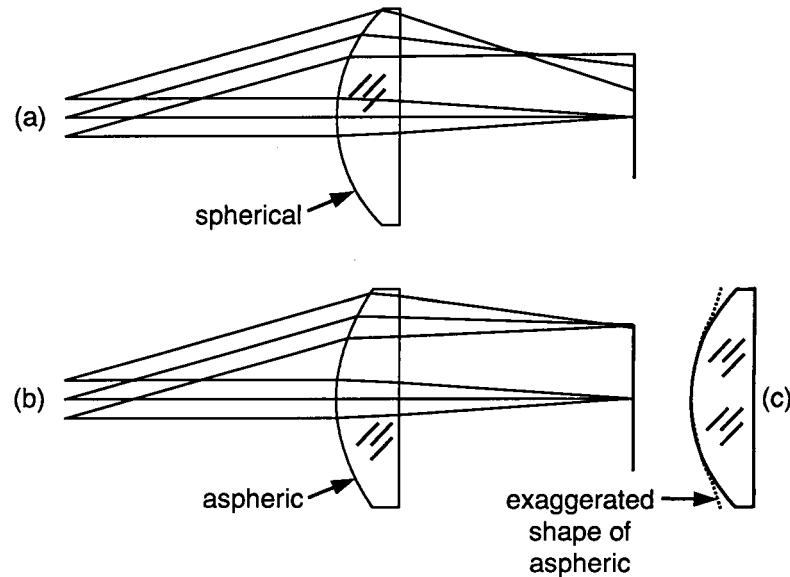
- The location of the component in the lens to minimize the beam size over the plastic component.
- The shape of the lens to keep the molding parting line inside the component.
- Athermalization of the lens.

In high-performance lenses, plastic components should be away from the aperture stop because it is very difficult to achieve diffraction-limited performance with injection-molded plastic components. However, they can be extremely useful in the correction of field aberrations such as astigmatism, field curvature, and also distortion.

Much of the discussion thus far with respect to aspheric surfaces and their benefits has related to the correction of spherical aberration. If an aspheric surface is located at or near the aperture stop of a system, it will primarily affect or benefit spherical aberration, which is an axial aberration which, for the most part, carries across the field of view. As aspheric surfaces are located further from the stop, they can help to minimize some or all of the off-axis aberrations such as coma and astigmatism.

A good example of the application of an aspheric surface used for astigmatism correction is shown in Fig. 7.7. In Fig. 7.7a, we see a single-element lens with its aperture stop located far to the left of the lens. If

Figure 7.7
Correction of Astigmatism with an Aspheric Surface



the curved lens surface is spherical, the oblique rays create a footprint on the surface, which is larger in the plane of the figure than the orthogonal plane in/out of the figure. As we discussed in Chap. 5, this tends to refract and pull the rays in the plane of the figure inward from where they would otherwise focus. We now need to ask ourselves what would it take to push the focus position outward and compensate for this inward focus shift. The answer is to create a more negatively powered surface in the plane of the figure at the outer periphery of the lens. This is shown in proper scale in Fig. 7.7b and in an exaggerated form in Fig. 7.7c. This more negatively powered surface shape in the plane of the figure has virtually no effect in the orthogonal plane, hence the highly efficient correction of astigmatism by the aspheric surface.

Another common use of aspherics is in the thermal infrared where the cost of materials is extremely high. With the use of aspherics, the number of elements can be reduced to a minimum.

Guidelines in the Use of Aspheric Surfaces

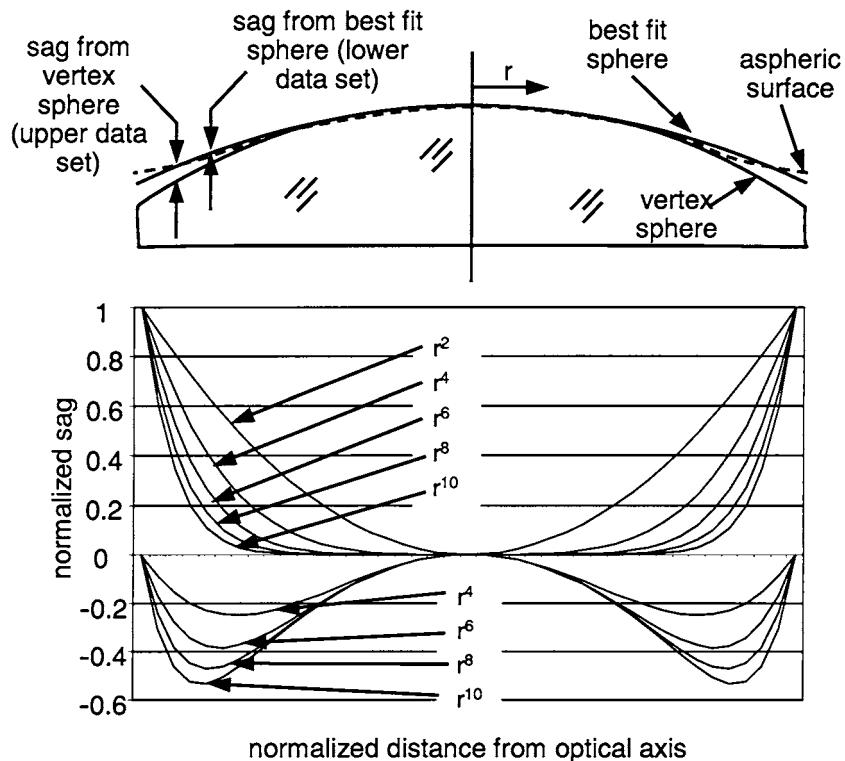
The proper usage of aspheric surfaces is extremely important. This includes which surfaces to make aspheric and whether to use a conic section or, alternatively, a higher-order aspheric. The conic sections include paraboloids, hyperboloids, and ellipsoids, as discussed earlier in this chapter. The higher-order terms are surface departures from conic, which are proportional to r^4, r^6, r^8, r^{10} , and so on, where r is the radial distance from the optical axis. The simpler forms of reflective systems, such as the classical Cassegrain (paraboloidal primary, hyperboloidal secondary) and the Ritchey-Chrétien Cassegrain (two hyperboloids), were discussed earlier. The classical Cassegrain is limited by coma and field curvature, and the Ritchey-Chrétien is, in effect, a coma-free Cassegrain which is limited by astigmatism and field curvature. Once the basic system is set up on your computer, varying the appropriate conic constants is all it takes to reach a viable solution.

But how do we decide which surface, or surfaces, in a lens system should be made aspheric, and how do we decide what form of aspheric to use? To answer this question, first consider Fig. 7.8, where we show:

- (1) the aspheric surface departures from the base spherical surface and
- (2) the departures from what we call the “nearest sphere” or “best-fit

Spherical and Aspheric Surfaces

Figure 7.8
Aspheric Sags from
Vertex Sphere (Top)
and Best-Fit Sphere
(Bottom)



sphere" to the aspheric surface. The nomenclature is shown in the top of Fig. 7.8 with accurate data below.

If we now compute for our baseline spherical optical system a plot of the optical path difference, we should look for a form matching the basic profile or character of these data. For example, if the axial OPD plot resembles the form of the sag from the nearest sphere for the r^6 case, then varying the r^6 coefficient on a surface near the aperture stop will likely be beneficial. If we find a sharp increase or decrease in the OPD off axis at the edge of the pupil, then varying a higher-order term or two on a surface away from the stop will likely be beneficial. There are some basic guidelines, and these are listed here:

1. Conic surfaces can be used for correcting third-order spherical aberration and other low-order aberrations.
2. If you have a nearly flat surface, then use an r^4 and higher-order terms rather than a conic.

3. If you have at least a somewhat curved surface, then you can use the conic along with higher-order terms if required.
4. It is generally best not to use both a conic and an r^4 surface, as they are mathematically quite similar. This is because the first term of the expansion of a conic is r^4 . While they can both literally be used, the optimization process often tends to beat one against the other, yielding artificially large coefficients, and this may have an effect on the convergence of the optimization.
5. Use aspherics beginning with the lower-order terms and working upward as required. If you can stay with conics, this may make testing more manageable. You should be able to assess the need for adding terms based on the character of the OPD plot.
6. It is very dangerous to use a large number of aspheric surfaces, especially with higher-order terms. This is because they will beat against each other. This means that as one surface adopts a certain aspheric profile or contour, it may increase in its asphericity, with its effect cancelled by adjacent surfaces. For example, if the first of two closely located aspheric surfaces has significant surface departure from sphericity, the neighboring aspheric surface could very likely cancel this effect. While the lens may perform well on paper, we now need to manufacture two highly aspheric surfaces, a difficult and expensive task which may not be necessary.
7. If possible, optimize your design first using spherical surfaces, and then use the conic and/or aspheric coefficients in the final stages of optimization. This may help in keeping the asphericities to a more manageable level.

Specification of Aspheric Surfaces

It is important to specify an aspheric surface sufficiently enough to convey to the shop both what you want and what you need. The following items are most often included in specifying aspheric surfaces:

1. The surface to be aspheric is labeled aspheric on the component drawing.
2. You should include an equation of the surface shape along with the aspheric coefficients. A small sketch indicating the nomenclature and sign convention is recommended.

3. A table listing the sag as a function of the radial distance from the surface vertex normal to the optical axis, r , is imperative. You should list a sufficient number of data points to adequately sample the surface profile.
4. You should list how close the actual surface must come to the ideal design prescription. The form of this can be "surface to match nominal surface to within four visible fringes (or 0.001 mm) over clear aperture."
5. You may need to call out higher-frequency surface irregularities and/or surface finish. The higher-frequency irregularities can be called out by indicating the maximum slope departure from nominal over the surface. Surface finish is normally called out by indicating the rms surface finish, in nanometers. This latter callout is generally used for diamond-turned surfaces, where surface roughness is sometimes a problem or where scattering and off-axis rejection is of major concern such as in space telescopes.
6. You should, if possible, indicate the form of testing to be used.

Do keep in mind that the more callouts you list and the more extensive the testing, the more costly your optics will be, and they will likely take longer to manufacture. Your callouts should indicate only what you need functionally for your system to work properly.

This page intentionally left blank

CHAPTER 8

Design Forms

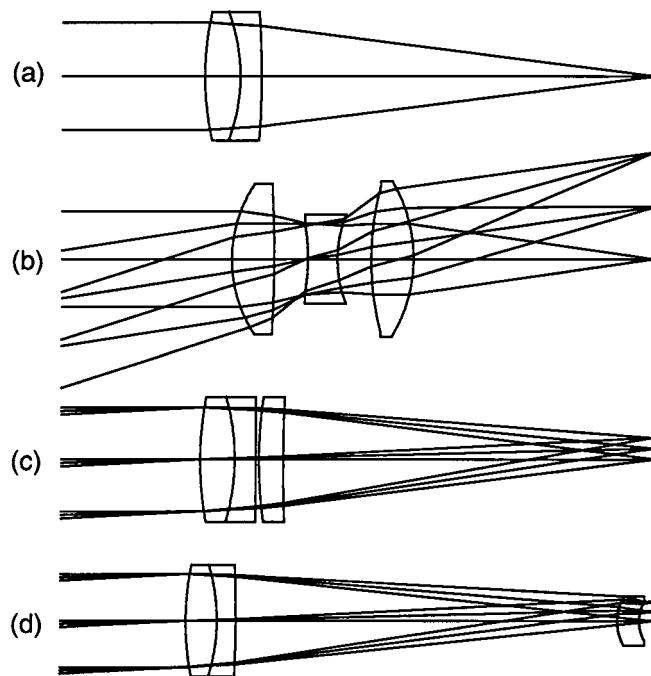
Introduction

In this chapter, we will discuss how we select the proper design form or configuration for both refractive and reflective image-forming systems. We will also consider fold mirrors and prisms since they have a significant influence on the system design configuration.

The proper system design form or “configuration” of an optical system is generally the key to a successful design effort. The term “configuration” here means the basic form of the system which includes not only the number of elements, but also the relative optical power and distribution of the elements within the lens system. For example, an achromatic doublet of two cemented elements, as shown in Fig. 8.1a is clearly different in form from a Cooke triplet, which consists of three separated elements, as in Fig. 8.1b, with two outer positive crown elements and a negative flint element at the center. The Cooke triplet can be used over wider fields of view than a doublet due largely to a reasonable degree of symmetry fore and aft of the central element, which is at or near the aperture stop. The doublet and triplet are very different configurations.

What if we were to add a single positively powered element immediately following a cemented doublet, as in Fig. 8.1c? Would the lens configuration be called a triplet? It certainly would not be a Cooke triplet as the symmetry is not present. This is a very different configuration or design form from a Cooke triplet. However, it is, of course, a three-element lens. The same is true for Fig. 8.1d, where we again have three elements, only here the third element is very near to the image and is serving to both flatten the field as well as correct astigmatism. All three of the

Figure 8.1
Doublets and Triplets



three-element configurations are quite different in configuration or form, each with their relative advantages.

The selection of an optimum configuration prior to initiating a design effort provides the starting point from which the design optimization proceeds. While lens design software has improved significantly over the years, the programs are rarely capable of changing configurations, and never add or delete elements. Most of the time the program will reach an optimum or local minimum in the error function for the input configuration. For more information on the optimization process, see Chap. 9. The configuration selection is driven by many factors. Nearly every system specification can have an influence on the configuration. The major factors influencing the configuration selection are:

- Field of view
- Performance requirement
- $f/\#$
- Packaging requirements
- Spectral range

System Configurations for Refractive Systems

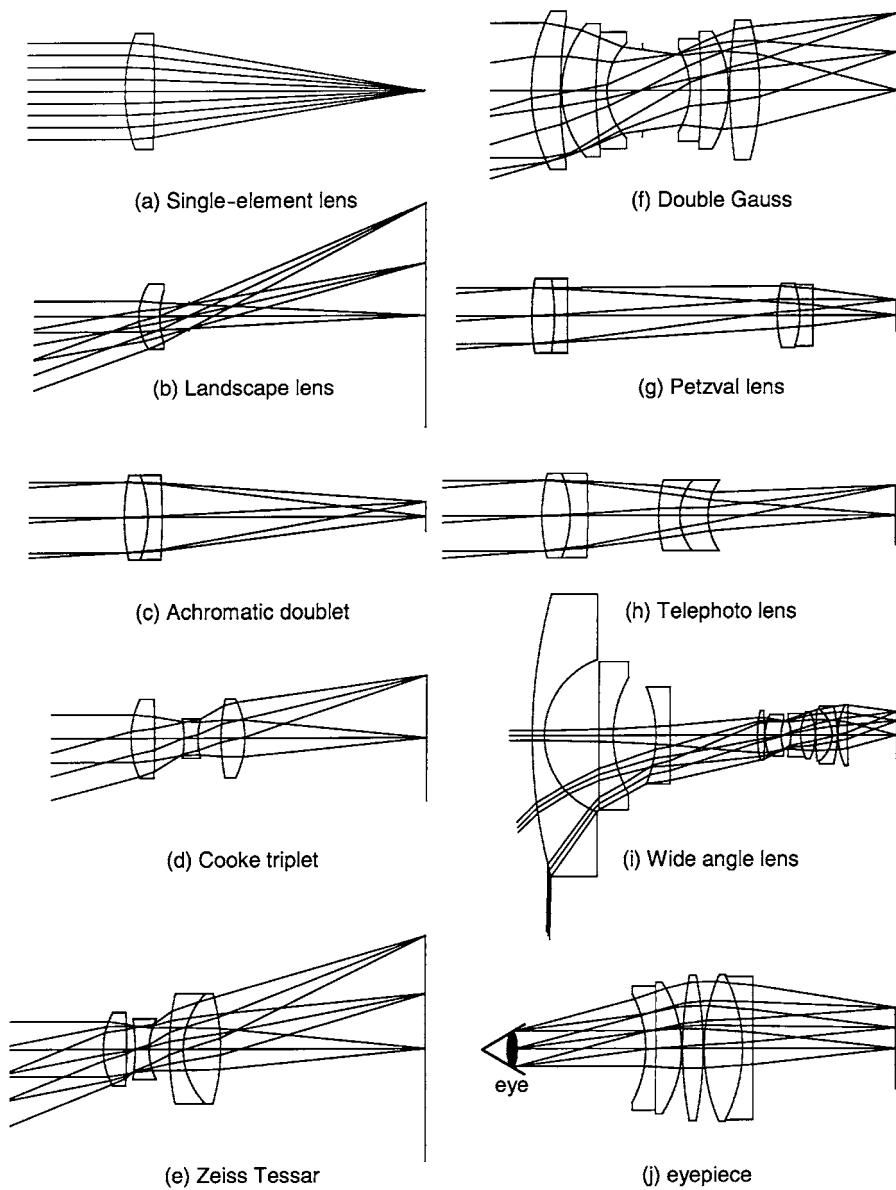
We will review a progression of configurations for lenses in order to illustrate just what differentiates one from another. The following configuration forms are shown in Fig. 8.2.

SINGLE-ELEMENT LENS (FIG. 8.2a) A single-element lens has generally poor image quality and a very small field of view. Further, it suffers from chromatic aberrations and it can only be used at a high $f/ \#$. We often think of a single element as of “magnifying glass quality.”

It is important to note that the performance, which may be poor for one application, may be just fine for another. For example, we have stated that a single element generally has poor image quality. This is, for the most part, true for most critical imaging applications such as camera lenses, machine vision optics, and other similar applications. However, if you are looking for a photon collector with little or no image quality requirements, then a single element may be quite adequate for the task. Another good example is the optics used for optical data storage and other microoptics applications. For data storage applications, a laser diode is imaged to a micron or submicron spot diameter. Since the laser is nearly monochromatic (there may be thermally induced shifts in wavelength), the field of view is nearly zero and the scale or size of the system is extremely small, we often find that a single aspheric element is sufficient for the task.

LANDSCAPE LENS (FIG. 8.2b) While a landscape lens is also a single element, it has an aperture stop which is remote or separated from the lens itself. Further, the lens is bent or “curled” around the stop for symmetry reasons. This reduces the angles of incidence on the surfaces and thereby reduces off-axis aberrations. Earlier in Chaps. 3 and 5 we discussed in greater depth how minimizing angles of incidence within a lens system reduces aberrations. A landscape lens can have its aperture stop either aft of the element as shown or in front of the element. It can be shown that the aberrations are somewhat reduced if the stop is in front of the lens, and many early box cameras were constructed this way. The landscape lens has chromatic aberration as well as residuals of many of the other third-order aberrations.

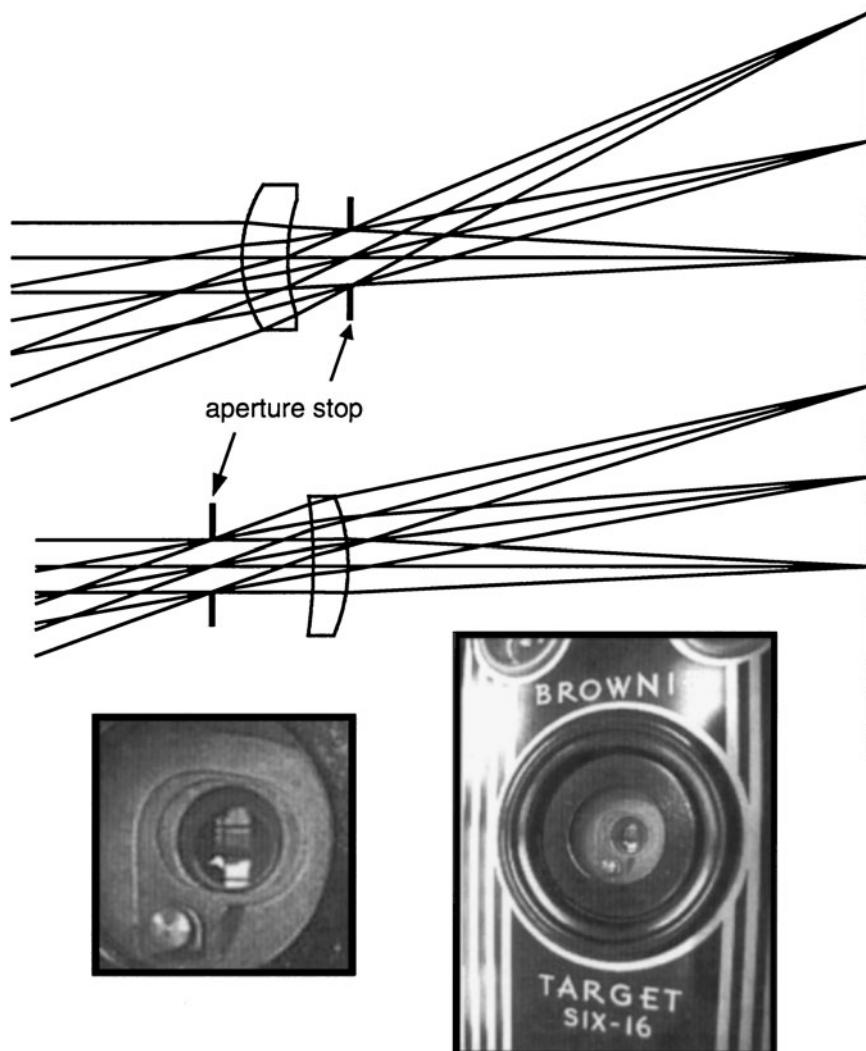
Figure 8.2
Progression of
Configurations



Design Forms

We show in Fig. 8.3 two forms of landscape lenses, one with the stop aft of the lens and the other with the stop forward of the lens. It can be shown that the performance is slightly improved with the stop in front. Also shown are two photos of the front of an early Kodak Brownie camera with a flat window followed by its aperture stop and finally the lens, which is aft of the stop.

Figure 8.3
Landscape Lens with
Stop Aft and Forward
of Lens



ACHROMATIC DOUBLET (FIG. 8.2c) The achromatic doublet is capable of bringing the red and blue wavelengths to a common focus, with the central green or yellow wavelength defocused slightly toward the lens. A typical achromatic doublet has a blur diameter approximately 25 times smaller than an equivalent single-element lens (based on an $f/5$ lens in the visible spectral band).

A cemented doublet performs well only over a small field of view, and it cannot be used at low $f/\#$ s due to higher orders of spherical aberration. In order to balance the inherent third-order spherical aberration of the cemented doublet, one can introduce a small airspace between the elements. This airspace will permit the balancing of fifth-order spherical aberration with the inherent third-order aberration for an improved overall level of performance. Further improvement can be realized by adding an additional element near the image, which can be used as a field flattener, and often it is possible to bend this element to balance and eliminate some or most of the astigmatism.

COOKE TRIPLET (FIG. 8.2d) This three-element lens form takes advantage of symmetry in order to minimize the angles of incidence of the rays as they proceed through the lens over the field of view, and hence it is capable of an acceptable level of performance for many applications. The Cooke triplet was first designed in England by H. D. Taylor at the “Cooke and Sons” optical company. The Cooke triplet is the first configuration we have presented in this review that allows for the optimization and balancing of the seven primary, or third-order, aberrations, as well as the control of focal length. There are eight “useful” variables in a Cooke triplet, the six radii and the two airspaces. Element center thicknesses are generally not of significant use in aberration control and, for the most part, element thicknesses that yield reasonable manufacturing ease are best. We can thus control or optimize the following:

1. Spherical aberration
2. Coma
3. Astigmatism
4. Axial color
5. Lateral color
6. Distortion
7. Field curvature
8. Control of the focal length

It is important to note that just because we have the same number of useful variables as the number of primary aberrations (along with the focal length), this does *not* at all mean that the aberrations can be brought to zero or even close to zero. What it means is that for the *f*/number and focal length is selected, the aberrations can be reasonably well balanced against one another, especially the third-order aberrations. Thus, for example, an *f*/6 Cooke triplet covering a 10° full field of view over the visible spectrum will likely be capable of a reasonable level of performance. However, an *f*/1.4 Cooke triplet covering a 30° full field of view will probably provide fair to poor performance, at best. The low *f*/number will lead to significant spherical aberration residuals, and the wide field of view will lead to coma, astigmatism, and other off-axis aberrations.

ZEISS TESSAR (FIG. 8.2e) The Zeiss Tessar is derived from, and is an improvement upon, the Cooke triplet. Paul Rudolph of Zeiss Jena replaced the original single rear lens in the Cooke triplet with a doublet lens, resulting in a better lens performance, with higher resolution, excellent contrast, and very low levels of distortion.

There is a rule of thumb in lens design which says “clip it in the bud.” What is meant here is that the best place to correct or eliminate aberrations is as close to where they are being introduced as possible. In the case of a Cooke triplet we see that the first positive lens element takes collimated light from infinity and bends or converges it into the second negative lens element. The second element takes the slightly converging light and diverges it into the third and positive element. And finally, the third positive element takes the slightly diverging light and bends the rays so as to create the required *f*/# of the lens. The amount of ray bending or redirection is greatest for the third element than for the first or second elements. Thus, the aberrations introduced by the third element will be greatest as well. This makes the third element an excellent candidate to convert to a doublet. As one of the more difficult aberrations to correct in a Cooke triplet is axial color (change in focal length with wavelength), making the third element a doublet (as in the Tessar) allows for a superior level of correction of this as well as other residual aberrations. Tessar designs can be effectively used at *f*/numbers down to *f*/4.5 or somewhat lower, depending on the relative mix and level of lens requirements. Further, as will become apparent, the Cooke triplet forms the basis of many more complex and high-performance configurations.

DOUBLE GAUSS (FIG. 8.2f) The double Gauss lens is yet a further extension toward improved performance at lower $f/\#$ s and wider fields of view. If we summarize what we have learned thus far, we can see more clearly the evolution of the double Gauss lens. The methodologies learned thus far include *splitting optical power* to minimize aberrations, using *negatively powered elements with smaller beam diameters* for field curvature correction as in the Cooke triplet, and using *symmetry fore and aft of the aperture stop* also for symmetry reasons, thereby minimizing the angles of incidence on the lens surfaces. As we will learn later, symmetry also allows for cancellation of several off-axis aberrations. The double Gauss lens uses at least two negatively powered elements near the stop and two or more positive elements on the outsides. Furthermore, there is a reasonably high level of symmetry surrounding the aperture stop.

The double Gauss lens is capable of good performance down to about $f/1.4$ and even lower. Indeed, there have been several $f/1.0$ double Gauss lenses in the 35-mm camera marketplace, including the famous *Nocitlux* designed by Walter Mandler of Ernst Leitz Canada. As with any lens, there are compromises, and at $f/1.0$ the lens is hardly diffraction limited; however, from an overall performance standpoint and light-gathering capability, the lens is a top performer. We should note that for the more demanding levels of performance (such as the Nocitlux), higher refractive index materials are often used as they can significantly reduce aberrations, as well as anomalous dispersion glasses for superior chromatic aberration correction. Glass selection is discussed in Chap. 6.

PETZVAL LENS (FIG. 8.2g) The Petzval lens represents a very different design philosophy. This lens is intended for smaller fields of view and only moderate $f/\#$ s such as $f/3.5$ or slower. The design philosophy here is to use two separated doublets with the power task shared between the two. This yields lower secondary chromatic aberrations than a single doublet of the same net $f/\#$. This form of design is used for high-performance small field-of-view lenses as one might encounter in aerial reconnaissance, for example. As noted, the Petzval lens is not well suited for wide-angle applications, as there is little opportunity for symmetry as with the Cooke triplet or the double Gauss.

TELEPHOTO LENS (FIG. 8.2h) The telephoto lens is a positive group of elements, followed by and separated from a negative group of

Design Forms

elements. As we showed earlier in Fig. 8.1, this form of lens has a focal length longer than the physical length of the lens, hence the name “telephoto.” The ratio of the physical length to the focal length is called the *telephoto ratio*. In the case of a fast lens (low *f*/number), with a low telephoto ratio below 0.6, for example, the lens configuration becomes quite complex. In the limit, if the light exits the second group collimated, we have a beam contractor or, in effect, a Galilean telescope, and the focal length is infinite. It is important to note that both the positive and the negative groups generally need to be separately achromatized in order to produce a complete lens with sufficiently low chromatic aberration.

It is interesting to think about the use of the words “telephoto lens.” To a photographer a telephoto lens is generally a lens whose focal length is longer than a standard lens for a specific film format. For example, the standard focal length for a 35-mm camera lens is in the order of 50 to 55 mm. A lens with a 100- to 135-mm focal length or longer is generally considered to be a telephoto lens. This is because objects appear closer due to the smaller field of view covered by the longer focal length lens. An achromatic doublet with a focal length of 135 mm might, to a photographer, also be a telephoto lens because it brings the object closer. However, to a lens designer a telephoto must be of a form or configuration as shown using a positively powered front group and a negatively powered rear group so that the focal length is longer than the physical length of the lens.

WIDE-ANGLE LENS (FIG. 8.2*i*) A lens covering a substantially wider field of view than a normal lens (in photography, for example) is called a *wide-angle lens*. Thus, with a standard focal length of 50 to 55 mm in 35-mm photography, a wide-angle lens is generally considered to be 35 mm or less. In order to cover the wider fields of view, we often use a strong negatively powered front element or group of elements to bend the rays outward to cover the wider field angles. In order to still image from infinity, the light in the space between the main body of the lens and the front negatively powered group needs to be converging toward the object as seen in Fig. 8.2*i*.

We show the wide-angle lens as having a negatively powered three-element front group with a multielement configuration for the prime lens group. What is happening here is that the prime lens group itself, is covering a smaller field of view, with the field angle increase happening only at the negatively powered front elements.

EYEPIECE (FIG. 8.2j) An eyepiece is used to visually view and magnify the image from a microscope objective or a telescope objective, or alternatively to view a display such as in a head-mounted display system. The eyepiece is a very different configuration of lens system in that the aperture stop is not only quite remote from the main part of the lens, but it is in reality the pupil of the eye. Eyepieces are normally designed by tracing rays from the eye, which is the aperture stop, to the image plane, as shown in Fig. 8.2j.

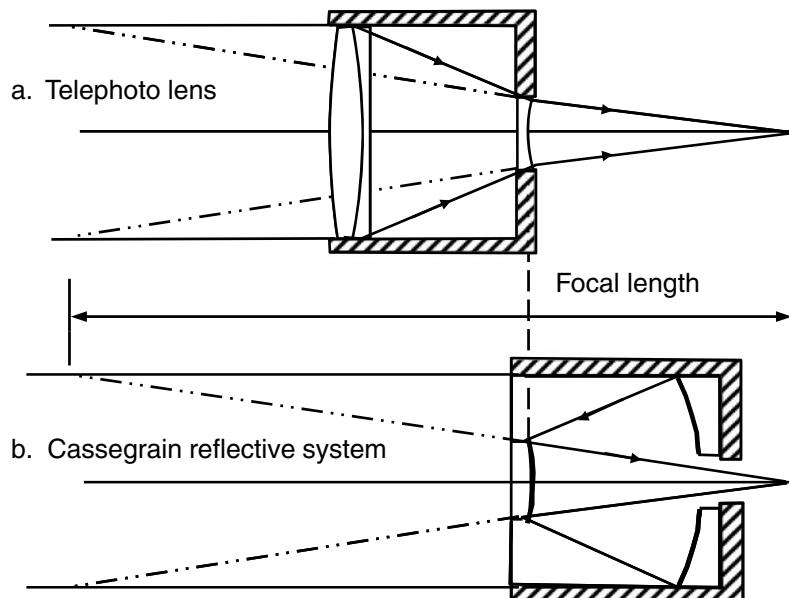
As can be seen in Fig. 8.2j the rays at the extreme field of view are primarily using the outer periphery of the lens elements. Unfortunately, this often results in significant amounts of astigmatism, lateral color, coma, and distortion. These field aberrations can be quite significant and difficult if not impossible to correct for in extremely wide field-of-view eyepieces with 60° field of view or more. The careful use of higher-index glasses combined with one or more aspheric surfaces (if possible) can help to mitigate the problems to some extent. Eyepieces represent a very different configuration form than other lenses discussed thus far.

System Configurations for Reflective Systems

As with lenses, reflective or mirror systems can be of many varied configurations. There are, however, some inherent and very basic differences between refractive or lens systems and reflective or mirror systems. Lens systems are most often straight through and use the full clear aperture of the entrance pupil, as shown in the simple telephoto lens example in Fig. 8.4a. Mirrors have the fundamental challenge that they get in each other's way as shown in the two-mirror Cassegrain system in Fig. 8.4b. The Cassegrain is the reflective analogy of the telephoto lens described earlier in this chapter. Note that the large objective lens shown as an achromatic doublet is the optical analogy of the large concave mirror generally called the *primary mirror*. Further, the smaller lens in the refracting telephoto lens is the analogy of the small convex mirror in reflective systems, and it is generally called the *secondary mirror*. Note that both lens and mirror systems are in the form of the telephoto lens and the focal lengths are in fact equal in both systems.

Design Forms

Figure 8.4
Telephoto and
Cassegrain
Configurations



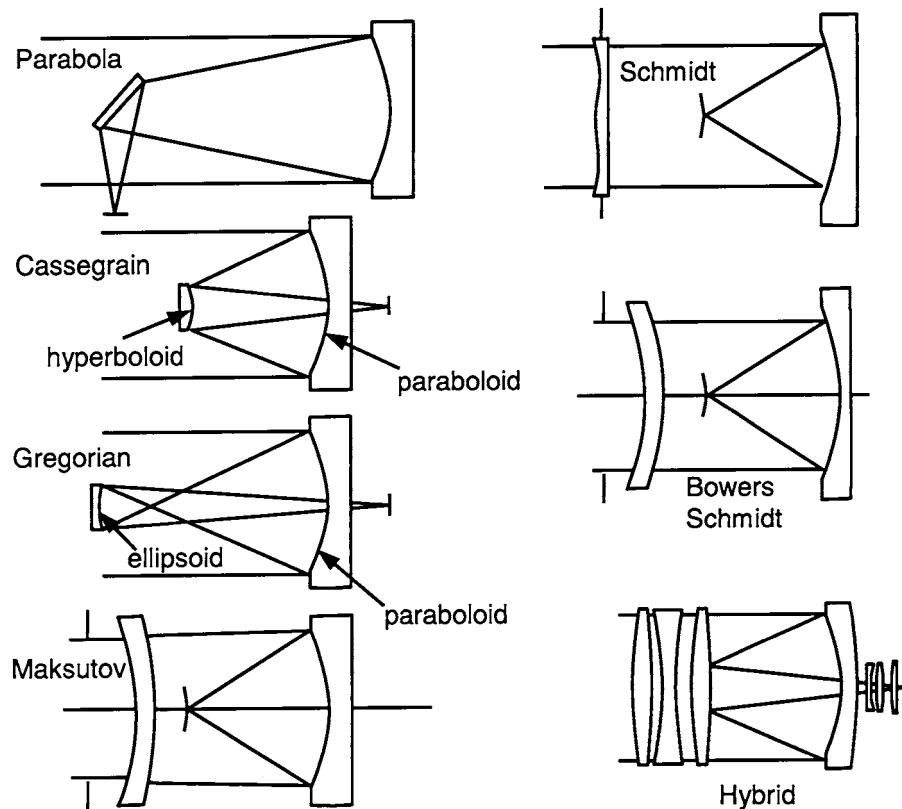
The fundamental issue of the mirrors getting in each other's way in reflective systems leads to many significant differences between reflective and refractive systems.

We will review a progression of configurations for reflective systems in order to illustrate just what differentiates one from another. The following configuration forms are shown in Fig. 8.5.

PARABOLOID A single parabolic mirror has zero spherical aberration on axis for an infinitely distant object. It is, however, limited by coma off axis. As shown, the image-forming light is often folded out to the side via a tilted flat-fold mirror sometimes called a diagonal mirror. In this case it is called a Newtonian telescope.

CASSEGRAIN The Cassegrain form of reflective system is perhaps the most common. In its "classical" design form, a parabolic primary mirror and a hyperbolic secondary are used. With this prescription, coma is the limiting aberration and is the same as a single parabolic mirror of the same f /number. An improved level of performance is achieved by allowing the primary mirror to be hyperbolic along with a hyperbolic secondary. This solution is called the Ritchey-Chrétien form

Figure 8.5
Reflective
Configurations



of Cassegrain. It is, in effect, a coma-free Cassegrain and is limited only by astigmatism and field curvature. The layout itself is virtually identical for the classical Cassegrain and the Ritchey-Chrétien forms. In most configurations the Cassegrain has an inward or concave curving image field due to the Petzval contribution of the convex secondary mirror predominating over the concave primary mirror.

GREGORIAN A concave parabolic primary mirror with a concave elliptical secondary mirror is the Gregorian form of telescope. In effect, the elliptical secondary mirror reimages the image formed by the primary mirror to its final position aft of the primary, as shown in Fig. 8.5. The Gregorian is not as common as it was some years ago, and the reason for this is that astronomers years ago did not believe that convex aspheric mirrors (as required by the Cassegrain) could be effectively

Design Forms

fabricated and tested. With the advent of interferometry and other advances in optical metrology, testing convex aspheric surfaces became more viable and the more compact Cassegrain is now more widely used.

MAKSUTOV The Maksutov uses a spherical primary mirror with a spherical weakly powered meniscus glass corrector plate to balance the spherical aberration of the mirror. This system has been popular with amateur astronomers for many years. A variation known as the *Maksutov-Cassegrain* locates the corrector closer to the primary, and an aluminized spot at the center of the convex surface acts like a Cassegrain secondary mirror, hence the name Maksutov-Cassegrain. This is the design form used in the well-known Questar telescope for many years.

SCHMIDT The Schmidt system uses a thin aspheric corrector plate located at the center of curvature of a spherical primary mirror to effectively correct all orders of spherical aberration. The aperture stop is at the corrector plate, and is located at the center of curvature of the primary mirror. Due to this geometry, the chief rays at all field angles will be incident onto the primary perpendicular to its surface. To third order, the spherical aberration will be eliminated at all field angles. The image is formed on a spherical image whose radius is equal to the focal length of the primary mirror. Further, the Schmidt system works quite well at low *f*/numbers, even as low as *f*/1 or less.

The fact that most aberrations are zero is due to what is known as the "Schmidt principal," which has as its basis the aperture stop being located at the center of curvature of a spherical mirror. It can be shown that in addition to zero spherical aberration, there is no third-order coma, astigmatism, or distortion. The major residual aberration of tangential oblique spherical aberration is due to the ray obliquity on the corrector plate as well as the foreshortening of the entrance pupil at off-axis field angles. The Schmidt principal is an exceptionally powerful technique, which has been successfully applied in the design of many well-corrected optical systems. It is especially useful in wide-angle applications. A variation of the Schmidt, invented by Baker, uses three separated aspheric Schmidt plates for minimization of the off-axis residual aberrations.

There is a variation of the Schmidt system known as a *shortened Schmidt Cassegrain*. In this configuration, we move the corrector plate closer to the primary mirror to a location where the ray bundle diameter is similar to what it would be in a Cassegrain system. Then a convex mirror is mounted on the interior of the corrector plate and the system

geometry now resembles a Cassegrain. The fact that the corrector plate and the aperture stop are no longer at the center of curvature of the spherical primary mirror are cause for coma and other off-axis aberrations; however, the overall performance is reasonably good for small fields of view. This form of system is extremely popular in contemporary amateur telescopes due to its robust performance combined with aggressive packaging.

BOWERS SCHMIDT This configuration uses a thin corrector plate, which is a shell concentric about the aperture stop. The aperture stop is at the center of curvature of the primary mirror as with the Schmidt telescope. The net result is a level of performance that comes close to the Schmidt telescope but without aspheric surfaces. If an aspheric corrector plate is now located at the aperture stop the aberration correction becomes incredible. This is because the shell corrector is concentric about the stop and only the weak aspheric corrector suffers from obliquity effects.

HYBRID A hybrid system is a combination of several pure or classical solutions. What we have here is a combination of the following:

- A multielement corrector group of a diameter equal to the entrance pupil. This near zero power group of elements is typically three to five lens elements and can be of the same glass type with no chromatic aberrations because it is of zero net optical power. The primary purpose of the corrector group is to balance and cancel the spherical aberration of the spherical primary mirror.
- The bulk of the optical power is from the primary mirror as well as the secondary, which itself is an aluminized area on the aft surface of the corrector group.
- Finally, there is a field-correcting group just before the image plane. This group of elements can effectively flatten the field of view and correct the residual off-axis aberrations such as coma and astigmatism.

The beauty of working with the hybrid system configuration described here is that each functional attribute of the system is quite independent and can be easily understood. The front corrector group corrects the spherical aberration of the primary mirror. The fact that it is of near zero power means that a single glass will produce a result essentially free

Design Forms

of chromatic aberration. And the rear field-correcting group can easily flatten the field curvature and simultaneously correct any residual off-axis aberrations, including coma and astigmatism.

UNOBSCURED APERTURE SYSTEMS Finally, there is a class of reflective optical systems generally known as “unobscured aperture systems.” These include the three-mirror anastigmat (TMA) as well as other forms of all reflective nonrotationally symmetric optical systems. These systems eliminate the performance degrading and difficult to support secondary mirror of the Cassegrain. However, we sometimes require nonrotationally symmetric aspheric surfaces which are difficult to manufacture and test and which are generally costly. Figure 8.6 shows a typical form of TMA. This configuration is an afocal telescope and is per US patent 5,173,801 by Cook. While the most obvious advantage of a TMA is in the hardware, especially as it relates to not needing to support a secondary mirror as in a Cassegrain form of system, the TMA can have significant advantages with respect to the suppression of unwanted diffracted light. Consider Fig. 8.7 where we show a system form developed for space applications. While there are variations on the basic theme, the essence of these systems is that they are used to image relatively close to a bright source as might be encountered in a space application when we are looking within several degrees of the sun. The sky is black, yet the intense solar radiation is just outside our field of view. In order to suppress the diffracted light, the aperture stop is reimaged to a location

Figure 8.6
Three-Mirror
Anastigmat

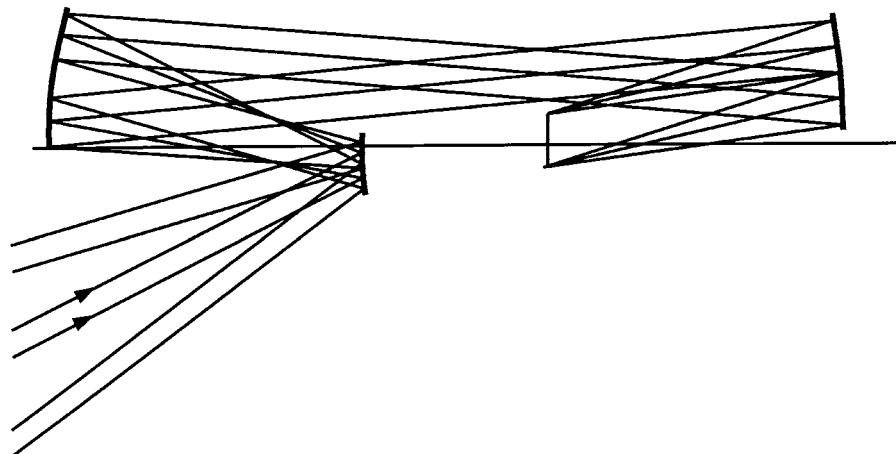
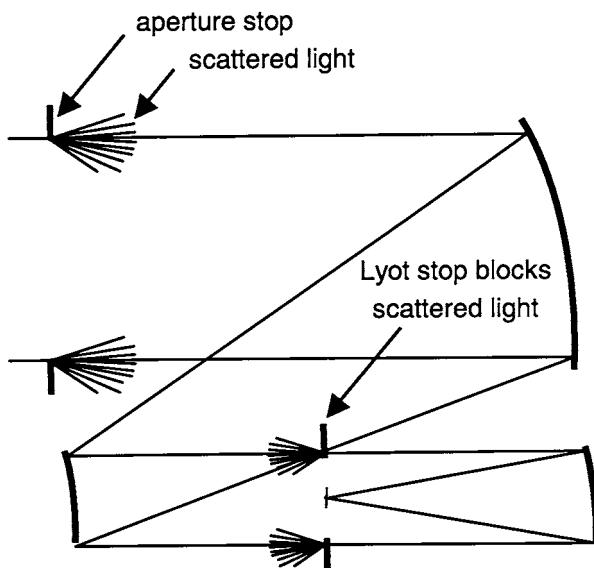


Figure 8.7

Three-Mirror All-Reflective System
Showing Lyot Stop
for Stray Light
Suppression



within the system where an aperture stop slightly smaller than that of the reimaged stop can be located so as to block the reimaged scattered light. This is known as a “Lyot stop” (pronounced “Leo,” after the French astronomer Bernard Lyot).

Reflective Systems, Relative Merits

- *No chromatic aberrations.* There are no chromatic aberrations whatsoever in all-reflecting systems. According to ray tracing theory and the use of Snell's law, the refractive index of a mirror is -1.0 for all wavelengths. For this reason reflective optics can be extremely well suited for multispectral applications or situations where refractive materials are either expensive or unavailable.
- *Central obscuration.* Since mirrors get in each other's way, there is often a central obscuration associated with reflective optical systems such as in the Cassegrain configuration. This obscuration affects the net photon throughput, affects the image contrast or MTF (Chap. 15), and is difficult to mount and align. Needless to say,

the support structure for this “secondary mirror” must be of a minimal obscuration to the incoming light, yet it must be strong and robust.

- *Aspheric surfaces required.* Due to the limited number of surfaces that can be effectively used in a reflective optical system, there are rarely enough surfaces to allow for minimization of aberrations as with refractive lens systems. For example, a typical double Gauss lens system may consist of seven elements. If there were no cemented elements, we would have 14 radii with which to minimize ray bendings and hence minimize the residual aberrations. In our Cassegrain reflective system configuration, we have only two surfaces, which are far too few for aberration control, thus leading us to require the use of nonspherical surfaces. One has only to ask the question of how a reflective system with 14 mirrors would look to appreciate the difficulty of working effectively with more than two to three mirrors. It is important to realize that aspheric surfaces are not necessarily bad. There are many contemporary methodologies for producing aspheric surfaces on both mirror as well as transmissive lens surfaces. For mirrors, we have ultraprecision machining or diamond turning as the most common.
- *Small number of elements.* The small number of elements together with the baffling issues, and the fact that the mirrors get in each other’s way, limits the field of view of reflective systems to be generally smaller than of the refractive systems.
- *Can be low weight.* Reflective systems can, in many cases, be made of aluminum, which is light in weight.
- *Inherently athermalized.* Reflective systems, if manufactured of a single material such as aluminum, are generally athermal. In other words, for a uniform temperature increase or decrease, the entire system expands or contracts by an amount dependent on the thermal coefficient of expansion. Since this is a uniform scaling of all system parameters, the image will still be in focus. If multiple materials and/or thermal gradients are used, then a careful assessment of the thermal properties of the imagery is critical. Zerodur is a glass material with almost zero coefficient of thermal expansion, and it is commonly used for large glass mirrors. The important message here is that reflective systems have the potential for being fully athermalized. To convince yourself that all reflective systems manufactured of the same material

throughout are indeed athermal, consider the following explanation: Assume that our design is a Cassegrain reflective system where all components, including the mirrors and the support structure, are aluminum. If we heat or cool the system uniformly then it will uniformly expand or contract according to the thermal coefficient of expansion of the material. This is, in effect, a scaling of the entire system. Imagine a drawing of our reflective system forming a perfectly focused image. Take this drawing to a copy machine and enlarge it by 20%. Now look at the drawing and ask yourself "is it still in focus?" Of course it is! Needless to say, if you have thermal gradients and/or different materials, your system may not be sufficiently athermal, and this may require active or passive athermalization.

- *Stray light susceptibility.* Reflective systems are often faced with problems associated with stray light. This stray light is often out-of-field light which directly or indirectly reaches the final sensor. The Cassegrain is an example of a system, which needs to be properly baffled to suppress unwanted stray light that may directly go through, missing both mirrors.

Refractive Systems, Relative Merits

- *Straight through.* The system operates straight through without any central obscuration. This results in a potentially higher photon efficiency with none of the degradations associated with a central obscuration.
- *Spherical surfaces, conventional manufacturing.* Since we can add lens elements and use the necessary techniques to minimize aberrations, we can most often use spherical surfaces and thus avoid expensive manufacturing methods often associated with aspheric surfaces.
- *Can add a lot of components.* This makes it possible to design high-speed systems with large fields of view.
- *Expensive materials and athermalization problems in the thermal infrared.* Refractive systems used in the thermal infrared (the MWIR which is the 3- to 5- μm spectral band and the LWIR which is the 8- to 14- μm spectral band) often require materials that are very expensive and have a high dn/dt . For example, germanium is extremely expensive, and furthermore it has a $dn/dt = 0.000396/\text{^{\circ}C}$.

Mirrors and Prisms

Mirrors and prisms are the optical components used in optical systems to

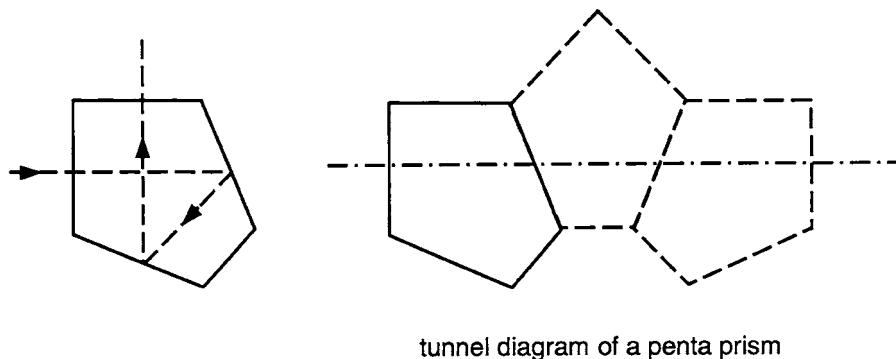
- Change the direction of light.
- Fold an optical system for better packaging.
- Provide a proper image orientation.
- Combine or split the optical beams using beamsplitter coatings.
- Disperse light with wedged prisms.
- Provide the means for interpupillary distance change in binocular systems.
- Expand or contract a laser's beam diameter, etc.

Flat mirrors fold the optical path in a system. Prisms also fold the optical path, except that the reflecting surfaces of the prisms behave like mirrors rigidly mounted with respect to each other. The optical designer has to be careful to leave enough space during the design of the optical system to place the mirrors and prisms where they are needed. Prisms have flat polished surfaces and have no optical power. If prisms are used for the proper image orientation and location, they use refraction at the input and output surfaces and reflection on the intermediate surfaces. Reflection is a total internal reflection if the incident cone of light is small enough and/or the magnitude of the prism angle is such that the condition for the total internal reflection is satisfied for all rays. Otherwise, surfaces are mirror coated.

Reflecting prisms are generally designed so that the entering and exit faces are parallel and perpendicular to the optical axis. This means that the prism can be represented as a plane parallel glass plate. The thickness of the glass plate is obtained by unfolding the prism around its reflecting surfaces. Unfolded prisms are shown in the form of a "tunnel diagram." The Penta prism with its tunnel diagram is shown in Fig. 8.8. We use a tunnel diagram so as to be able to use a single block of glass with no folding of the optical path during the design. The prism of refractive index, n , and the glass thickness, d (thickness of the unfolded prism), has its equivalent path length in air, also called the prism *apparent thickness*, and is equal to

$$d_{\text{apparent}} = \frac{d}{n}$$

Figure 8.8
Penta Prism



tunnel diagram of a penta prism

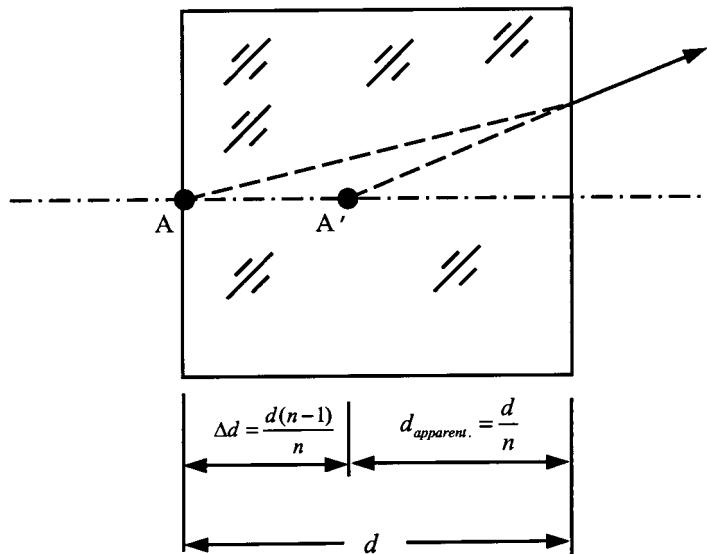
If the prism is inserted in a convergent beam, the image will be shifted by

$$\Delta d = \frac{d(n - 1)}{n}$$

Prism thickness and the apparent prism thickness are shown in Fig. 8.9.

In the case of mirrors, the optical designer has to leave sufficient space in the optical path for the mirrors to be mounted. Prisms do not introduce aberrations only if they are located in a collimated beam.

Figure 8.9
Prism Thickness and
the Apparent Prism
Thickness



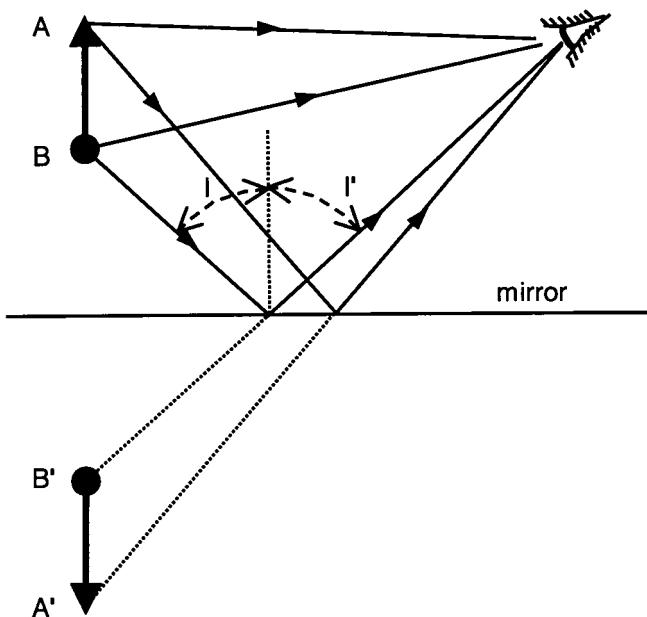
When the optical designer works on a system which contains, among other powered components, a prism of thickness d , he or she usually designs the system using a block of glass of thickness d because of the simplicity, smaller number of surfaces, and, consequently, faster ray tracing. In the case when the prism is located in a convergent or divergent beam, this block of glass introduces both monochromatic and chromatic aberrations, and has to be present during the optimization of the system. The whole process is iterative. The optical designer starts with the rough size of the prism, and optimizes the system with the prism equivalent block of glass. At some point in the design, the designer checks the diameter of the input and the output beam, as well as the ray angles through the prism, and makes the necessary adjustments to the prism size and positioning.

In a real system, the length of the prism along the optical axis is often significantly shorter than the glass equivalent plate, since the optical path in the prism is folded a few times. A good practice is to enter the real prism surfaces in the optical prescription when the design of the system is nearly finished and check the ray footprints on each prism surface as well as the optical performance of the whole system. It is also convenient to export a computer file with all the optical components in the system in a format that is suitable for the mechanical designer to design the mechanics around the optical components.

When a ray is reflected from a flat mirror, the incident ray, the normal to the mirror at the point of incidence, and the reflected ray, all lie in a single plane, which is called the *plane of incidence*.

Let us examine the orientation of the image after reflection from a flat mirror. If the observer looks directly at the object AB shown in Fig. 8.10, the point A appears to be the highest point of the object. If you look at the same object reflected from the mirror, it appears that the image is located behind the mirror in $A'B'$. However, this time it appears that point A' is at the bottom of the image. One reflection changes the orientation of the image in the plane of incidence, which means in our example that the image is upside down. If the object is a two-dimensional object, there is no change in the image orientation in the plane perpendicular to the plane of incidence. In our example, this means that the left side of the object after reflection is seen as coming from the left side. Let us imagine that object AB has no symmetry. It could be, for example, letter R. When viewed directly, the letter is oriented properly, and we can read it. After reflection in the mirror, letter R is upside down, and even if we rotate it around the direction of image

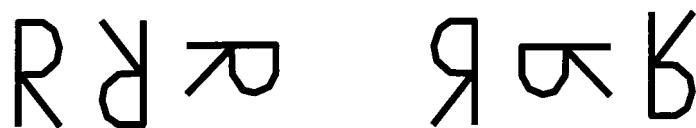
Figure 8.10
Image Orientation
after a Single Reflec-
tion from a Flat Mirror



propagation, we can never orient it so that it will be readable. In the case when the image is readable, which is shown in Fig. 8.11a, we call it a “right-handed image.” If the image is as shown in Fig. 8.11b, where letter R is backwards, regardless of the orientation of the image, it is called a “left-handed image.”

After multiple reflections off flat mirrors, or reflections inside a prism, a general rule says that an even number of reflections gives a right-handed image and an odd number of reflections gives a left-handed image. However, whether the number of reflections is even or odd does not tell us anything about the image orientation.

Figure 8.11
Right- and Left-
Handed Images



(a) Right-handed image

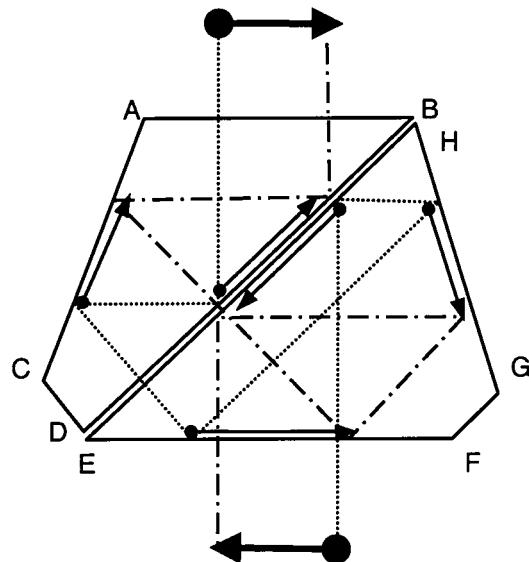
(b) Left-handed image

Reflecting prisms are used to:

- Erect the image in telescopes, which means that the top to bottom as well as the left to right are inverted, and the image is right-handed.
- Invert the image in one plane, either top to bottom or left to right. The image is left-handed.
- Deviate the optical axis, with inversion, erection, or no change in the image orientation.
- Displace the optical axis, with inversion, erection, or no change in the image orientation.
- Keep the unchanged image orientation, rotating the prism around the optical axis (these are prisms with no axis deviation or displacement) while the input image rotates around the center of the field of view (or around the input optical axis).

As an example of an inverting prism, we will look at the *Pechan prism*. A Pechan prism is shown in Fig. 8.12. It consists of two prisms with a small air gap between them. There are a total of five reflections, which means that a right-handed image entering the prism is changed to a left-handed image after the prism. All five reflections have a common plane

Figure 8.12
Pechan Prism



of incidence, and the image is inverted in that plane. In the direction which is normal to the common plane of incidence, there is no change in the image orientation.

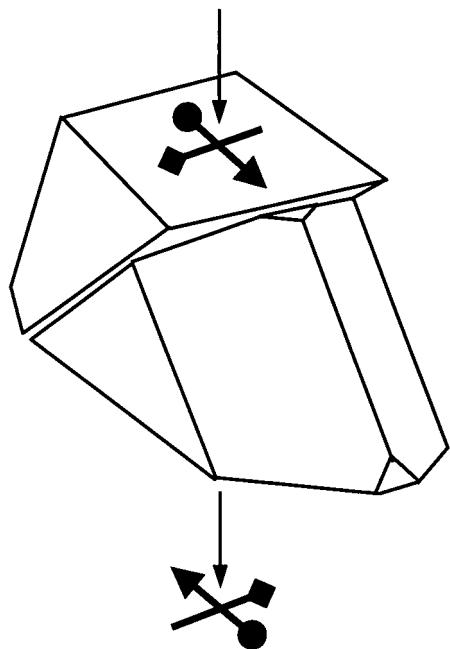
Tracing of an image through a Pechan prism is shown in Fig. 8.12. The object chosen for tracing is a circle-arrow object. After refraction through the surface, *AB*, the image is totally internally reflected from surface *BD*. This TIR limits the field of view of the Pechan prism, which is about 10° for medium-index glasses. After surface *BD*, the rays reach the surface, *AC*, which has to be mirror coated. Small arrows on each surface show the orientation of the circle-arrow object as it falls on a given surface. The image is then refracted by the two surfaces with the air gap between them, and then totally internally reflected off surface *EE'*. The next surface, *GH*, also has to be mirror coated for the rays to be reflected, and after the final TIR on surface *EH*, the image exits the prism as an inverted image with no deviation. Although the Pechan prism consists of two prisms, which means that the alignment and mounting of the prism is rather complicated, it is very commonly used to invert the image. Most of the systems that require image rotation or derotation, have a Pechan prism that rotates half the rotation angle of the scanning entrance mirror. It can be used in a convergent beam, and the clearance for the prism rotation is the smallest of all inverting prisms with no deviation.

If surface *GH* of the Pechan prism is converted into a roof, where two sides of the roof are normal to each other, the prism is called a roof Pechan prism (Fig. 8.13). It has six reflections, so that the image is right-handed. The added reflection on the roof is in the plane which is normal to all other planes of reflection in the prism. This means that the image is right side up. Binoculars with a straight axis, which are Newton-type telescopes, use this type of prism for image erection. The prism is located between the objective and the image plane formed in the focus of the eyepiece. The light beam coming from one point in the field is split on the roof in two beam segments, each undergoing reflections on both sides of the roof. These two segments join after the roof, and form one image. However, if there is an error in the 90° roof angle, it may introduce the image blur or a double image. The tolerance on a roof angle is, in most cases, only a few arc-seconds.

An optical designer has to be careful when designing the prisms, or using standard prisms in an optical system, because the prisms have to work properly for a specified field of view, or a given cone angle. Mistakes can result in the appearance of ghost images. For example, in the case of the Pechan prism, there are three surfaces where the beam is

Design Forms

Figure 8.13
Roof Pechan Prism



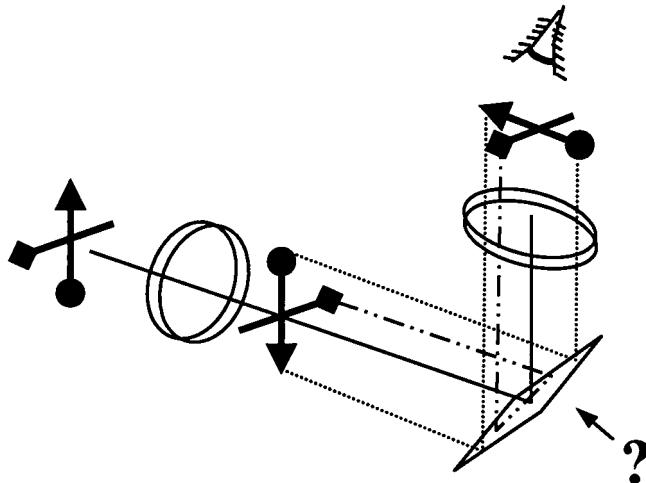
totally internally reflected. If the incident cone angle is too large for a chosen index of refraction of the prism, there will be one part of the field, which will not be properly reflected through the prism as the rest of the image. It will only be refracted and pass directly through the prism. The other case when the ghost images can appear is when the prism is not large enough and some skew rays undergo one extra reflection off the side surfaces. This can happen in the right-angle prism. The ghost reflections may be eliminated making the prism larger, and cutting notches in the prism.

Let us now look at a right-angle telescope, determine what kind of prism can be used to deviate the optical axis by 90° , and also erect the image so that the viewer sees a noninverted right-handed image through the eyepiece. A Newtonian telescope creates an inverted right-handed image. If we put a screen in the focal plane of the objective, the image is both inverted and reverted. The objective creates an image which is right-handed, with an altered orientation both left to right and top to bottom. The eyepiece acts like a magnifier and it does not change the image orientation.

In order to determine the image orientation at any location along the optical system, it is convenient to write on a small piece of paper an object with no symmetry. This can be a letter R, and this piece of paper should be moved through the space, simulating reflections off the prism mirror surfaces. You should always be located such that the image is moving toward your eyes. Do not forget to rotate the image left to right and top to bottom in the case of the objective lens or a relay lens, and do not rotate it through the eyepiece.

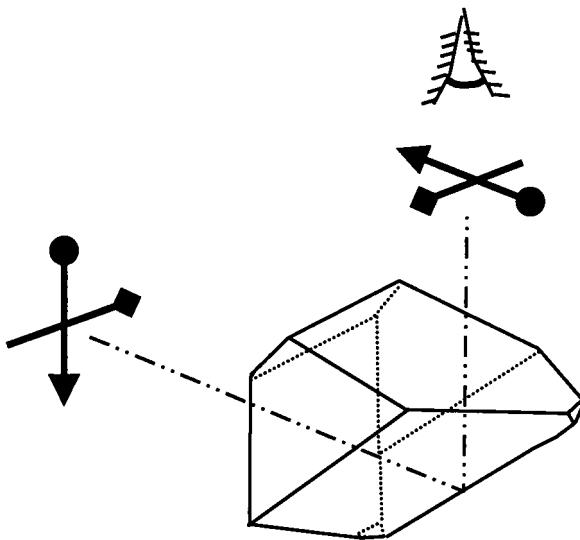
Our problem is sketched in Fig. 8.14. The object is shown as a circle at the bottom, an arrow at the top and a square on the left. After the objective, the image changes its orientation in both directions. Since the eyepiece does not change the image orientation, corrections in the image orientation in the horizontal and the vertical plane have to be done with the prism. If we would have only a plane mirror, the arrow and the circle would be properly oriented after the reflection. However, the square, which was turned to the right side after the objective, would stay on the right side after the reflection off the mirror. This means that we have to find a way to invert the image left to right, keeping only one reflection in the vertical plane. This can be accomplished using the right-angle prism and adding the roof on the hypotenuse surface. This prism is called the *Amici prism*, and it is shown in Fig. 8.15.

Figure 8.14
Right-Angle Telescope



Design Forms

Figure 8.15
Amici Prism



There are many standard prism types. Useful information about prisms, their function, and the important dimensional relationships, can be found in the MIL-HDBK-141 (1962).

Design of Visual Systems

Visual optics includes the wide variety of optical systems creating imagery to be viewed directly by the human eye. This includes telescopes, microscopes, binoculars, riflescopes, camera viewfinders, head-mounted displays, magnifiers, and others. Common to all of these systems is that the eye is looking into some form of viewing optics such as an eyepiece.

Basic Parameters of the Human Eye

The basic optical parameters of the human eye are listed in Table 8.1. Note that most of the data are listed as "approximately" due to the natural variation from person to person.

TABLE 8.1

Typical Human Eye
Optical Parameters

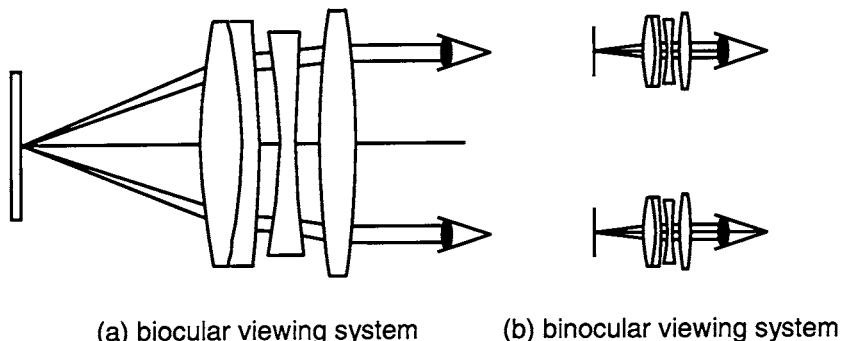
Parameter	Value for Human Eye
Entrance pupil diameter (mm)	≈ 2.5 to ≈ 7
Focal length (mm)	≈ 16.9
f/number	≈ 2.4 to ≈ 6.8
Distance from cornea to point of rotation (mm)	≈ 13.5
Radius of cornea (mm)	≈ 8
Interpupillary distance (IPD) (mm)	≈ 63.5 average, 46 to 80 range
Accurate seeing area (degrees)	≈ 1
Normal viewing angles (horizontal) (degrees)	$\approx \pm 5$ to ± 30
Total visual limit (horizontal) (degrees)	$\approx \pm 108$

The eye of an average young person can accommodate or focus to a distance in the order of 250 mm. Due to geometry, as a person looks at a progressively closer object, the two eyes must progressively increase their angle of convergence. In the limit, for an interpupillary distance or IPD of 63.5 mm and an object at a 250-mm distance, the full convergence angle is 14.5°. If we work with a monocular system such as a riflescope, then convergence is not an issue; however, for a binocular, or even more for biocular systems, it can be quite important. A *biocular* system is a large field-of-view “eyepiece” about 100 mm in diameter, used for viewing a screen or display with both eyes. Figure 8.16 shows a binocular and a biocular system. The design of biocular systems is extremely difficult, and complex multielement designs often are required. One of the difficulties to the designer is that each eye is looking through the extreme edges of the biocular system, and this is the region of the pupil which generally is the most difficult to correct for aberrations.

The user of a pair of binoculars will manually focus the binoculars. Since there is generally no convergence in the optical paths, the user will most often focus the binoculars at infinity, especially when looking at distant objects. This means that collimated light will exit the eyepiece and enter into the eye. If we were to design a riflescope, convergence is not an issue, so the user will focus the device to the most comfortable

Design Forms

Figure 8.16
Biocular and Binocular Systems



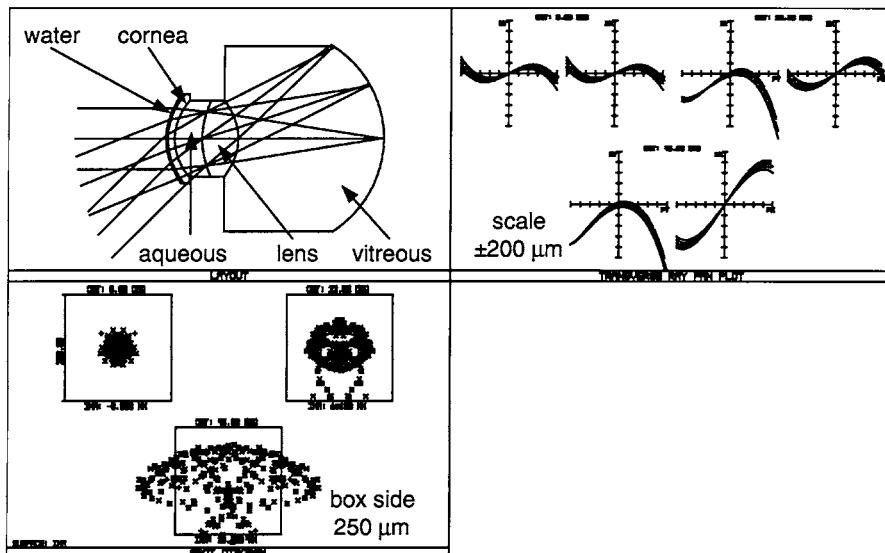
distance. This is often in the range of 2.5 to 6 m, although some research has shown that the resting state of the eye in a dark condition is more like 1 m. It is important to keep in mind that if you require the user to accommodate to some close distance, then you should consider converging the two optical paths (diverging them into the eyes).

You can appreciate the situation if you consider providing diverging light into the two eyes from a virtual object 250 mm in front of the user. While the user can easily accommodate to a 250-mm distance, if the two optical paths were parallel to each other, then there would be potentially significant eye strain, since when accommodating to 250 mm, a person will naturally converge his or her eyes by about 14.5° as noted previously.

There are a number of reasonably reliable eye models, and Figs. 8.17 and 8.18 show one of the more common models, the Lotmar eye model, for eye pupils of 7 and 3 mm in diameter, respectively. The data are shown for fields of view of 0° , 22.5° , and 45° off axis. You can see that there is a residual of spherical aberration which is, of course, more prominent with the larger pupil diameter. For the 7-mm-diameter pupil the spherical aberration equates to an rms blur diameter of about 8 min of arc on the retina, and this reduces to about 1.3 min of arc for the 3-mm-diameter pupil. The eye was permitted to refocus for each pupil diameter. Note the significant off-axis coma and astigmatism residuals.

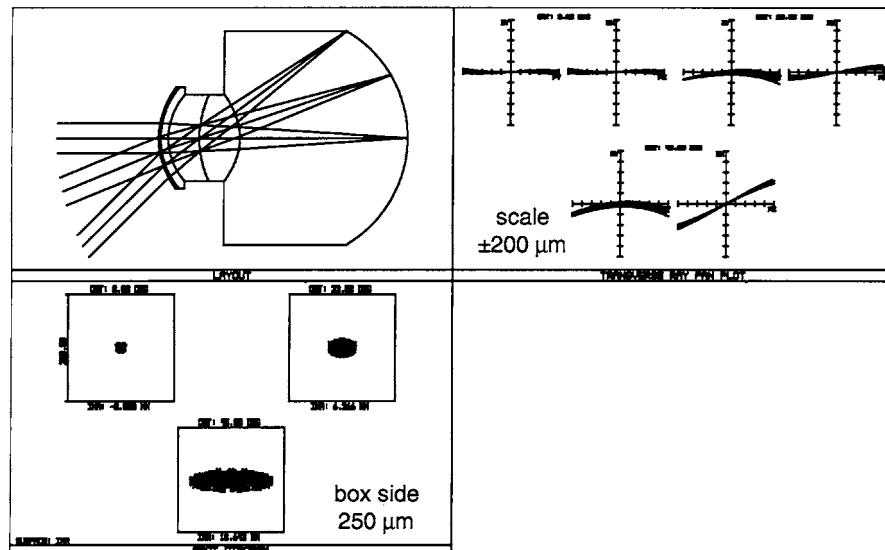
It has been our experience that the residual eye aberrations are somewhat different for different eye models. This leads to the following question: Should the design of a visual system include the effects of the eye? In other words, should your optical design attempt to cancel the eye's residual aberrations based on one of the eye models? In all likelihood

Figure 8.17
Lotmar Eye Model
with 7-mm-Diameter
Pupil



you could do a reasonably good job of accomplishing this from a lens design standpoint; however, the disadvantage is that the eye's aberrations vary from person to person, and even the eye models available give somewhat different aberration residuals. It is generally accepted that it is most prudent to design visual optical systems assuming a perfect eye.

Figure 8.18
Lotmar Eye Model
with 3-mm-Diameter
Pupil

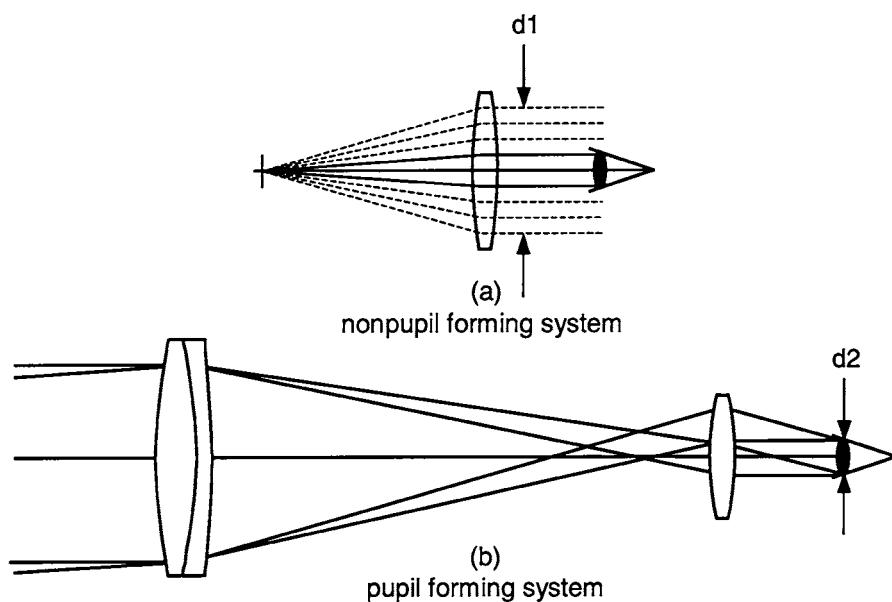


Pupil-Forming Systems

Most (but not all) visual systems are known as “pupil-forming systems.” The term can be best illustrated by considering Fig. 8.19, where we show a magnifying glass in Fig. 8.19a and a simple telescope in Fig. 8.19b. The magnifying glass is *not* a pupil-forming system. What this means is that there is no well-defined exit pupil from the magnifying glass which needs to be mated or matched to the entrance pupil of the eye. On the other hand, the exit pupil of the telescope (Fig. 8.19b) is well defined, and in order to see the imagery, the entrance pupil of the eye *must* be lined up with the exit pupil of the telescope.

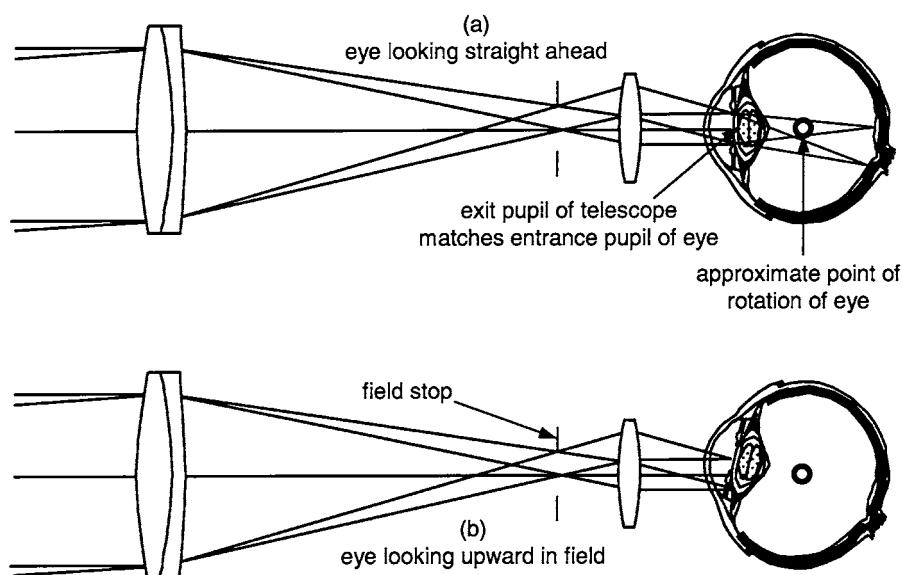
Another way to understand pupil-forming systems is to imagine placing a white card in place of the eye in both Figs. 8.19a and b. Further, assume a bright object which is being viewed. In the case of the magnifying glass, light from the object being viewed will fill a large area of diameter d_1 , which is effectively the diameter of the lens. In the case of the telescope, there will be a bright disk of diameter d_2 at the exit pupil of the telescope. If the telescope were similar to a pair of 7×50 binoculars, this bright disk at d_2 will be 7 mm in diameter. The important aspect of pupil-forming systems is that special attention must be given to assure that the light exiting the optical system does enter the pupil of the eye.

Figure 8.19
Pupil-Forming Optical Systems



There is a very interesting subtlety, and this can be understood by looking at Fig. 8.20. In Fig. 8.20a the eye is looking straight ahead to the left. The center of the field of view will be at or near the fovea, which is the highest acuity of the retina, and the user will see in his or her peripheral vision the field stop of the eyepiece. In effect, the imagery will fall within a well-defined circle, and there is black outside of the circle. If the person now looks upward to the left by rotating his or her eye about the center of rotation of the eye, some of the light from the edge of the field will not enter the eye's pupil. In situations where we have a wide apparent field-of-view telescope with a 7-mm or smaller exit pupil diameter, when the person looks upward toward the edge of the field stop, the light may even disappear completely! It is possible to move your eye closer to the eyepiece so that the exit pupil of the telescope is located at the point of rotation of the eye. Now the person can see clearly the field stop in the eyepiece when looking toward the field stop; however, when looking straight ahead, the field stop and the outer periphery of the field of view may completely disappear from view! This is a very striking, as well as a weird, effect, and if you ever have the opportunity to see it, it is worthwhile to do so.

Figure 8.20
Telescope Showing
Effect of Eye Rotation



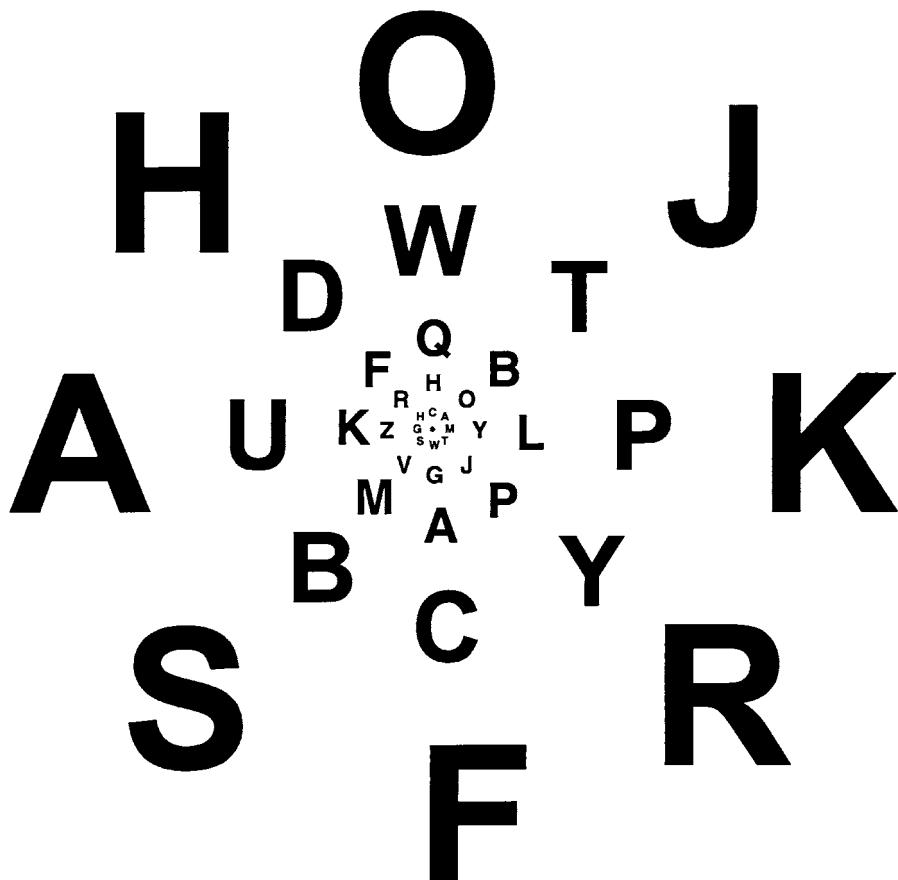
Requirements for Visual Optical Systems

When designing visual systems, a number of parameters must be considered which are unique to these systems, and these are listed here:

- For many hand-held systems such as binoculars, the user will most often bring the object of greatest interest to the center of the field of view. For this reason, off-axis performance is generally more forgiving for these systems.
- The *eye relief* is the clearance or distance from the last optical element in the viewing optics to the front of the cornea of the eye. For eyeglass users the generally accepted minimum eye relief is in the order of 25 mm. Larger values are, of course, helpful, but the diameter of the eyepiece grows with the eye relief.
- It is clear from the transverse ray aberration curves for the Lotmar eye model that the resolution of the eye degrades quite severely as we move away from the center of the field of view. The visual acuity of the eye is maximum at the fovea, which is very close to the center of the field of view of the eye. As an object moves away from the fovea, the visual acuity decreases dramatically, and at $\pm 20^\circ$ the visual acuity is only about 10% of that at the fovea. This is well illustrated in Fig. 8.21 where we show a series of letters of different sizes. The increase in letter size is inversely proportional to the decrease in visual acuity as we move from the fovea. Thus, if you look at the small spot at the center of the figure, all of the letters should be approximately equally resolved.

Although it is true that the visual acuity drops quickly away from the fovea, visual systems should have relatively good imagery at the edge of the field. How good depends on the application of the system. For example, consider a binocular. A viewer can rotate his or her eyes to look toward the edge of the field, in which case the field periphery is the sharpest area the eye sees. This could lead us to a conclusion that the edge of the field should be very well corrected for aberrations. However, we all know that any object in the field can be brought to the center by simply rotating the whole binocular. So the quality of the imagery at the edge of the field in a binocular does not have to be as good as in the center, but it should not be so bad that the image blurring and coloring is immediately noticed when the eyes are pointed to the field edge.

Figure 8.21
Illustration of Visual Acuity As a Function of Distance from the Fovea



- The most general and accepted metric that we often hear is that *the eye resolves 1 min of arc*. What this means is that the eye can resolve the capital letter E when each of the dark or bright bars forming the horizontal portions of the letter subtend 1 min of arc, as shown in Fig. 8.22. This means that the eye resolves 2 min per line pair, or 0.5 line pair per minute of arc.
- With respect to visual systems design, providing an image blur diameter from 1 min of arc to perhaps 3 min of arc is generally considered an acceptable level of performance in the center of the field of view. At the edge of the field, 20 to 40 arc min of image blur maximum may be acceptable.

Design Forms

Figure 8.22
Eye Resolution

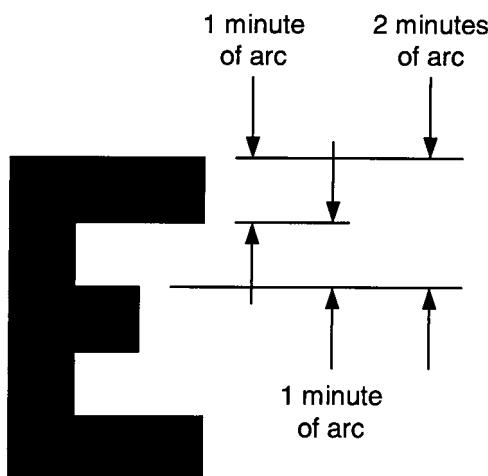


Table 8.2 shows a list of the typical tolerances associated with optics for the eye, as developed by Mouroulis. These data are for binocular viewing with two eyes. Note that items that the eye normally does not do are specified with tighter tolerances. This includes dipvergence (one eye looking upward and the other eye downward) and divergence for example.

There is a rule of thumb that says that the eye can resolve a contrast of 5%. Thus, it is not uncommon to determine where the MTF drops to 0.05 and to conclude that the eye will resolve this spatial frequency. There is perhaps a better way, and this is to use the so-called aerial image modulation (AIM) curve. This is the relationship between the contrast and the number of line pairs per millimeter that the eye can resolve. Walker shows these data, and they are summarized in Fig. 8.23. The AIM curve is a relationship between the modulation required to resolve a given target and the spatial frequency of the target. Using the Lotmar eye model with a 7-mm pupil diameter, we see that if the eye were diffraction limited in Fig. 8.23a, we would resolve 1.6 min/line pair, or 0.8 min/line; however, due to the spherical aberration, we can only resolve in the order of 3 min/line pair, or 1.5 min/line (Fig. 8.23b). In the case of a 3-mm pupil diameter the diffraction-limited eye would resolve about 1.6 mm/line pair (Fig. 8.23c) and the Lotmar model predicts about 1.7 mm/line pair (Fig. 8.23d). It is important to realize that these data are only as accurate as the Lotmar eye model and the referenced AIM threshold data. While they may not be precisely accurate, they do give us a good indication of the resolution of the eye.

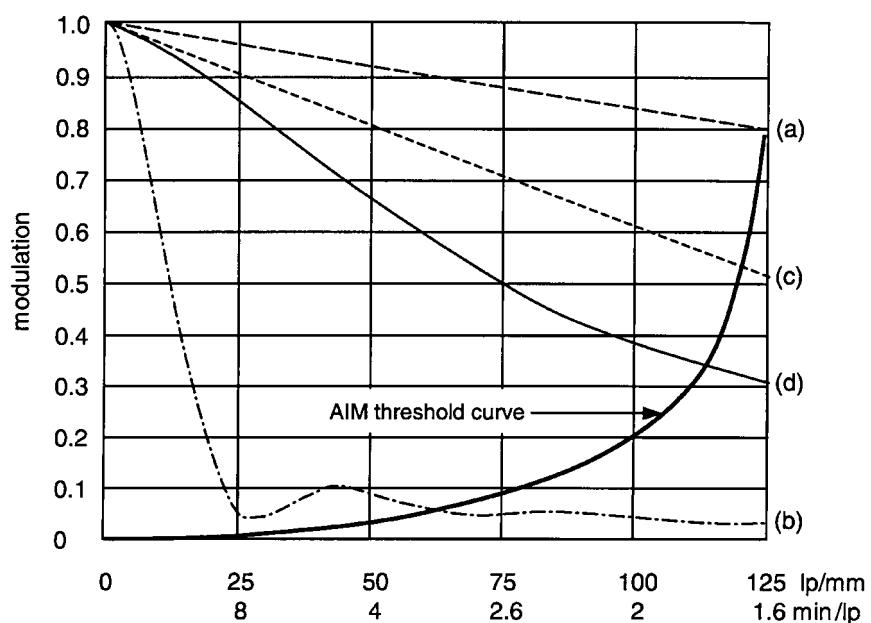
TABLE 8.2

Typical Tolerances
for Visual Systems

Parameter	Typical Specification
Divergence (degrees)	0
Convergence (degrees)	1.5
Dipvergence (min)	8
Magnification difference (%)	± 0.5 to ± 1
Brightness difference (%)	± 10
Best focus if fixed	-1 diopter*
Accommodation	-3 diopter to infinity is OK -1 to -1.5 diopter is best
Axial chromatic aberration	-1 to -1.5 diopters
Lateral color (min)	2
Image quality	Integral of MTF from 0 to 20 line pairs/degree

*A diopter is the reciprocal of focal length, in meters. As used in the table, -1 diopter means that the eye will be focusing to a distance of 1 m in front of the user.

Figure 8.23
Use of AIM Threshold Curve to Predict Visual Resolution



Distortion is an important criterion in visual systems. As we learned in Chap. 5, distortion is more of a mapping error rather than an image-degrading aberration. Its primary effect is to the cosmetic appearance of the image. Distortion is a change in magnification with field of view. Generally, positive or negative distortion in the order of 2 to 2.5% is small enough to be almost imperceptible. Large amounts of distortion can be annoying in any visual system. Telescopes with a large apparent field of view often have up to 10% of distortion at the edge of the field.

Summary on Design of Visual Systems

It is best in the design of visual optical systems to assume a perfect eye model. Most of the computer design programs have the ability to model a so-called perfect lens or paraxial lens, which is used following the system being designed to emulate the eye (or any form of focusing lens). What these so-called perfect lenses do is to simply compute the angular aberrations in the optical system and multiply them by the focal length of the perfect lens which is input by the user. So if you are interested in the actual image blur on the retina, then use a perfect lens of focal length 16.9 mm, which is the approximate focal length of the eye.

If you are designing a visual system which must be fixed focus, it is probably best to design and produce your system to provide light which is either collimated to the user or appears to come from a distance of from 3 to 6 m. Some data suggest a value closer to 1 m. If you can allow the user to refocus, this is generally preferred.

The image quality, or blur diameter, should be in the order of 1 min of arc to 3 min of arc. The specific application will have a lot to do with how good the imagery needs to be. You should design your system for an exit pupil diameter of at least 7 mm (10 to 12 mm is better), and then evaluate the performance both at the 7-mm-diameter pupil as well as with reduced pupil diameters. Some detailed specifications include the eye being decentered to various positions within the exit pupil of the optical system. If the system is to be used in bright conditions, you may even evaluate your design in the 2.5- to 3-mm pupil diameter region.

Our final note on the design of visual systems is to keep in mind that the eye is quite forgiving. Persons who get a new eyeglass prescription

often notice color fringing or lateral color when looking toward the outer periphery of their field of regard (30° to 40° from straight ahead, for example). After several days to several weeks, this color fringing often seems to disappear. The eye, along with the rest of the human visual system, is, in effect, a very powerful computer with impressive image-processing capabilities. This does not mean that you can ignore the aberrations and image quality of visual systems, what it does mean, however, is that, in many situations, the optics for visual systems can be more forgiving than you might think.

The Optical Design Process

The optical design process includes a myriad of tasks that the designer must perform and consider in the process of optimizing the performance of an imaging optical system. While we often think primarily of the robustness of the optimization algorithm, reduction of aberrations, and the like, there is much more to do. The designer must be at what we sometimes call “mental and technical equilibrium with the task at hand.” This means that he or she needs to be fully confident that all of the following are understood and under control:

- All first-order parameters and specifications such as magnification, focal length, f /number, full field of view, spectral band and relative weightings, and others.
- Assure that the optical performance is being met, including image quality, distortion, vignetting, and others.
- Assure that the packaging and other physical requirements, including the thermal environment, is being taken into account.
- Assure that the design is manufacturable at a reasonable cost based on a fabrication, assembly, and alignment tolerance analysis and performance error budget.
- Consider all possible problems such as polarization effects, including birefringence, coating feasibility, ghost images and stray light, and any other possible problems.

Once every one of these items has been addressed and is at least recognized and understood, we start with the sketch of the system. First, the system is divided into subsystems if possible, and the first-order parameters are determined for each subsystem. For example, if we are to design a telescope with a given magnification, the entrance pupil diameter should be chosen such that the exit pupil size matches the eye pupil. A focal length of the objective and the eyepiece should be chosen such that the eyepiece can have a sufficiently large eye relief. Now, when the specs for each subsystem are defined, it is time to use the computer-aided design algorithms and associated software to optimize the system, which will be discussed in the rest of this chapter. Each subsystem can be designed and optimized individually, and the modules joined together or, more often, some subsystems are optimized separately and some as an integral part of the whole system.

What Do We Do When We Optimize a Lens System?

Present-day computer hardware and software have significantly changed the process of lens design. A simple lens with several elements has nearly an infinite number of possible solutions. Each surface can take on an infinite number of specific radii, ranging from steeply curved concave, through flat, and on to steeply curved convex. There are a near infinite number of possible design permutations for even the simplest lenses. How does one optimize the performance with so many possible permutations? Computers have made what was once a tedious and time-consuming task at least manageable.

The essence of most lens design computer programs is as follows:

- First, the designer has to enter in the program the starting optical system. Then, each variable is changed a small amount, called an *increment*, and the effect to performance is then computed. For example, the first thickness may be changed by 0.05 mm as its increment. Once this increment in thickness is made, the overall performance, including image quality as well as physical constraints, are computed. The results are stored, and the second thickness is now changed by 0.05 mm and so on for all variables that the user has designated. Variables include radii, airspaces,

The Optical Design Process

element thicknesses, glass refractive index, and Abbe number. If you are using aspheric or diffractive surfaces, then the appropriate coefficients are also variables.

- The measure of performance as used here is a quantitative characterization of the optical performance combined with a measure of how well the system meets its first-order constraints set by the user such as focal length, packaging constraints, center and edge thickness violations, and others. The result of the computation is a single number called an *error function* or *merit function*. The lower the number, the better the performance. One typical error function criteria is the rms blur radius, which, in effect, is the radius of a circle containing 68% of the energy. Other criteria include optical path difference, and even MTF as described in Chap. 15.
- The result is a series of derivatives relating the change in performance (P) versus the change in the first variable (V_1), the change in performance (P) versus the change in the second variable (V_2), and so on. This takes on the following form:

$$\frac{\partial P}{\partial V_1}, \frac{\partial P}{\partial V_2}, \frac{\partial P}{\partial V_3}, \frac{\partial P}{\partial V_4} \dots$$

- This set of partial derivatives tells in which direction each parameter has to change to reduce the value of the sum of the squares of the performance residuals. This process of simultaneous parameter changes is repeated until an optimum solution is reached.

A lens system consists of a nearly infinite number of possible solutions in a highly multidimensional space, and it is the job of the designer to determine the optimum solution.

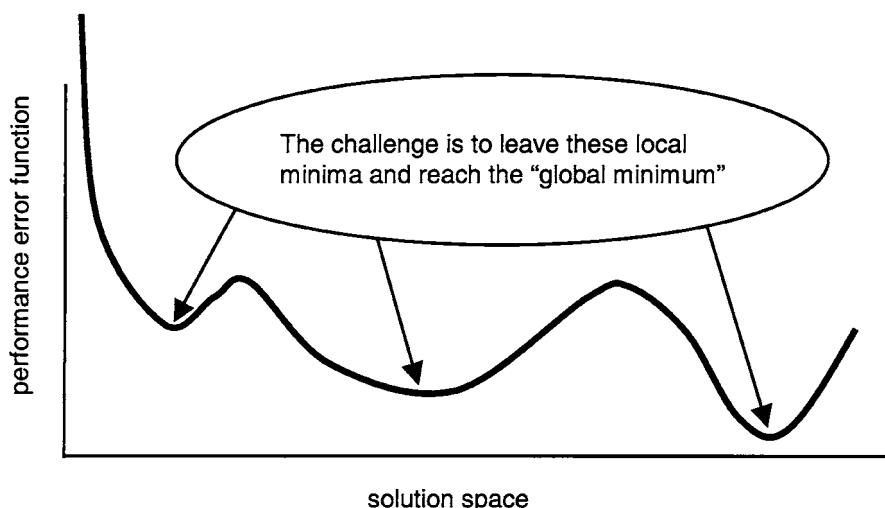
Designers have used the following analogy to describe just how a lens design program works:

- Assume that you cannot see and you are placed in a three-dimensional terrain with randomly changing hills and valleys. Your goal is to locate the lowest elevation or altitude, which in our analogy equates to the lowest error function or merit function. The lower the error function, the better the image quality, with the “goodness” of performance being inversely proportional to the elevation.
- You are given a stick about 2 m long, and you first stand in place and turn around tapping the stick on the ground trying to find which direction to walk so as to go down in elevation.

- Once you determine the azimuth resulting in the greatest drop in elevation, you step forward in that direction by 2 m.
- You now repeat this process until in every direction the elevation goes up or is level, in which case you have located the lowest elevation.
- But what if just over a nearby hill is an even lower valley than you are now in? How can you find this region of solution? You could use a longer stick, or you could step forward a distance several times as long as the length of your stick. If you knew that the derivative or slope downward is linear or at least will continue to proceed downward, this may be a viable approach. This is clearly a nontrivial mathematical problem for which many complex and innovative algorithms have been derived over the years. But the problem is so nontrivial as well as nonlinear that software algorithms to locate the so-called global minimum in the error function are still elusive. Needless to say, the true global minimum in the error function may be quite different or distant from the current location in our n -dimensional terrain.

Figure 9.1 shows a two-dimensional representation of solution space as discussed previously. The *ordinate* is the error function or merit function, which is a measure of image quality, and the *abscissa* is, in effect, solution space. We may initiate a design on the left and the initial

Figure 9.1
Illustration of Solution Space in Lens Design



optimization brings the error function to the first minimum called a *local minimum* in the error function. We then change glasses and/or make other changes to the design and ultimately are able to move the design to the next lower local minimum. Finally, we add additional elements and make other changes and we may be able to reach the local minimum on the right. But how do we know that we are at, or even close to, a global minimum? Here lies the challenge as well as the excitement of lens design!

It is important here to note that reaching global minimum in the error function is not necessarily the end goal for a design. Factors including tolerance sensitivity, packaging, viability of materials, number of elements, and many other factors influence the overall assessment or “goodness” of a design. Learning how to optimize a lens system is, of course, quite critical to the overall effort, and learning how to reach a viable local or near-global minimum in the error function is very important to the overall success of a project.

How Does the Designer Approach the Optical Design Task?

The following are the basic steps generally followed by an experienced optical designer in performing a given design task. Needless to say, due to the inherent complexity of optical design, the processes often become far more involved and time consuming. Figure 9.2 outlines these basic steps:

1. The first step in the design process is to *acquire and review all of the specifications*. This includes all optical specifications including focal length, *f*/number, full field of view, packaging constraints, performance goal, environmental requirements, and others.
2. Then we *select a representative viable starting point*. The starting point should, wherever possible, be a configuration which is inherently capable of meeting the specifications for the design. For example, if the specifications are for an *f*/10 monochromatic lens covering a very small field of view and having an entrance pupil diameter of 5 mm, then the lens may very well be a single element. However, if the requirements call for an *f*/1.2 lens over a wide spectral band covering a 40° full field of view, then the solution may very well

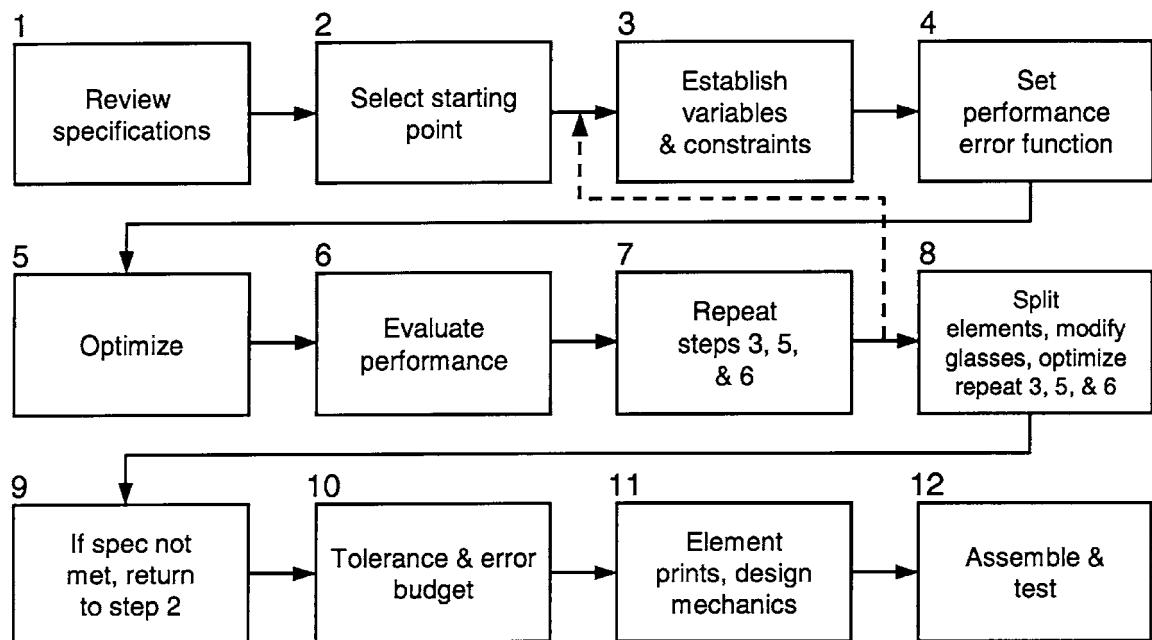


Figure 9.2
Lens Design and Optimization Procedure

be a very complex six- to seven-element double Gauss form. If we were to use a single element for this latter starting point, there would be no hope for a viable solution. Finding a good starting point is very important in obtaining a viable solution. The following are viable sources for starting points:

You can use a *patent* as a starting point. There are many sources for lens patents including Warren Smith's excellent book *Modern Lens Design*. There is also a CD-ROM called "LensView," which contains over 20,000 designs from patents. These are all searchable by a host of key parameters. While the authors of this book are not patent attorneys, we can say with confidence that you may legally enter design data from a patent into your computer and work with it in any way that you would like to. If your resulting design is sold on the market, and if the design infringes on the patent you used (or any other for that matter), you could be cited for patent infringement. It is interesting to note that the purpose

of our patent system in this country is to promote inventions and innovation. This is done by offering an inventor a 17-year exclusive right to his or her invention in exchange for *teaching* in the patent how to implement the invention. Thus, you are, in effect, invited to use the design data and work with it with the goal of coming up with a better design, which you can then go out and patent. By this philosophy, inventors are constantly challenged to improve upon an invention, which, in effect, advances technology, which is what the patent process is all about. Needless to say, we urge you to be careful in your use of patents.

You could use a so-called hybrid design. We mean a hybrid to be the combining of two or more otherwise viable design approaches so as to yield a new system configuration. For example, a moderate field-of-view Tessar lens design form can be combined with one or more strongly negatively powered elements in the front to create an extremely wide-angle lens. In effect, the Tessar is now used over a field of view similar to its designed field, and the negative element or elements bend or "horse" the rays around to cover the wider field of view. An *original design* can, of course, be a viable starting point. As your experience continues to mature, you will eventually become comfortable with "starting from scratch." With today's computer-aided design software, this works most of the time with simple systems such as doublets and triplets; however, with more complex systems, you may have problems and will likely be better off resorting to a patent or other source for a starting point.

3. Once you have entered your starting point into the software package you are using, it is time to *establish the variables and constraints*. The system variables include the following: radii, thicknesses, airspaces, surface tilts and decenters, glass characteristics (refractive index and Abbe number), and aspheric and/or other surface variables, including aspheric coefficients. The constraints include items such as focal length, *f*/number, packaging-related parameters (length, diameter, etc.), specific airspaces, specific ray angles, and virtually any other system requirement. Wavelength and field weights are also required to be input. It is important to note that it is not imperative (nor is it advisable) to vary every conceivable variable in a lens, especially early in the design phase. For example, your initial design optimization should

probably be done using the glasses from the starting point, in other words do not vary glass characteristics initially. This will come later once the design begins to take shape and becomes viable. You may also want to restrict the radii or thicknesses you vary as well, at least initially. For example, if adjacent elements have a very small airspace in the starting design, this may be for a good reason, and you should probably leave them fixed. Also, element thicknesses are very often not of great value as variables, at least initially, in a design task, so it is usually best to keep element thicknesses set to values which will be viable for the manufacturer.

4. You now will *set the performance error function and enter the constraints*. Most programs allow the user to define a fully “canned” or automatically generated error function, which, as discussed earlier, may be the rms blur radius weighted over the input wavelengths and the fields of view. In the Zemax program the user selects the number of rings and arms for which rays will be traced into the entrance pupil (rays are traced at the respective intersection points of the designated number of rings and arms). Chapter 22 shows a detailed example of how we work with the error function.
5. It is now time to *initiate the optimization*. The optimization will run anywhere from a few seconds for simple systems to many hours, depending on just how complex your system is and how many rays, fields of view, wavelengths, and other criteria are in the system. Today, a state-of-the-art PC optimizing a six- to seven-element double Gauss lens with five fields of view will take in the order of 5 to 10 s per optimization cycle. Once the computer has done as much as it can and reaches a local minimum in the error function, it stops and you are automatically exited from the optimization routine.
6. You now *evaluate the performance* using whatever criteria were specified for the lens. This may include MTF, encircled energy, rms spot radius, distortion, and others.
7. You now *repeat steps 3 and 5 until the desired performance is met*. Step 3 was to establish the variables and constraints, and step 5 was to run the optimization, and these steps are repeated as many times as necessary to meet the performance goals. You will often reach a solution that simply does not meet your performance requirements. This is very common during the design evolution, so do not be surprised, depressed, or embarrassed if it happens

to you... it happens to the best of us. When it does happen, you may need to add or split the optical power of one or more of the lens elements and/or to modify glass characteristics. As we have discussed previously, splitting optical power is extremely valuable in minimizing the aberrations of a lens.

8. There is a really simple way of splitting an element in two, and while it is not “technically robust,” it does work most of the time. What you do is insert two surfaces in the middle of the current element, the first of which will be air and the second is the material of your initial element. The thickness of each “new” element is one-half of the initial element and the airspace should be small, like 0.1, for example. Now simply enter twice the radius of the original element for both s_1 and s_2 of the new elements. You will end up with two elements whose net power sum is nearly the same as your initial element. You can now proceed and vary their radii, the airspace, and, as required, the thicknesses.
9. If you still cannot reach a viable design, then at this point you will need to *return to step 2 and select a new starting point*.
10. Your final task in the design process is to *perform a tolerance analysis and performance error budget*. We will be discussing tolerancing in more depth in Chap. 16. In reality, you should be monitoring your tolerance sensitivities throughout the design process so that if the tolerances appear too tight, you can take action early in the design phase and perhaps select a less sensitive design form.
11. Finally, you will need to *generate optical element prints, contact a viable lens manufacturer, and have your elements produced*. You will also need to work with a qualified mechanical designer who will *design the cell or housing* as well as any required interfaces. It is important to note that while we list the mechanical design as taking place at this point after the lens design is complete, it is extremely important to work with your mechanical designer throughout the lens design process so as to reach an optimum for both the optics as well as the mechanics. Similarly, you should establish a dialog with the optical shop prior to completing the design so as to have time to modify parameters which the shop feels needs attention such as element thicknesses, glass types, and other parameters.
12. Once the components are in house, you will need to *have the lens assembled and tested*. Assembly should be done to a level of precision and cleanliness commensurate with the overall

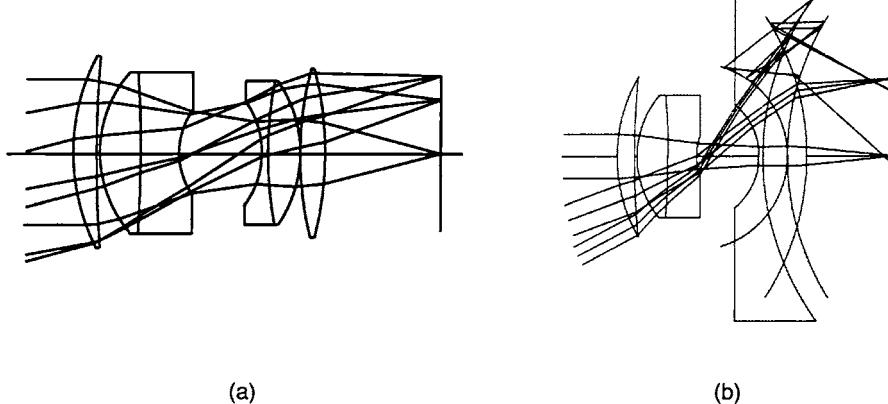
performance goals. Similarly testing should be to a criterion which matches or can be correlated with your system specifications and requirements. We discuss testing in Chap. 15.

Sample Lens Design Problem

There was a very interesting sample lens design problem presented at the 1980 International Lens Design Conference. The optimized design for an $f/2.0$, 100-mm focal length, 30° full field-of-view double Gauss lens similar to a 35-mm camera lens was sent out to the lens design community. One of the tasks was to redesign the lens to be $f/5$ covering a 55° full field with 50% vignetting permitted. Figure 9.3 shows the original starting design, as well as the design after changing the f /number and field of view, without any optimization.

Sixteen designers submitted their results, and they spent from 2 to 80 h working on the problem. We will present here three representative solutions in Fig. 9.4. The design in Fig. 9.4a is what we often call a *happy lens*. What we mean is that the lens is quite well behaved with no steep bending or severe angles of incidence. The rays seem to “meander” nicely through the lens. It is a comfortable design. We show to the right of the layout a plot of the MTF. MTF will be discussed in detail in Chap. 10. For the purpose of this discussion, consider the MTF to be contrast plotted in the ordinate as a function of the number of line pairs per millimeter

Figure 9.3
Starting Design for
Sample Lens Design
Problem



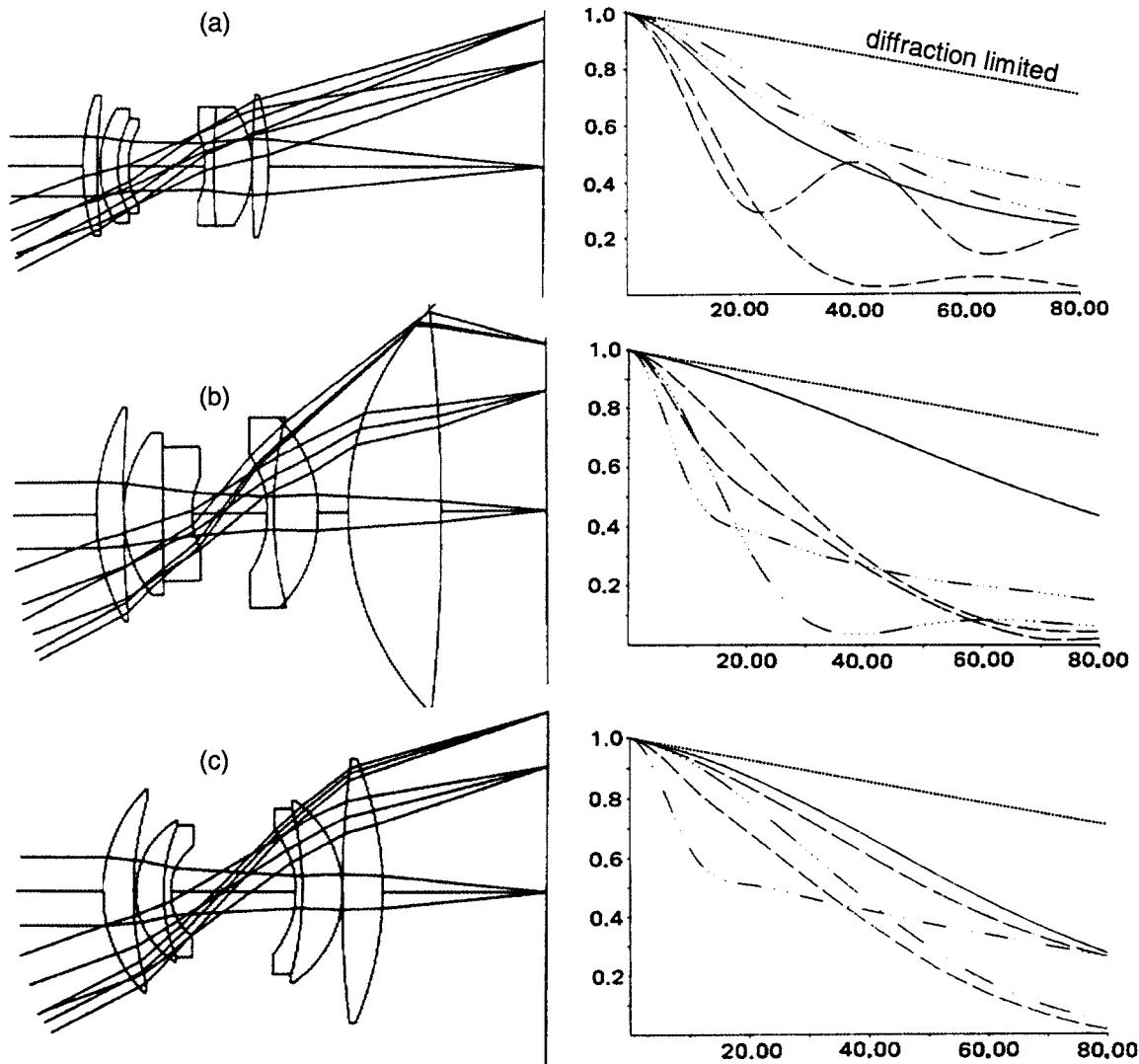
The Optical Design Process

Figure 9.4
Representative Solutions to Sample Lens Design Problem

in the abscissa. The different curves represent different positions in the field of view and different orientations of the resolution patterns. The higher the curves, the better the contrast and the overall performance. The MTF is reasonable for most of the field positions. As will be discussed in Chap. 22, a good rule of thumb for the MTF of a 35-mm camera lens is an MTF of 0.3 at 50 line pairs/mm and 0.5 at 30 line pairs/mm.

The design in Fig. 9.4*b* has a serious problem; the rays entering the last element are at near-grazing angles of incidence. Notice that the exit pupil at full field is to the right of the lens (since the ray cone is descending toward the axis to the right), and at 70% of the field the exit pupil is to the left of the lens (since the ray cone is ascending to the right and therefore appears to have crossed the axis to the left of the lens). This is a direct result of the steep angles of incidence of rays entering the last element. The variation in exit pupil location described here would not itself be an issue unless this lens were used in conjunction with another optical system following it to the right; however, it does indicate clearly the presence of the severe ray bending which will inevitably lead to tight manufacturing and assembly tolerances. Further, the last element has a near-zero edge thickness which would need to be increased. The lens is large, bulky, and heavy. And finally, the MTF of this design is the lowest of the three designs presented.

Finally, the design in Fig. 9.4*c* is somewhat of a compromise of the two prior designs in that it is somewhat spread out from the design in Fig. 9.4*a* but does not have the problems of the design in Fig. 9.4*b*. The MTF of the design in Fig. 9.4*c* is the best of the three designs.

Comparing the three designs is very instructive as it shows the extreme variability of results to the same problem by three designers. The question to ask yourself is what would you do if you subcontracted the design for such a lens, and after a week or two the designer brought you a stack of paper 200-mm thick with the results of the design in Fig. 9.4*b*. And what if he or she said "wow, what a difficult design! But I have this fabulous solution for you!" Prior to reading this book, you might have been inclined to congratulate the designer on a job well done, only to have problems later on during manufacturing and assembly. Now, however, you know that there may be alternate solutions offering superior performance with looser tolerances and improved packaging. Remember that even a simple lens has a near infinite number of possible solutions in a multidimensional space.

CHAPTER

10

Computer Performance Evaluation

What Is Meant by Performance Evaluation

The performance characteristics of an imaging optical system can be represented in many ways. Often the final optical performance specification is in terms of the modulation transfer function (MTF), encircled energy, rms blur diameter, or other image quality criteria. These criteria relate in different ways to the image quality of the system. Image quality can be thought of as *resolution* or how close two objects can approach each other while still being resolved or distinguished from one another. Image quality can also be thought of as image sharpness, crispness, or contrast.

As discussed earlier, imagery is never perfect. It is limited by geometrical aberrations, diffraction, the effects of manufacturing and assembly errors, and other factors. The characterization of image quality by the methods described in the following sections will help you to assess just how your system performs with respect to its imagery.

It is important to realize that the image quality or resolution of the entire system is not totally dependent on the optics, but may include the sensor, electronics, display device, and/or other system components

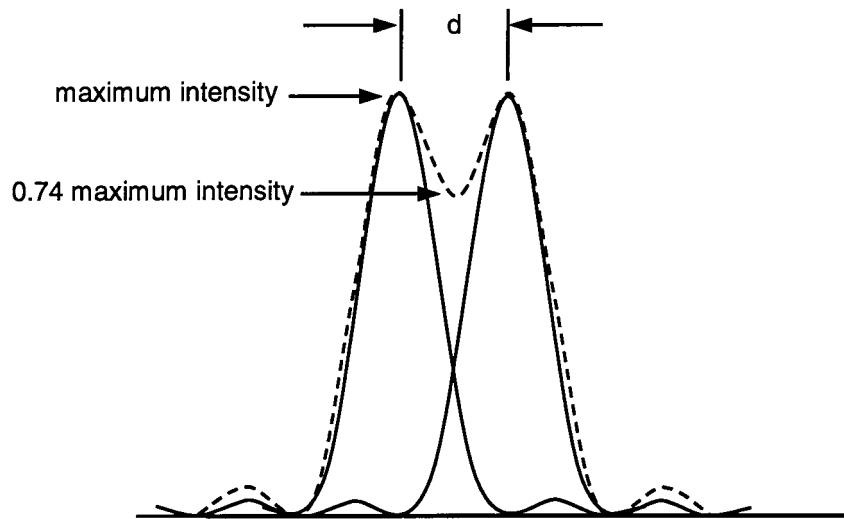
making up the system. For example, if the eye is the sensor, it can accommodate for both defocus and field curvature, whereas a flat sensor such as a CCD cannot. In this chapter, we will be discussing only the optics contribution to image quality.

What Is Resolution?

When we think about the image quality of an image-forming optical system such as a camera lens, the first parameter that often comes to mind is *resolution* or *resolving power*. Classically, the ability of an optical system to separate two closely spaced point sources at the nominal object distance is generally considered to be the resolution.

Consider a perfect optical system which has an entrance pupil diameter, D and focuses to an image with a given $f/ \#$. Two point sources closely spaced will be imaged through the optical system, each of them forming a diffraction pattern. If two perfect diffraction patterns as in Fig. 10.1 are separated by the radius of the Airy disk (the radius of the first dark ring in the diffraction pattern), then the intensity midway between the two peaks in the pattern drops to 0.74 of the maximum intensity, and the two point images are said to be *resolvable*. This is the Rayleigh criterion

Figure 10.1
Two Resolvable
Images of Closely
Spaced Point Sources



for resolution. This, of course, assumes that the ultimate media or sensor is not the limiting factor. The separation, d , of the two points in the image plane is

$$d = 1.22 \lambda f/\#$$

in units of wavelength. In object space, in radians, this becomes (λ and pupil diameter in same units)

$$\alpha = \frac{1.22 \lambda}{\text{entrance pupil diameter}}$$

or very frequently used as a rule of thumb, the resolution, in arc seconds, for the visible spectral range is

$$\alpha = \frac{136}{\text{entrance pupil diameter}}$$

where the entrance pupil diameter is given in millimeters. This is interesting and certainly of value in understanding the limiting resolution of the optical system with given first-order optical parameters, but it really does not help us to understand the performance of a specific optical system design. As will be shown in this chapter, there is far more to the characterization of optical performance than the theoretical resolution.

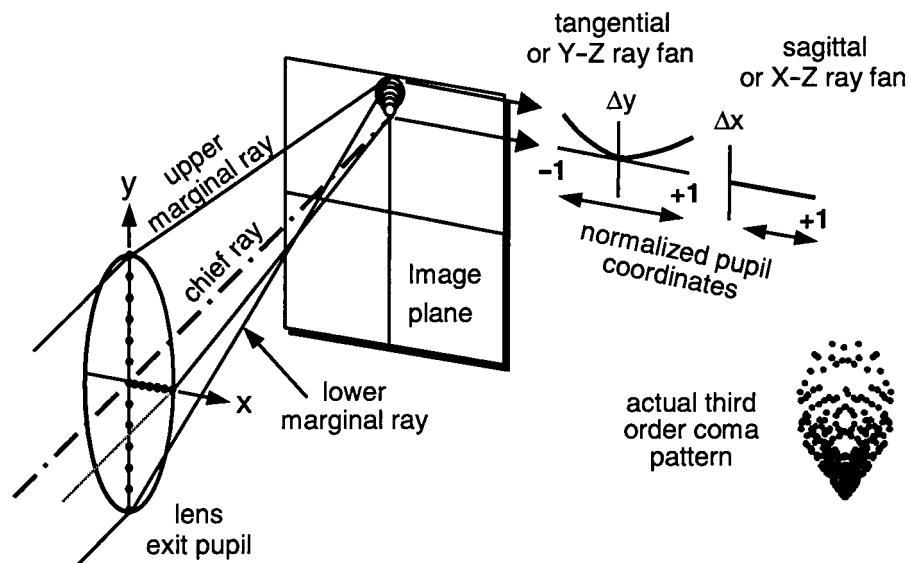
Ray Trace Curves

Most of the methods used in computing image quality, such as the modulation transfer function, spot diagrams, encircled energy, and the like, are functionally robust and represent different, yet similar representations of the net performance of the optical system as designed. However, there are two disadvantages with these metrics. First, they can sometimes take too much time to compute. This, however, is less and less of a problem as PCs have become faster and faster. Second, the real problem is that while these metrics do help to show the overall net resulting image quality, they do not provide a detailed indication to the designer of the specific aberrations present in the design over the field of view and over the spectral bandwidth. While some information can at times be derived, more often the user really cannot tell what aberrations are present and at what magnitudes. These data are important to the designer as an aid in correcting the residual aberrations.

The solution is to generate what are called *transverse ray aberration curves* or simply *ray trace curves*. With these graphical data, a reasonably experienced designer can immediately tell just how much spherical aberration, coma, astigmatism, field curvature, axial color, lateral color, and field curvature are present. In addition, in many cases the user can also tell what orders of these aberrations are present. Finally, with this knowledge, the designer can often make a reliable judgment as to what to do next regarding further optimization of the lens. In spite of some fabulous advances in performance simulation and modeling, transverse ray aberration curves are still invaluable to the serious designer.

We show in Fig. 10.2 the basic formation of the ray trace curve. This perspective figure (Fig. 10.2) shows a lens exit pupil with the lens imaging to an off-axis image position. First, consider tracing the *chief ray* to the image. The height on the image of the chief ray is our reference point, and is generally taken to be the image height. Now let us trace a ray through the top of the exit pupil. This ray, which is called the *upper marginal ray*, hits the image higher than the chief ray for the aberration shown, which is coma. Now let us trace a ray through the bottom of the exit pupil. This ray, which is called the *lower marginal ray*, also hits the image higher than the chief ray, and in fact for classical third-order coma it hits the image the same distance above the chief ray as the ray from the top of the pupil. In

Figure 10.2
Explanation of Ray Trace Curves



other words, both of the rays from the top and the bottom of the exit pupil hit the image vertically displaced by the same amount.

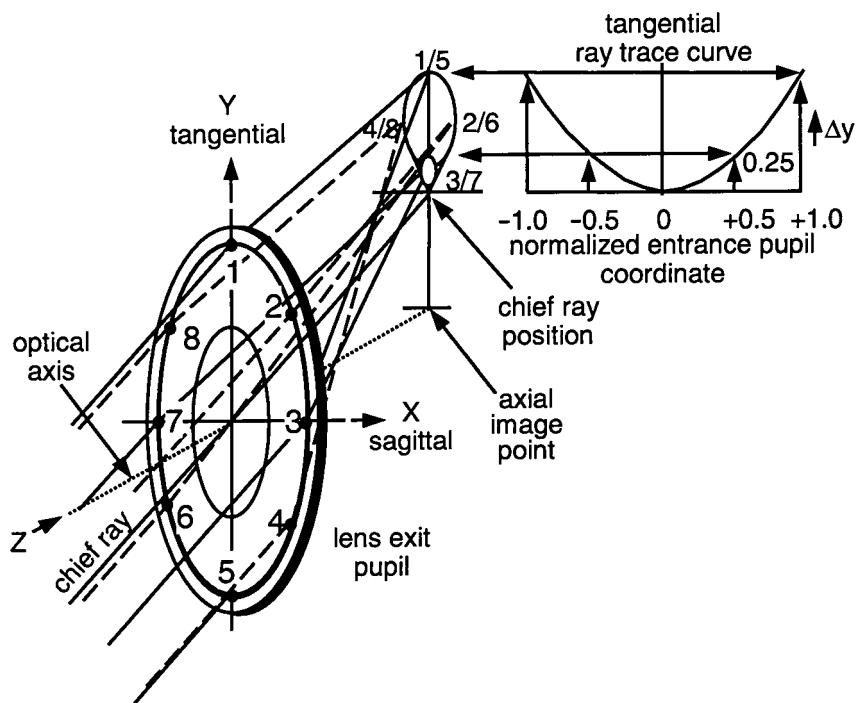
We will now proceed to establish a set of coordinate axes for our ray trace curves. In the first set of coordinates (on the left in Fig. 10.2), the abscissa is the normalized exit pupil radius in the y direction, and the ordinate is the distance above or below the chief ray on the image that our ray intersects the image plane (Δy). Thus, both the upper and lower marginal rays form the end points on the curve. We now proceed to trace rays through each of the black dots from $y = +1$ to $y = -1$, with the intersection points relative to the chief ray plotted on the curve. For third-order coma the result will be a quadratic or parabolic curve since third-order coma is quadratic with aperture.

Now we establish a second set of coordinate axes, as shown on the right in Fig. 10.2. Here we have the normalized X coordinate in the exit pupil in the abscissa, and the displacement of the ray in the x direction (Δx) as the ordinate. As it turns out, for third-order coma there is no x departure at all for these rays in the exit pupil. We will show why this is the case shortly. For now, you will see that for third-order coma a quadratic curve in the “tangential” ray fan and zero departure for the x rays in the “sagittal” ray fan are the results. If any lens designer who is “worth his or her salt” sees this form of ray aberration curves with a quadratic in the tangential ray fan and virtually zero in the sagittal curve, then he or she should conclude instantly that the lens has third-order coma.

Since these ray trace curves are so fundamentally important to the optical designer’s work, a more in-depth discussion is in order. As you will see, there are here, as with many other areas of optical design, subtleties that could easily be misleading if not fully understood. Consider our coma pattern where the ray trace curves suggest zero x departure of the rays hitting the image, which implies or suggests zero x width to the image blur. Yet we all know that coma does have width in the x direction. Just what is going on, and why are the data misleading?

Figure 10.3 will explain the situation. Here we trace rays around the periphery of the exit pupil from positions 1 through 8. From our prior discussion, we know that the chief ray is our reference, and that rays 1 and 5 from the top and bottom of the exit pupil both hit the image high, above the chief ray. If you follow the numbers in Fig. 10.3, you will see how *one* rotation around the exit pupil results in *two* rotations around an ellipse in the image, and since positions 1 and 5 are both high, then positions 3 and 7 which are 180° opposed will be at the bottom of the elliptical pattern. Neither of these rays will have any x departure at

Figure 10.3
Formation of Comatic
Image Blur

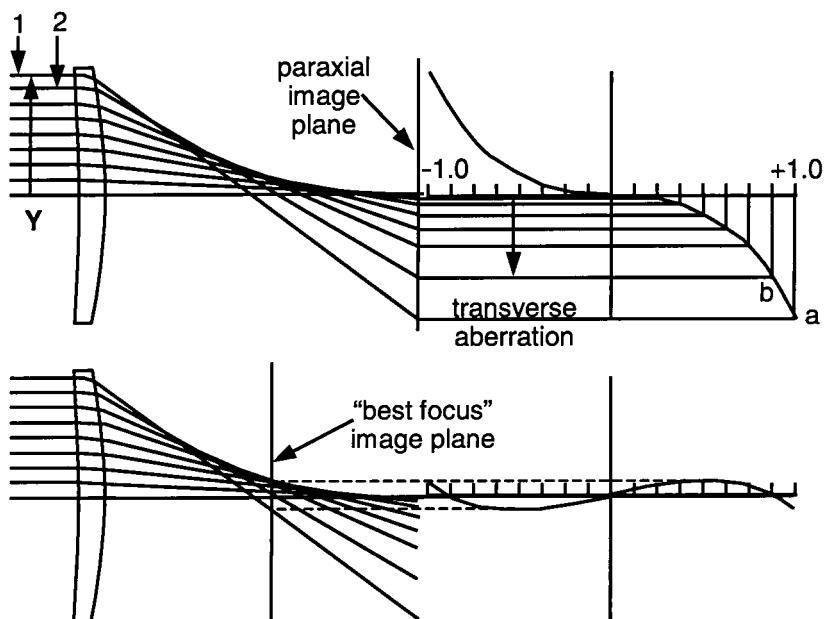


all! Thus, the ray trace curve for rays traced in the x direction was a horizontal line in Fig. 10.2. So where is the x spreading of the coma pattern coming from? The answer is from rays at positions 2, 4, 6, and 8, which are called *skew rays*. Since the rays making up the ray trace curves contain only the y (or *tangential rays*) and the x (or *sagittal rays*), the designer sees no indication or evidence whatsoever of the x spreading of the imagery. This is a real subtlety, and it is a fine example why one should never be totally dependent on only one form of image evaluation or analysis. By looking only at the ray trace curves, one could easily conclude that such a system had virtually zero x spreading of the off-axis imagery, and this could make its performance ideal for some system applications. For the most part use of the ray trace curves are wonderfully helpful and revealing; however, do be aware of subtleties as pointed out earlier.

A further illustration of the ray trace curves, Fig. 10.4 shows how spherical aberration is formed and how the ray trace curves are derived. In the top of Fig. 10.4 the image is located at paraxial focus and it should be clear how each of the rays entering the lens from the left results in a

Computer Performance Evaluation

Figure 10.4
Formation of Ray
Trace Curves for
Spherical Aberration



corresponding intercept on the image plane and how this is plotted as the ray trace curve. Ray 1 strikes the image lowest and results in point *a* in the plot. Ray 2 is the next ray lower down entering the entrance pupil and it results in point *b* in the plot, and so on. Since third-order spherical aberration is cubic with aperture, the resulting curve is cubic. Note the symmetry above and below the optical axis. Now consider what happens if we relocate the image plane to the "best focus" position. Following the same logic in generating the ray trace curves, we see a much lower departure of the ray intercept points making up the curve. This is true and quite real, and it tells us that the image blur diameter when we refocus the image will be significantly reduced from that at paraxial focus. As an exercise, what will the ray trace curve be for a perfect image where the image plane is intentionally defocused toward the lens? The answer is a straight line sloped upward to the right. So let's use this as an aid in further understanding ray trace curves. Since defocus yields a sloped but otherwise straight line, we can easily determine what any ray trace curves for any lens will look like as we go through focus by simply drawing or imagining a sloped straight line as a new coordinate axis. This is an invaluable tool as you can now immediately assess the relative improvement after refocusing a given lens. And since field curvature is a quadratic

change in focus with field of view, you can with a little practice assess immediately the benefits of curving your sensor if this is possible.

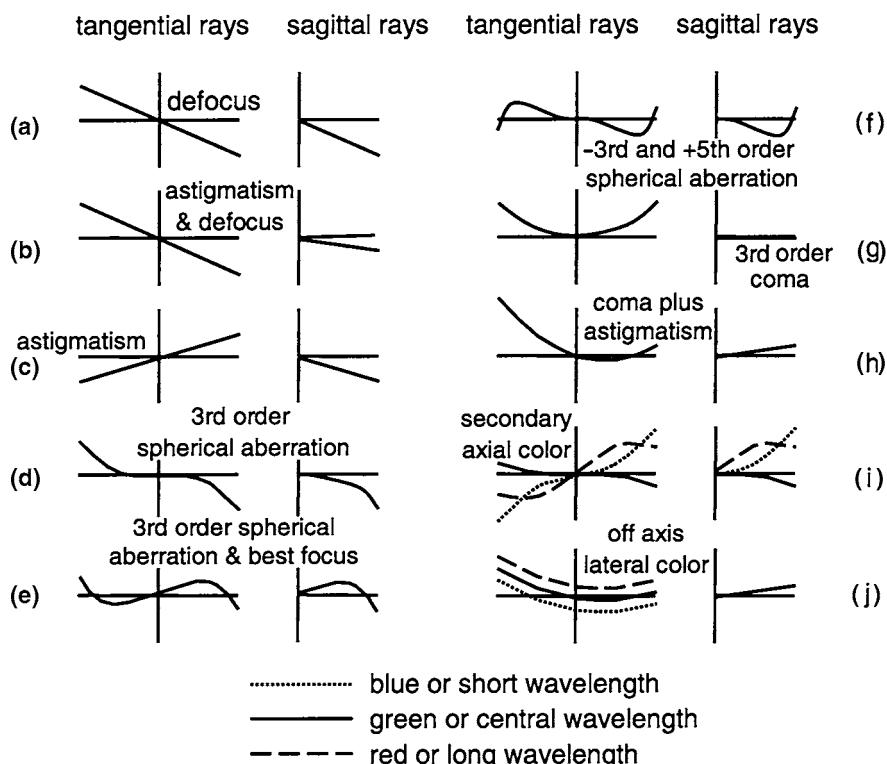
We show in Fig. 10.5 ray trace curves for various typical aberrations and combinations of aberrations:

Figure 10.5a is pure defocus. As noted earlier, defocus will produce equal sloped straight lines in the sagittal and tangential ray fans. Recall that the tangential ray fan is in the y - z direction and typically oriented parallel to the field-of-view direction. The sagittal ray fan is orthogonal to the tangential ray fan, and typically, the sagittal fan is fully symmetrical which is why we sometimes show only one-half of the fan.

Figure 10.5b shows straight lines at different slopes. This is a combination of astigmatism (which is the difference between the slopes of the two curves) and defocus.

Figure 10.5c shows that if we best focus for the residual astigmatism off axis as we might do with a curved image surface, we find the result

Figure 10.5
Typical Transverse Ray
Aberration Curves



here where an equal and opposite ray fan slope results in the tangential and sagittal directions.

Figure 10.5d is negative or undercorrected third-order spherical aberration, which is, of course, a cubic with aperture.

The data in Fig. 10.5e are the same third-order spherical aberration as in Fig. 10.5d, only we have refocused the image to a more optimum focus position to minimize the residual blur diameter.

Figure 10.5f shows negative third-order spherical aberration, which is being balanced by positive fifth-order spherical aberration.

The data in Fig 10.5g are for pure third-order coma, which, as we know from before, is quadratic with aperture. We also know from before that the sagittal curve indicates zero image blurring in the sagittal direction. This may be misleading and is due to the nature of coma formation and the fact that the ray aberration curves show only the rays along two lines in the pupil plane.

The data in Fig. 10.5h are for a combination of third-order coma and astigmatism.

Figure 10.5i shows a combination of some negative third-order spherical aberration at the central wavelength as well as secondary axial color and spherochromatism. The secondary axial color is the focus difference between the central wavelength and the common red and blue focus, which together are focused beyond the central wavelength. The spherochromatism is the change in spherical aberration with wavelength.

Finally, Fig. 10.5j shows an off-axis ray trace curve with primary lateral color or color fringing along with a small amount of coma and astigmatism.

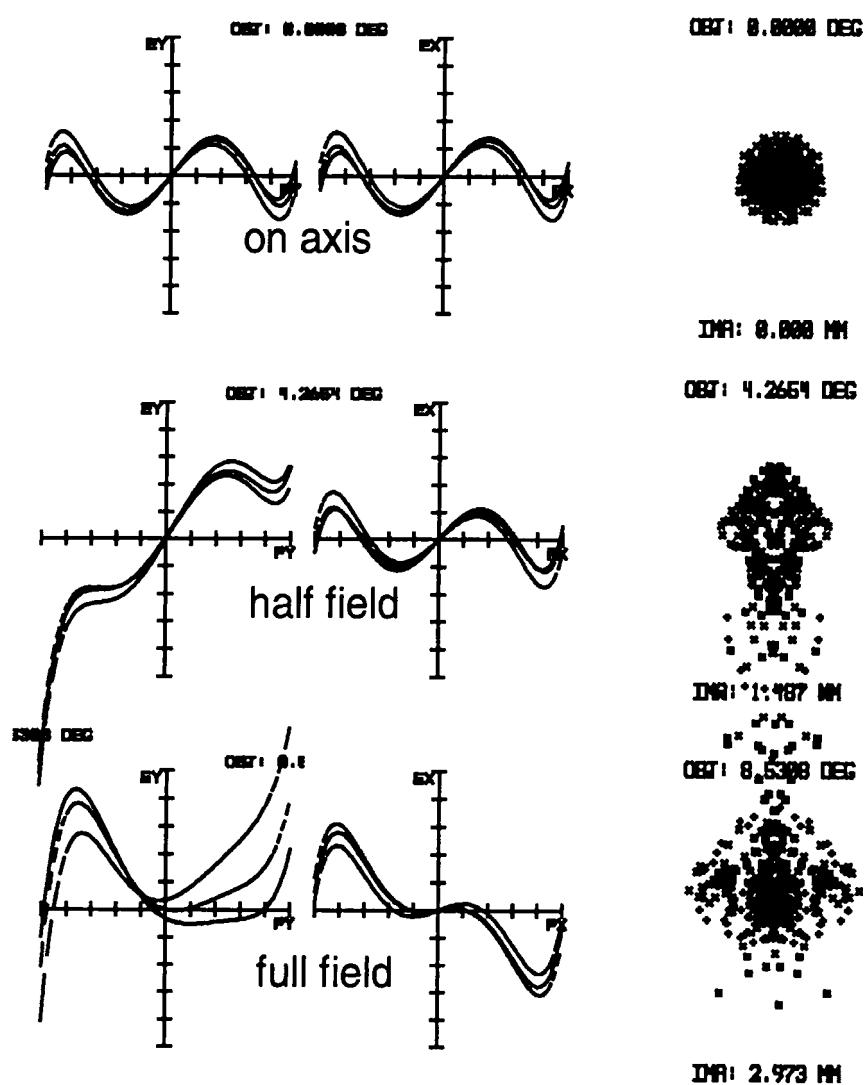
It should be quite apparent that the ray trace curves for each of the aberrations has its own distinctive form, and this is what makes them so useful. The aberrations, in effect, add algebraically, so it is easy to tell almost immediately what aberrations are present in a given design at the different field positions.

Spot Diagrams

Spot diagrams are the geometrical image blur formed by the lens when imaging a point object. This is a more functionally useful form of output; however, it is sometimes difficult to distinguish the specific aberrations present.

Figure 10.6 shows the spot diagrams for a Cooke triplet form of lens and shows both the transverse ray aberration curves as well as spot diagrams for the same field positions. Generally, the rms spot radius or diameter is output with the spot diagrams. The rms spot diameter is the diameter of a circle containing approximately 68% of the energy. This metric can be of great value, especially when working with pixelated sensors where one often wants the image of a point object to fall within a pixel.

Figure 10.6
Geometrical-Based
Spot Diagrams and
Transverse Ray Aberra-
tion Curves for
Cooke Triplet
Example



Note that in optical design software we often come across the terms “spot radius,” “spot diameter,” and “spot size.” While the designer is most interested in spot diameter, the software generally outputs spot radius. The use of the words “spot size” is fine for relative comparison (for example, “the spot size has increased by a factor of 2”); however, the term can cause undo confusion when tied to a specific value. For example, “the spot size is 50 μm ” does not really tell us whether this is the radius or diameter of the image blur, nor does it tell us whether this is for 100% of the energy or some other value such as the rms. Be careful in interpreting these forms of data from the software you are using.

Optical Path Difference

As we know from Chap. 4, if the peak-to-valley optical path difference (P-V OPD) is less than or equal to one-fourth of the wavelength of light, the image quality will be almost indistinguishable from perfect. It is known as diffraction limited. Just like transverse ray aberration curves, we can plot the optical path difference, and Fig. 10.7 shows a typical OPD plot for our sample Cooke triplet. In addition to the plotted data, we show also perspective views of the three-dimensional wavefront departure from perfect. Note how the curve and the perspective view correlate so well for the on-axis field position with the bump in the center as well as the turned up edge of the wavefront clearly evident in both data.

It is a little difficult without extensive experience to quickly determine the residual aberrations from an OPD plot, so you will generally want to compute the more standard ray trace curves for this purpose.

Encircled Energy

Encircled energy is energy percentage plotted as a function of image diameter. One good example of how we might use encircled energy is to specify an imaging optical system using a CCD sensor. Let us assume that the pixel pitch of our sensor is 7.5 μm . A good reliable and simple specification is that 80% of the energy from a point object shall fall within a diameter of 7.5 μm . Figure 10.8 shows an encircled energy plot for our sample Cooke triplet. Eighty percent of

Figure 10.7
Optical Path Difference for Cooke Triplet Example

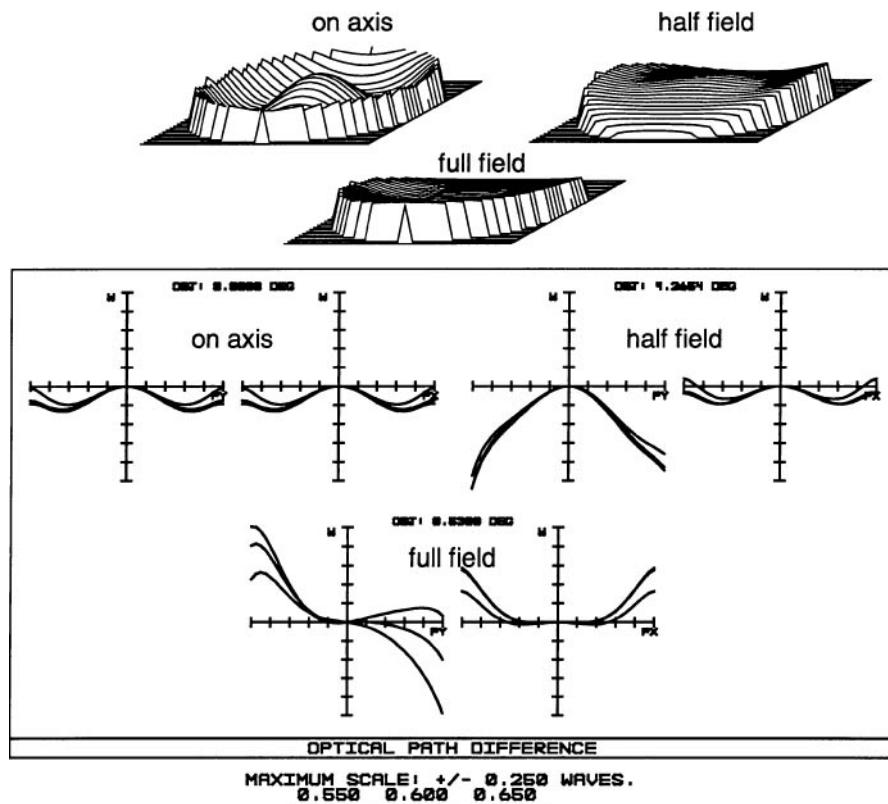
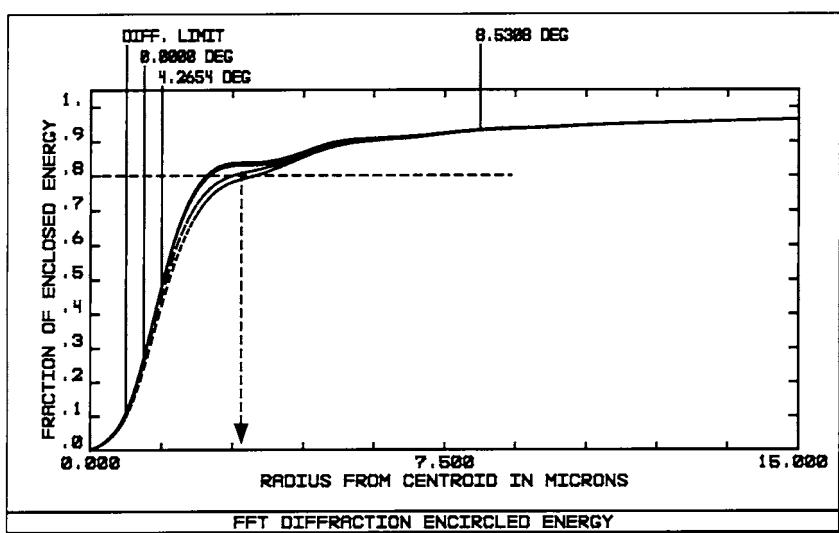


Figure 10.8
Encircled Energy for
Cooke Triplet
Example



Computer Performance Evaluation

the energy is contained within a diameter of approximately $6 \mu\text{m}$, which is a good match to the sensor. It also leaves some margin for manufacturing tolerances.

MTF

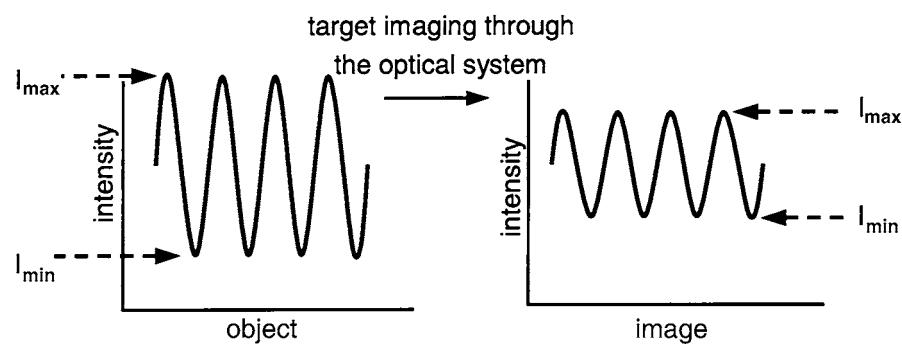
MTF is perhaps the most comprehensive of all optical system performance criteria, especially for image forming systems. Figure 10.9 is a representation of what is happening. We begin with a periodic object or target, which is varying sinusoidally in its intensity. This target is imaged by the lens under test, and we plot the resulting intensity pattern at the image. Due to aberrations, diffraction, assembly and alignment errors, and other factors, the imagery will be somewhat degraded and the brights will not be as bright and the darks will not be as dark as the original pattern.

Let us define some terms:

$$\text{Modulation} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad \text{MTF} = \frac{\text{modulation in image}}{\text{modulation in object}}$$

The modulation is simply the maximum intensity minus the minimum intensity divided by the maximum plus the minimum. The MTF is the ratio of the modulation in the image to the modulation in the object as a function of spatial frequency, which is generally in the form of line pairs per millimeter. Thus, the modulation transfer function represents *the transfer of modulation from the object to the image by the lens as a function of spatial frequency*.

Figure 10.9
Illustration of the
Meaning of the
Modulation Transfer
Function



There is another term, *contrast* (sometimes called *contrast ratio*), which is given by

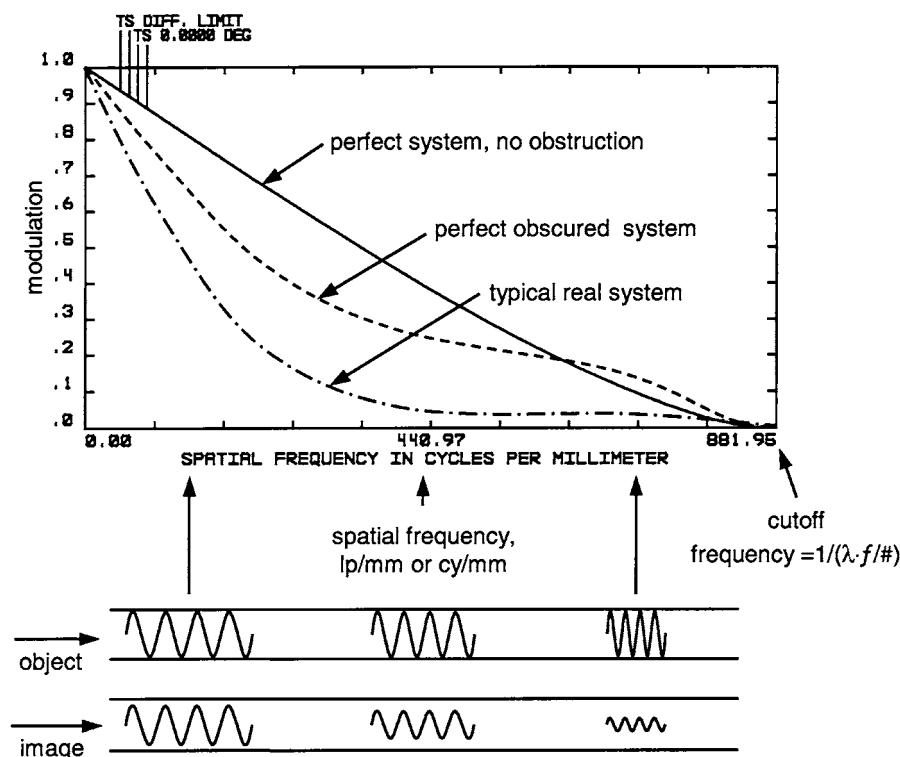
$$\text{Contrast} = \frac{I(\max)}{I(\min)}$$

Figure 10.10 shows several typical MTF curves. We show the MTF of a perfect optical system, a perfect system with a central obscuration (such as a Cassegrain telescope), and a typical real system. The MTF of the perfect obscured system has more diffraction due to its obstruction, and thus a lower MTF. The cutoff frequency, which is where the MTF goes to zero, is

$$\nu_{\text{cutoff}} = \frac{1}{\lambda(f/\#)}$$

The example shown is an $f/2$ lens in the visible ($0.55\text{-}\mu\text{m}$ wavelength), and the cutoff frequency is approximately 882 line pairs/mm. A good

Figure 10.10
Typical MTF Curves

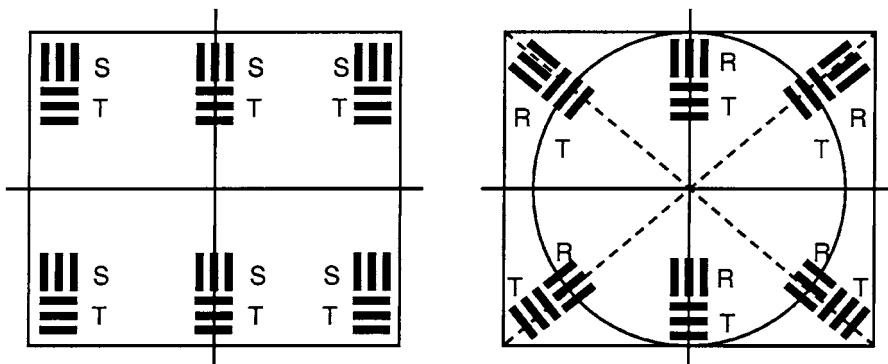


rule of thumb to remember is that an $f/2$ lens used at $0.5 \mu\text{m}$ has a cutoff frequency of 1000 line pairs/mm, and it is very easy to scale from here. For example, an $f/4$ lens has twice the Airy disk diameter and thus half the cutoff frequency of 500 line pairs/mm.

Recall that the MTF tells us how well the modulation in the object is transferred to the image by the lens. We show below the MTF curve in Fig. 10.10 a graphical representation of an object and of the resulting image at a low spatial frequency, a midspatial frequency, and at a high spatial frequency. Think of the low spatial frequency as being very large tree trunks with a bright sky between them, and the high spatial frequency as being tiny close tree branches with the same bright sky between them. The sky is the same for both and the darkness of the bark on the tree trunks and the branches is the same, hence the modulation of the objects is identical. However, the modulation of the image will be far lower at the higher spatial frequency of the branches since the MTF is lower.

The MTF is generally computed or measured for bars that are *radial* and for bars that are *tangential*. The radial bars are like the spokes on a bicycle, and the tangential bars are tangential to the edge of the bicycle tire as well as tangential to the edge of a circular field of view. The tangential and radial bars are orthogonal to each other. There is a subtlety of some significance with respect to the preceding nomenclature and how we handle it in the various software packages. Consider a lens designed to cover a rectangular format with a $3 \times 4 \times 5$ aspect ratio. We said earlier that the radial bars are parallel to spokes on a wheel and tangential bars are orthogonal to the radial bars. This is fine and true, and the definition is valid. However, given the pixelated nature of many of today's sensors such as CCD or CMOS detector arrays, or alternatively pixelated display devices as used in LCD and similar projection systems, it makes far more sense if the resolution bars and their associated nomenclature were consistent with the pixel rows and columns. What we mean is that it will be far more appropriate to refer to bars that are vertical and bars that are horizontal, rather than radial and tangential as shown in Fig. 10.10. Fortunately, most of the software packages use the terminology *sagittal* and *tangential*, and these refer to entrance and exit pupil coordinates. Tangential ray aberrations in the exit pupil are up and down in the plane of the paper in the y direction and sagittal ray aberrations are in and out of the plane of the paper in the x direction. We show in Fig. 10.11 the orientation of the bars using both nomenclatures. Regardless of which software package you use, it is imperative to understand what assumptions are being made with respect to target orientation.

Figure 10.11
Target Orientation Conventions



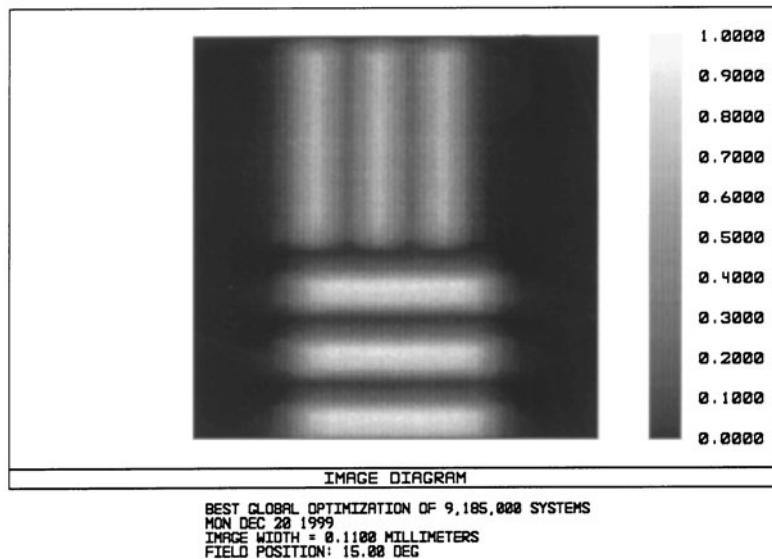
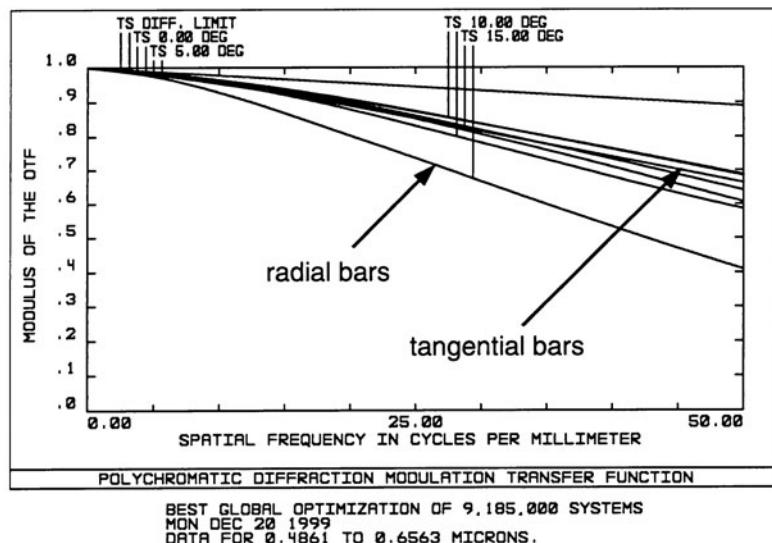
The bottom line here is as follows:

- Tangential aberrations are in the y direction and blur horizontal bars. Tangential bars are horizontal.
- Sagittal aberrations are in the x direction and blur vertical bars. Sagittal bars are vertical.
- Radial bars are bars parallel to spokes on a wheel.
- Tangential bars are tangent to the rim of a wheel.

A good example of a computed MTF and the resulting image quality is shown in Fig. 10.12. Here we plot the MTF of a double Gauss lens at four fields of view (on axis, 0.33 field, 0.67 field, and at the edge of the field). Note how the radial (vertical) bars at the edge of the field of view are more degraded than the tangential (horizontal) bars. The simulated 3-bar imagery is at 50 line pairs/mm, and the modulation at this spatial frequency is about 0.65 for the tangential bars and about 0.4 for the radial bars. The simulated imagery clearly shows this difference in the two target orientations. If you ever have a question with respect to this, we recommend that you read the manual for the software package you are using, phone the user support person, or, even better, run several examples.

Figure 10.13 shows the MTF of a perfect system as a function of obstruction ratio in a system such as a Cassegrain telescope. As the central obscuration increases, the amount of light that is diffracted away from the central maximum of the Airy disk increases and there is a corresponding reduction in MTF. As this is happening, the diameter of the central maximum actually decreases somewhat, resulting in a high-frequency

Figure 10.12
MTF and Simulated
Bar Target Imagery of
Double Gauss Lens



MTF that is actually slightly above the unobscured perfect system MTF. This can sometimes be used to advantage if you are trying to separate two close point objects or stars as in astrometric work. Figure 10.14 shows point-spread functions for aberration-free systems with no central obscuration, 0.2, 0.4, 0.6, and 0.8 diameter central obstructions, respectively.

Figure 10.13
MTF of a Perfect System As a Function of Central Obscuration

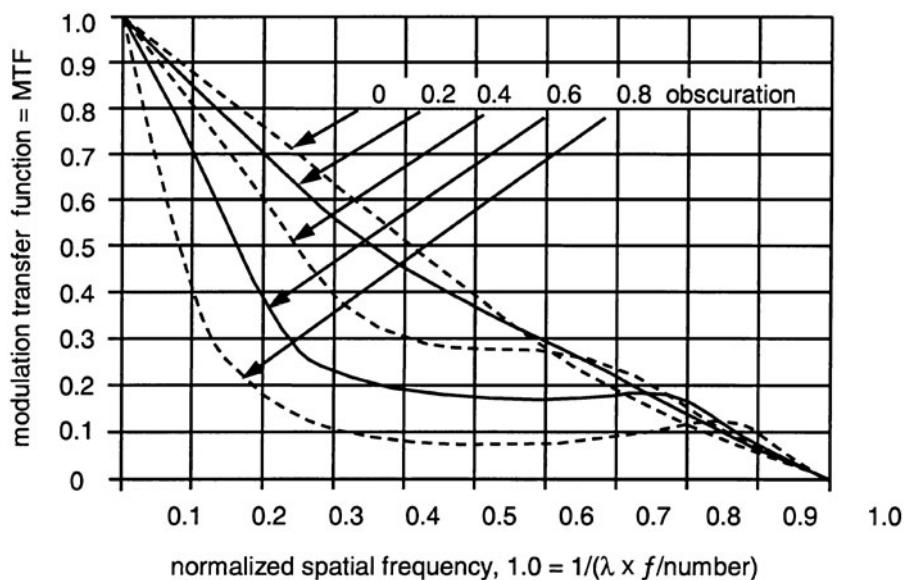
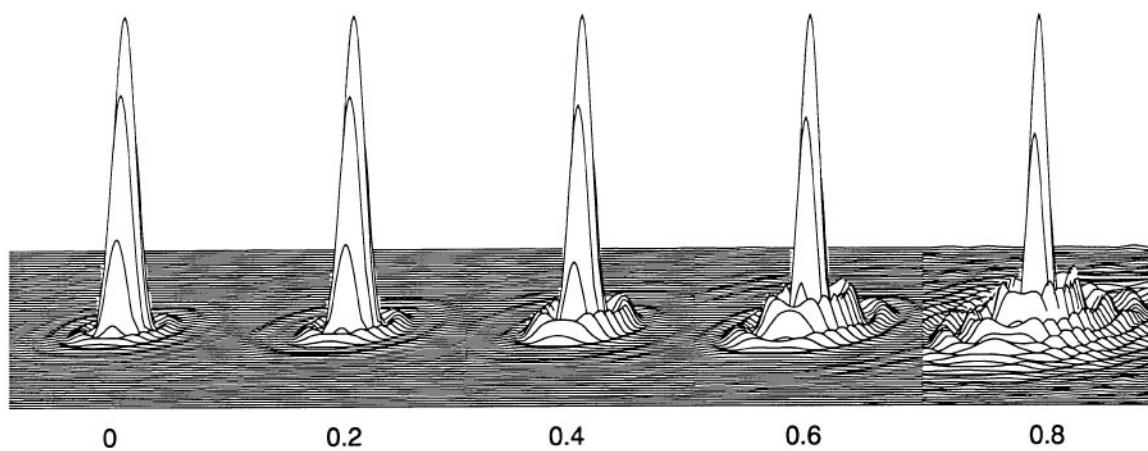


Figure 10.14
Point Spread Functions of an Aberration-Free System As a Function of Central Obscuration



It can be shown that the MTF of an image-forming optical system can actually be less than zero. As an otherwise well-performing system is degraded in performance due to defocus, aberrations, and/or manufacturing errors, the MTF degrades. As the performance continues to degrade, eventually the MTF may drop below zero. In extreme cases, the MTF will oscillate above and below zero. Any time the MTF is less than zero, that constitutes a phase reversal, which is where the dark bars become bright and the bright bars become dark. This can be seen visually by covering one eye and looking at Fig. 10.15. Hold the page at a distance closer than you can comfortably accommodate or focus. Do not try to focus on the pattern. This might range from 25 to 75 mm or more, depending on your accommodation. Then move the page closer or further from your eye very slowly, and you should be able to see very clearly the phase reversal. What is especially interesting here is what you are observing is the phase reversal of the imagery formed by your own cornea and eye as imaged onto your own retina... there is no other optics whatsoever! The phase reversal should be very clear and striking. If for some reason you cannot see it, Fig. 10.16 is a digital photograph of the effect. The camera lens was focused at infinity and located several inches from the figure to obtain the photo. Figure 10.17 shows the MTF

Figure 10.15
Radial Bar Pattern for
Demonstration of
Spurious Resolution

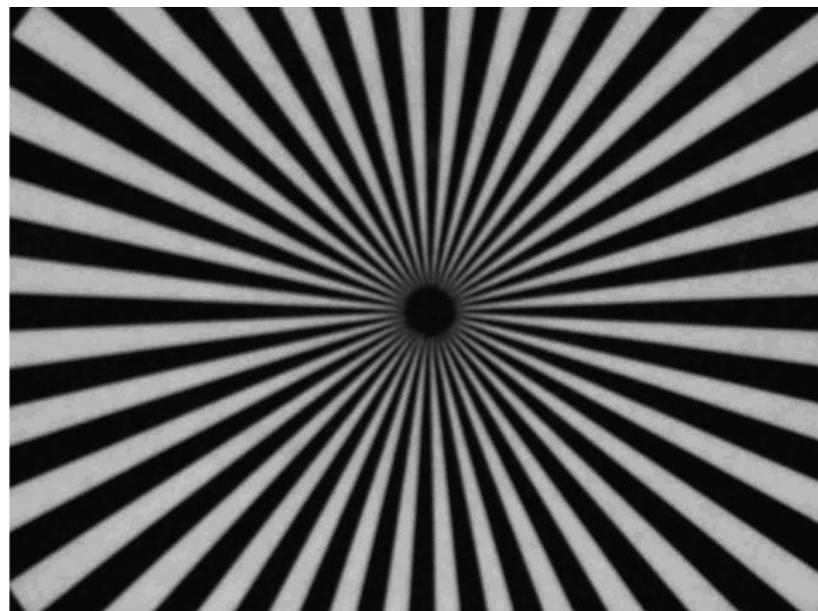
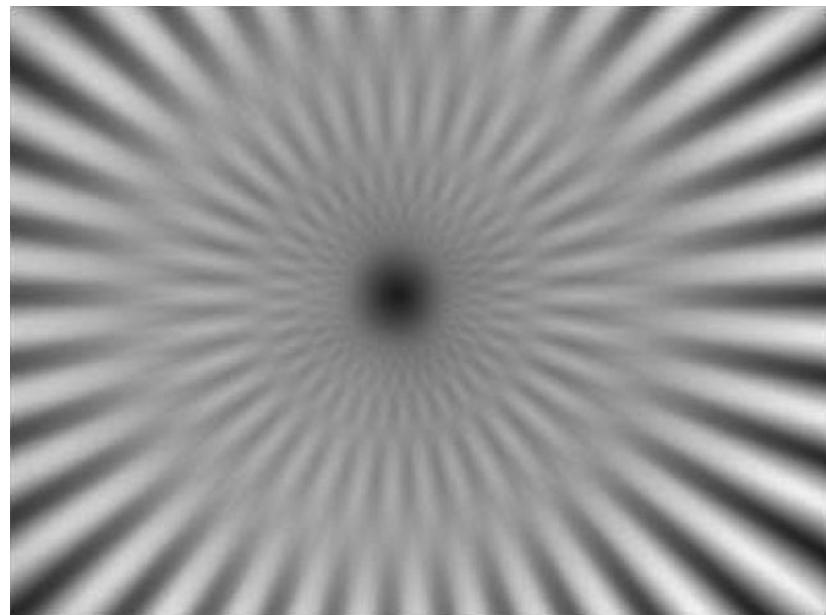
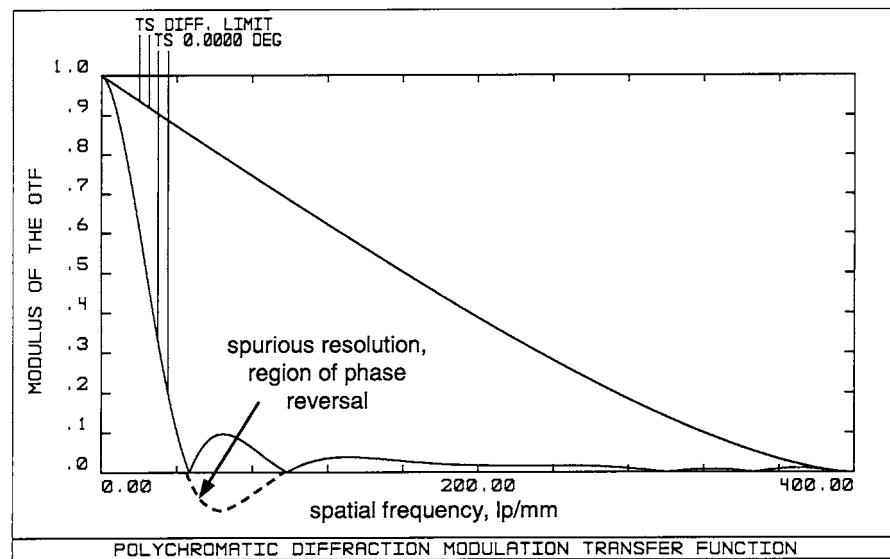


Figure 10.16
Spurious Resolution
Using Defocused
Digital Camera



of an otherwise perfect aberration-free $f/5$ system with approximately 1.5 waves of defocus. Note how at about 50 line pairs/mm the MTF is negative at about 0.1. In this spatial frequency range, there will be a phase reversal, as shown and discussed earlier.

Figure 10.17
Spurious Resolution
Due to Approximately
1.5 Waves of
Defocus



CHAPTER

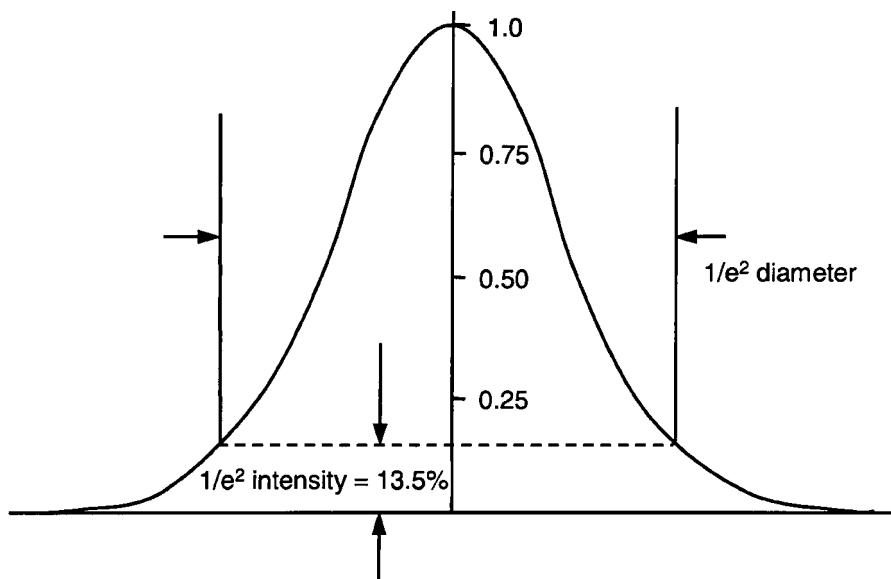
11

Gaussian Beam Imagery

Coherent light generated by lasers has properties different from light generated by other sources which we usually deal with in more conventional optical systems. If we look through a telescope at a distant object, the light intensity across the entrance pupil and aperture stop is uniform, and this is generally known as a “top-hat” intensity profile or distribution. A telescope objective, if it is free of aberrations, focuses a point object into an Airy disk pattern, with the diameter determined by the f/number or numerical aperture of the objective lens. In this case, a uniform top-hat distribution in pupil space transforms mathematically (by the optical system) to an Airy disk in image space. Laser beams emitted from rotationally symmetric resonators, such as HeNe or YAG lasers with a TEM₀₀ output, have an intensity distribution across the beam which is in the form of a gaussian intensity profile, as shown in Fig. 11.1. A gaussian intensity distribution in pupil space will mathematically transform to a gaussian in image space if the beam is not truncated by the aperture of the optical system, which is, of course, different from the uniform pupil transformation. Note that all of the material in this chapter assumes an aberration-free optical system. It is important to include the effects of lens aberrations in the final assessment of image quality and spot size.

The optical design of systems through which laser beams propagate is, therefore, very different from the design of conventional nonlaser systems used in either the visible or some other wavelength region. First, color correction of the laser-based optical system is much easier

Figure 11.1
Gaussian Intensity Distribution



because the wavelength band of the laser light is extremely narrow. For HeNe lasers fully monochromatic light at $0.63282 \mu\text{m}$ can be used. Some lasers emit multiple spectral lines, or wavelengths can be changed. In these cases the optics must be designed to cover the functional spectral bandwidth. Laser diodes sometimes have a wavelength shift with temperature, which also must be taken into account by the optics.

Laser systems are often corrected for a small field of view because the laser beam enters the system parallel to the optical axis. Even if the real usable field of view is zero, the optical design must be optimized over a small, yet finite field of view in order to accommodate assembly and alignment tolerances. If one were to design such a system at identically zero field of view, it is possible that the performance may seriously degrade within 1 or 2 mrad off axis. However, there are systems, such as laser scanning systems, where some components of the system have to be designed for a large field of view. There are also systems where it is required to focus a laser beam to a very small spot, which requires the design of diffraction-limited optics with very large numerical apertures. In some cases, the laser beams have high-power densities, and they can cause damage to the optical components. Transparency of the material

Gaussian Beam Imagery

from which the components are made can be degraded, and the material used for cementing of components and coating of the components can be damaged. The choice of optical materials is thus very important. In some cases, dust particles on the components in the system can also absorb enough energy to damage the surface of the component. Scattering from surface defects is a greater problem in laser systems than in visible and other incoherent systems.

The coherence length of a gaussian laser beam is large, and it appears as if the wavefront emerges from one point. If a gaussian beam is not truncated by the optical system, it emerges from the system as a gaussian beam. The narrow spectral line width of a laser beam and a well-defined wavefront permit very precise focusing and control of the beam.

Beam Waist and Beam Divergence

The beam emitted from a laser in TEM00, or fundamental mode, has a perfect plane wavefront at its beam waist position and a gaussian transverse irradiance profile that varies radially from the axis, which can be described by

$$I(r) = I_0 \exp\left(-2 \frac{r^2}{r_0^2}\right)$$

or by the beam diameter

$$I(d) = I_0 \exp\left(-2 \frac{d^2}{d_0^2}\right)$$

where I_0 is the axial irradiance of the beam, r and d are radius and diameter of a particular point in the beam, and r_0 and d_0 are the radius and diameter of the beam where irradiance is $(1/e^2)I_0$, or 13.5% of its maximum intensity value on axis. To define the propagation characteristics of a laser beam, the value, r_0 , is accepted as a definition of the radial extent of the beam. Finite apertures in the optical system or inside the laser itself, along with diffraction, cause the beam to diverge or converge. There is no such thing as a perfectly collimated beam without any spreading. Spreading of a laser beam is defined by diffraction theory. A laser beam converges to a point where the beam is smallest, called the *beam waist*, and it then diverges from this point

with a full angular beam divergence, θ , which is the same as the convergence angle. The beam divergence angle, θ , is the angle subtended by the $1/e^2$ diameter points in the far field, where the irradiance surface asymptotically approaches the full divergence angle, θ , shown in Fig. 11.2. At the location of the beam waist, the laser beam wavefront is flat, and it quickly acquires curvature on both sides of the waist. The beam waist diameter, d_0 , depends on the full divergence angle, θ , as

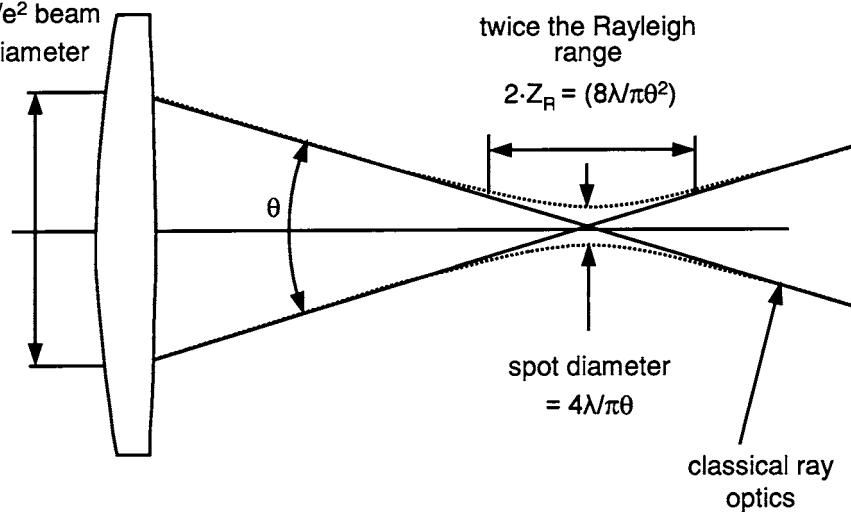
$$d_0 = \frac{4\lambda}{\pi\theta}$$

where λ is the wavelength of the laser beam and θ is in radians. It is important to know that the product of the divergence and the beam waist diameter is constant. The beam diameter grows with the distance, z , from the beam waist according to

$$d(z) = d_0 \sqrt{1 + \left(\frac{4\lambda z}{\pi d_0^2}\right)^2}$$

The radius of the wavefront at the beam waist location is infinite. As we move away from the beam waist, it then passes through a minimum at some finite distance, after which it rises again, approaching the value of

Figure 11.2
Divergence of a Laser Beam in the Far Field



Gaussian Beam Imagery

z as z approaches infinity. The radius of the laser beam wavefront can be expressed as

$$R(z) = z \left[1 + \left(\frac{\pi d_0^2}{4\lambda z} \right)^2 \right]$$

The minimum value of the wavefront radius occurs at what is known as the *Rayleigh range*, where the beam diameter has the value $\sqrt{2}d_0$. At the extremes of the Rayleigh range, the image diameter is thus 41% larger than at the center of the beam waist. The Rayleigh range is sometimes used as a depth of focus number; however, do keep in mind that this reasonably large increase in spot diameter may not be acceptable. The Rayleigh range as measured from the beam waist is

$$z_R = \frac{d_0}{\theta} = \frac{4\lambda}{\pi\theta^2} = \frac{\pi d_0^2}{4\lambda}$$

The wavelength of the laser radiation, the beam waist diameter, the beam divergence angle, and the Rayleigh range are four parameters that completely describe a gaussian beam.

Collimation of Laser Beams

There is no perfectly collimated beam, since the beam divergence and the beam waist are defined by diffraction theory. It can be shown that the minimum beam spread between two points at a distance, z , happens when the starting beam waist is equal to

$$d_0 = 2 \left(\frac{\lambda z}{\pi} \right)^{1/2}$$

This relationship is equivalent to the one that gives the Rayleigh range, which tells us that the Rayleigh range is the distance inside which a gaussian beam has the minimum spreading. The diameter of the beam at the Rayleigh range is

$$d = \sqrt{2}d_0$$

If we use a beam expander at the output of a laser, we can increase the distance of the minimum spreading. If we additionally adjust the

beam expander so that the beam waist is located in the middle of the starting Rayleigh range, then the beam will spread to $\sqrt{2}d_0$ over a distance twice as long.

Propagation of Gaussian Beams and Focusing into a Small Spot

In 1983 S.A. Self derived a simple algorithm for tracing a gaussian beam through an optical system. The formula that he used had a very similar form to the paraxial lens formula that gives the relationship between the object position, focal length of a lens, and the image position. For a gaussian beam, Self calculates the Rayleigh range and the beam waist transformation by each lens in the system:

$$\frac{1}{s + [1/(s - f)]z_R^2} + \frac{1}{s'} = \frac{1}{f}$$

where s is the distance from the lens to the waist on the object side and s' is the distance from the lens to the new waist position. When the incident beam has its waist located in the front focal plane of a positive lens, the emerging beam has its waist at the rear focal plane.

Optical design programs generally use the gaussian beam propagation algorithm derived by A.E. Siegman. These programs usually calculate the radial beam size (semidiameter), the narrowest radial waist, surface coordinates relative to the beam waist, the semidivergence angle, and the Rayleigh range. This assumes the TEM00 fundamental mode, that is, a gaussian irradiance distribution. Lasers may be able to produce a number of other stable irradiance distributions, or modes, but they are not as compact as a gaussian beam, and they all have regions or holes in the beam, that is, where the irradiance drops to zero within the irradiance distribution. The mixed mode beams, defined by the beam quality factor, M , can also be traced in the current generation of optical design programs, in which case the M factor scales the embedded TEM00 gaussian mode.

In many applications, the goal is to focus the laser beam down to a very small spot. If the optical system focusing the beam is diffraction limited, the spot diameter at $1/e^2$ of the peak irradiance is defined by

$$d = \frac{4\lambda f}{\pi d_0}$$

The depth of focus is proportional to the square of the $f/\#$ of the focusing lens. If we define the allowable increase in the diameter of the focused spot, the depth of focus can be calculated using the formula for the spot diameter as the function of the distance from the waist location

$$d(z) = d_0 \sqrt{1 + \left(\frac{4\lambda z}{\pi d_0^2} \right)^2}$$

Truncation of a Gaussian Beam

Let us assume that we have a diffraction-limited lens with a given aperture diameter. The intensity profile of the focused spot is dependent on the intensity distribution of the radiation filling the aperture of the lens. For a uniformly illuminated aperture, the diffraction pattern is the classical Airy disk. It has a central bright spot and progressively weaker rings, with the first dark ring (intensity falls to zero) at a diameter, d :

$$d = 2.44 \lambda \cdot f/\#$$

When the illumination in the aperture is not uniform, the intensity profile in the focused spot does not have zero-intensity points, and the measure of the spot diameter is usually accepted as the diameter at which the intensity drops to $1/e^2$ of the peak intensity. When a gaussian beam falls onto the aperture of a lens, it may be truncated by the lens aperture. Let us define the truncation ratio as

$$T = \frac{d_0}{D}$$

where d_0 is a gaussian beam diameter measured at $1/e^2$ of the peak intensity and D is the aperture diameter of the lens. In the case of the lens aperture being two times the gaussian beam waist diameter ($T = 0.5$), the beam is truncated only below the intensity level of 0.03%. We can say that the effect of truncation is negligible, and the gaussian beam after transformation by the lens remains gaussian. On the other hand, when T , the truncation ratio, becomes a large number, and only the narrow central portion of the gaussian beam is transmitted through the lens, this case corresponds to the uniform illumination of the lens aperture, and the transmitted beam has the intensity distribution similar to the Airy disk. The beam waist diameter, d_0 , given in units of $\lambda \cdot f/\#$ for a few values of

TABLE 11.1

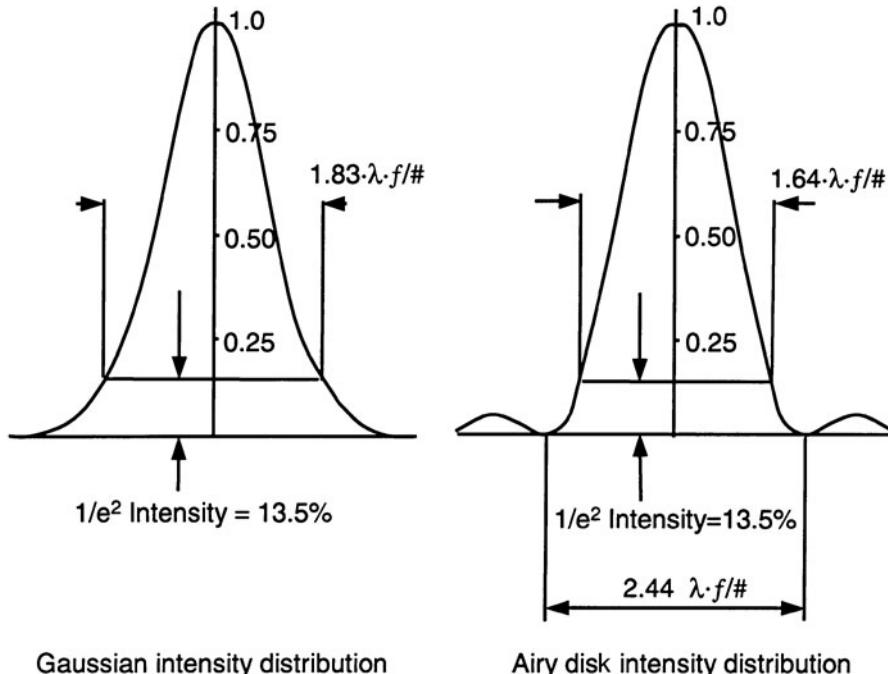
Beam Waist Diameter in Units of $\lambda \cdot f/\#$ for Several Values of the Truncation Ratio

Truncation Ratio T	d_0 at $1/e^2$ Intensity (in units of $f/\#$)	d_0 (Intensity Goes to Zero)	Truncation Intensity Level (%)
$\approx\infty$	1.64	2.44	100
2	1.69	—	60
1	1.83	—	13.5
0.5	2.51	—	0.03

truncation ratio is given in Table 11.1. Two cases of the intensity distribution for $T = \infty$ and $T = 1$ are shown in Fig. 11.3.

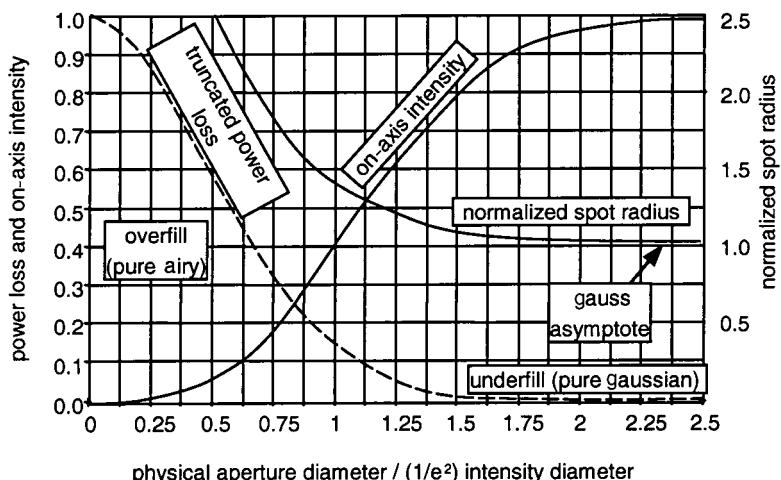
In Fig. 11.4, we show how the truncation ratio affects the performance of an otherwise perfect system with a gaussian intensity profile incident onto the entrance pupil of an imaging optical system. The abscissa is the ratio of the physical aperture to the $1/e^2$ beam diameter. The ordinate on the right is the normalized spot radius (normalized to $1/e^2$ spot radius).

Figure 11.3
Truncation of a Gaussian Beam and the Intensity Distribution at the Image Plane



Gaussian Beam Imagery

Figure 11.4
Truncated Gaussian Beam Imagery. Aperture Changes Relative to a Constant $1/e^2$ Diameter

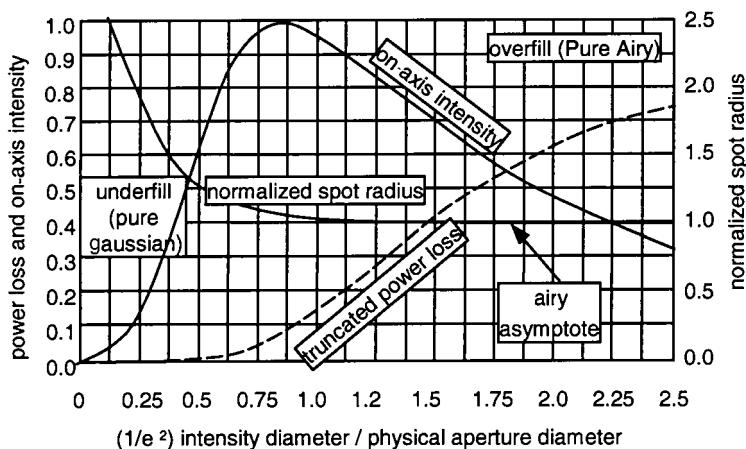


The easiest way to understand this somewhat complex data is to think of a constant $1/e^2$ beam diameter of, say, 25 mm. If the physical aperture diameter were 1 m, there would be virtually zero truncation of the beam, and a gaussian intensity distribution in the entrance pupil would transform into a perfect gaussian image profile at the image. The normalized spot radius asymptotically approaches unity as the physical aperture diameter continues to increase. The far right data point in the abscissa in Fig. 11.4 is for the physical aperture being 25 times the $1/e^2$ beam diameter. If, on the other hand, the physical aperture diameter were in the order of 3 mm for our 25-mm $1/e^2$ beam diameter, the ratio of the physical aperture diameter to the $1/e^2$ beam diameter would be about 0.125 and the intensity profile would be nearly a top-hat, which means that we will acquire diffraction and the imagery will be close to a classical airy disk pattern. It should be clear that if we truncate the beam at the $1/e^2$ beam diameter, the $1/e^2$ diameter of the image will be approximately 40% larger than an untruncated beam.

The ordinate on the left in Fig. 11.4 is for both the power loss and the on-axis intensity with the abscissa having the same meaning as before. As the ratio of the physical aperture to the $1/e^2$ beam diameter increases, the on-axis intensity increases, and the loss in power due to truncation decreases.

In Fig. 11.5 we show a similar plot, only here we have the $1/e^2$ beam diameter changing relative to a constant physical aperture diameter. The interpretation and rationale is the same as for the prior data.

Figure 11.5
Truncated Gaussian Beam Imagery. $1/e^2$ Diameter Changes Relative to a Constant Aperture Diameter



There are several important messages from Figs. 11.4 and 11.5. First, a small beam truncation will affect the $1/e^2$ beam diameter of the resulting image. Truncating the beam to the $1/e^2$ beam diameter will increase the $1/e^2$ spot diameter by approximately 40%. In order to increase the spot diameter by less than 10%, we require a physical aperture about 50% larger than the $1/e^2$ beam diameter. The other data which are of significance are the intensity and power loss associated with a given truncation factor. If our physical aperture were 50% larger than the $1/e^2$ beam diameter, our on-axis intensity would be 80% of that from an untruncated beam.

Application of Gaussian Beam Optics in Laser Systems

Gas lasers have their place in many applications in the large consumer market. HeNe lasers, which emit red light at $\lambda = 632.82$ nm, are used in bar code readers, the printing industry, machine vision, etc. They are used wherever the packaging constraints are not too tight and the use of visible light is convenient. The beam out of the HeNe laser is TEM₀₀ rotationally symmetric gaussian beam, which can be easily transformed with the properly designed optics to the required spot size.

Solid-state lasers can generate very powerful beams. They can emit TEM00 gaussian beams. There are a significant number of types of solid-state lasers that emit in the near IR spectral region. A YAG laser, which emits light at the wavelength of $1.064\text{ }\mu\text{m}$, can be frequency doubled to produce green light at 532 nm. It can also be frequency tripled or quadrupled. YAG lasers are used in industrial applications where high power is needed, such as writing on metal, welding, cutting, hole drilling, etc. One disadvantage of solid-state lasers is that they are generally expensive. One field that makes an extensive use of lasers is medicine, especially in the area of diagnostics, cancer treatment, and eye surgery. The laser wavelengths used range from the UV below 200 nm to $10.6\text{ }\mu\text{m}$ in the far infrared. CO₂ gas lasers emit radiation at $10.6\text{-}\mu\text{m}$ wavelength, and they are widely used both in medical and industrial applications.

UV lasers are used in lithography to achieve the submicrometer imagery needed to make integrated circuit chips. The minimum spot size is determined by the wavelength and the numerical aperture of the focusing optics. The shorter the wavelength, the smaller the spot to which the laser beam can be focused. This pushes the microprocessor industry to use increasing shorter wavelengths. Currently, a common wavelength is 193 nm.

Laser diodes are undergoing an extremely rapid development. Advantages of laser diodes include their low cost and small physical size. They are also available in a wide range of wavelengths. Laser diode devices provide continuous output power or may be analog or digitally modulated. Pulsed laser diodes typically operate with pulses shorter than 100 ns. Laser diodes can be packaged in the form of a linear diode array or a two-dimensional array. Laser diodes may be temperature tuned by approximately $0.3\text{ nm/}^{\circ}\text{C}$. Applications of laser diodes range from telecommunications, data communications, sensing, thermal printing, laser-based therapeutic medical systems, and, satellite telecommunications to diode pumping of solid-state lasers.

The primary optics-related disadvantage of laser diodes is that they emit nonrotationally symmetric beams, which are more difficult to collimate or focus into a small spot. Most beams from laser diodes have a gaussian intensity profile along one axis, often called the *fast axis*, and a nongaussian intensity profile along the *slow axis* perpendicular to it. The beam diverges much faster in the fast axis than in the slow axis. Achieving collimation of the beams from laser diodes is not a trivial task because of the lack of the rotational symmetry in the beam. It requires

anamorphic optics in most cases treating the fast and the slow axes separately. A good way to collimate a beam from a laser diode is to use two crossed glass or gradient index fibers, perpendicular to the optical axis of the laser beam shown in Fig. 11.6, each fiber having a different focal length. This is an elegant solution because of the small packaging. Two perpendicular fibers have small focal lengths. The fiber that collimates the fast axis has a focal length of a few hundred micrometers, and can be a gradient index fiber with a radial gradient index distribution, or a planohyperbolic homogeneous fiber. The fiber that collimates the slow axis has a longer focal length. In this way, the output of most laser diodes can be transformed into near rotationally symmetric beams. The beam can then be focused with a rotationally symmetric lens into a single-mode or a multimode fiber core that transmits the laser signal along the fiber axis.

A very convenient way of transmitting laser beams is through fibers. The laser beam from a laser diode can be coupled into a fiber with a coupling efficiency as high as 90% and transformed to a clean beam, easier to collimate at the output of the fiber. In telecommunications, the laser light-carrying signals are transmitted through hundreds of kilometers of fiber.

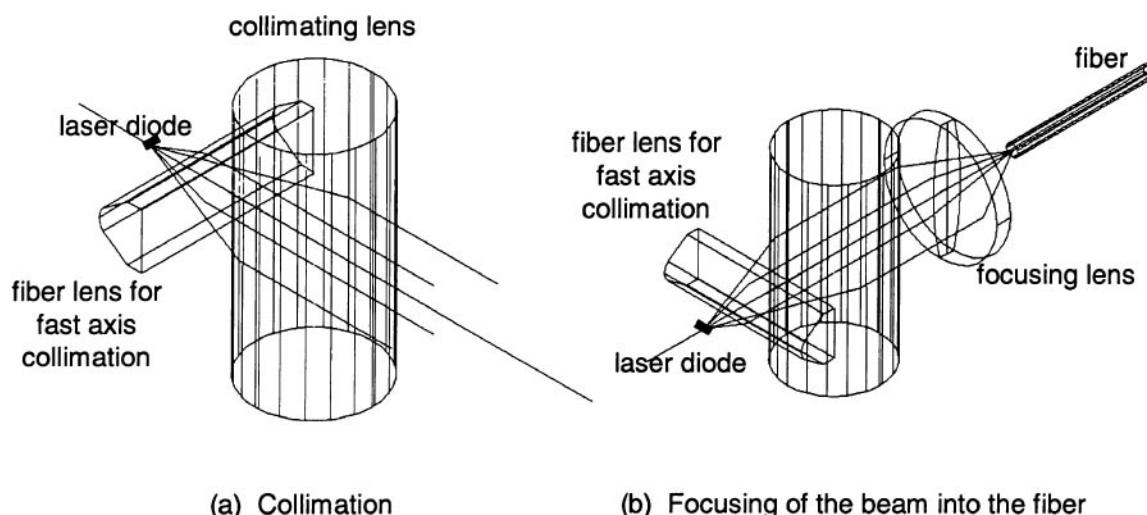


Figure 11.6

Collimation and Focusing of the Beam from the Laser Diode

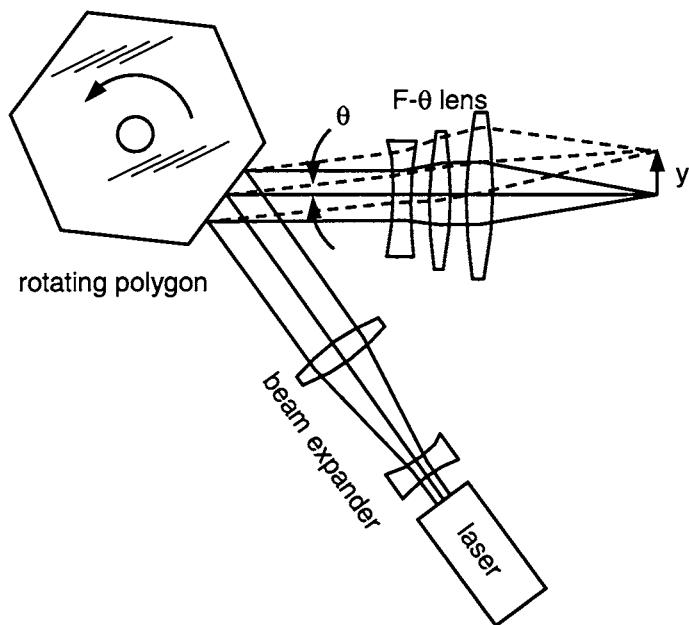
Gaussian Beam Imagery

F-θ Lenses in Laser Scanners

Laser scanners represent one of the more common applications of gaussian beam optics, and a very special lens form is required in these laser scanning systems. Figure 11.7 shows a very generic laser scanning system. An HeNe or similar laser beam is first expanded by a beam expander, and the expanded laser beam is directed toward a multifaceted polygon scan mirror. As the mirror rotates at a constant rotational speed, angle θ changes linearly.

Most lenses follow the relationship that the image height $Y = f \tan \theta$. This is literally true for lenses with zero distortion. If such a lens were used in our laser scanner, then the velocity of the scanned laser spot would increase in proportion to the tangent of the scan angle. Since most laser scanners require a linear spot velocity, conventional lenses are not viable. Lenses which follow the relationship $Y = f \theta$, will produce a linear scan velocity, and these lenses are called *F-theta*, or *F-θ*, lenses. In effect, they are lenses in which negative or barrel distortion is intentionally introduced in order to counteract the increased image

Figure 11.7
Laser Scanner with
F-θ lens



velocity in conventional low-distortion optics. Fortunately these forms of remote-aperture stop lenses (the stop is on the polygon facet) tend to inherently have negative distortion. It is critical, however, to assure uniform spot size through scan, and this is sometimes challenging. In order to loosen the tolerance on pyramidal error of the polygon, the collimated beam from the laser is often focused with a cylindrical lens to a line on the polygon facet. This requires an anamorphic $F\text{-}\theta$ lens to focus the beam to a spot. This form of optical system is used in laser printers.

CHAPTER

12

Basics of Thermal Infrared Imaging in the 3- to 5- and 8- to 12- μm Spectral Bands (Plus UV Optics)

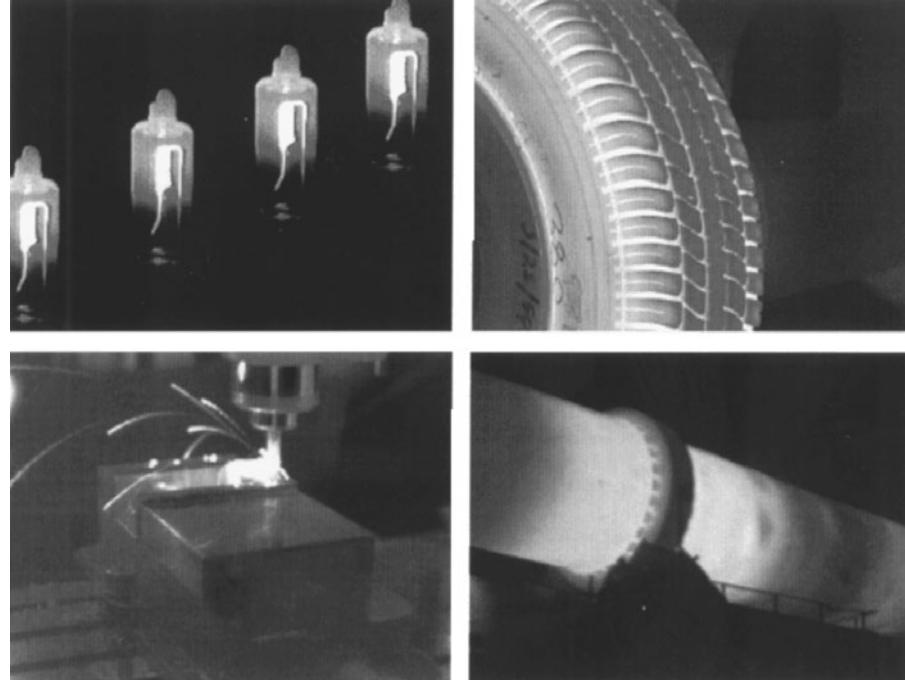
The Basics of Thermal Infrared Imaging

Thermal infrared imaging is generally considered to be in the medium or midwave IR (MWIR) that extends from 3 to 5 μm and in the long-wave IR (LWIR) that extends from 8 to 14 μm . In these wavelength bands, we are looking at thermal or heat sources rather than visible light. There are many different applications for thermal infrared imaging such as nondestructive testing whereby an IR camera can image a machine, such as a CNC lathe, to look for overheating of the bearings, or we can image houses looking for heat losses in the winter. In the medical arena,

doctors can look for various abnormalities indicated by localized skin temperature variations. In nuclear power plants IR cameras are invaluable to quickly search for thermal leaks in the cooling system. Boarder control and security are other areas where IR imaging has become crucial. Figure 12.1 shows several industrial and commercial examples of thermal infrared imaging. The applications of thermal infrared imaging are continuing to develop at a rapid pace.

The human eye is sensitive to the spectral band from approximately 0.4 to 0.7 μm , and thus the eye cannot see this longer-wavelength thermal energy. It takes special detectors or sensors to record the energy, and, needless to say, the imaging optics must efficiently transmit these wavelengths. While there are many applications for the near infrared (NIR), which includes the regions from 0.85 to 1.6 μm , for telecommunications as well as 1.06 μm , which is the Nd: YAG wavelength often used in the applications where the higher power is needed, for the most part ordinary optical materials can be used. The near IR will thus not be considered in this chapter, but rather the MWIR and LWIR where special optical materials and other design considerations are mandatory.

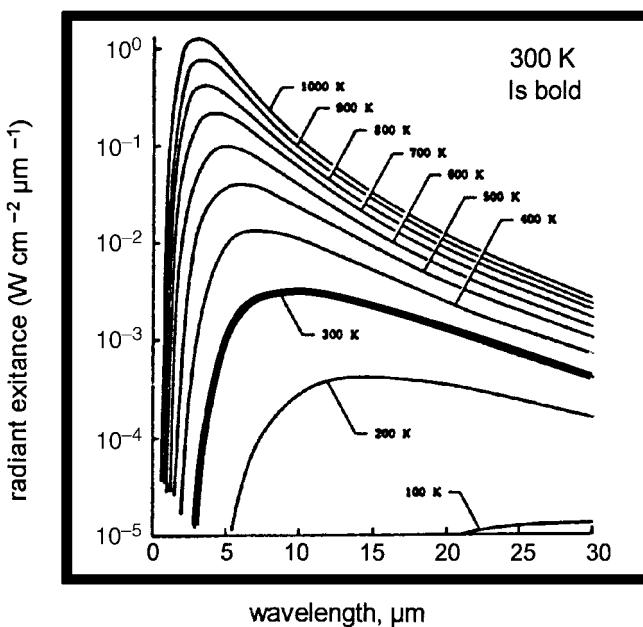
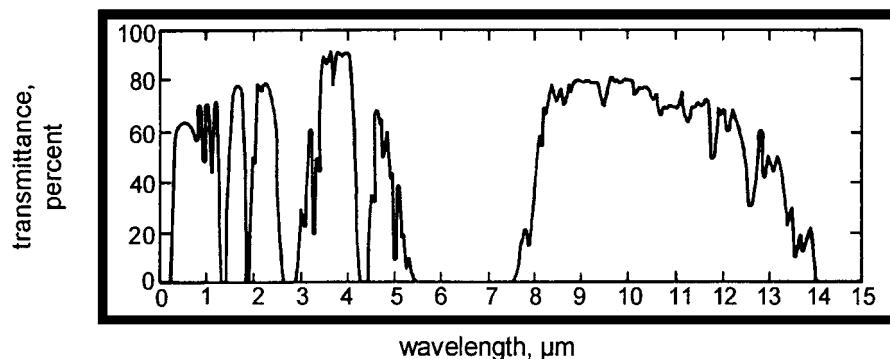
Figure 12.1
Examples of Industrial Applications of Thermal Infrared Imaging
(Courtesy of FLIR Systems, Boston)



Basics of Thermal Infrared Imaging

We see in Figure 12.2 the spectral transmittance of a 1.8-km air path in the 3- to 5- μm MWIR and 8- to 14- μm LWIR spectral bands. Water and CO_2 absorption bands limit the use of wavelengths to these two bands within the atmosphere. We also see the radiant exitance for black bodies ranging in temperature from 100 to 1000 K. For reference, ambient temperature is about 300 K, and you can see that the peak radiant exitance at this temperature occurs at about 10 μm . For this reason, LWIR systems tend to have the highest sensitivity. However, LWIR detectors

Figure 12.2
Top: Transmittance of
1.8-km Horizontal Air
Path at Sea Level;
Bottom: Radiant
Exitance of Blackbody
As Function of
Temperature



are more expensive and difficult to produce than their MWIR counterparts. In addition, with today's image-processing hardware and algorithms, excellent MWIR imagery can be obtained.

In a simplified form shown by Riedl, the signal-to-noise ratio for an IR system is

$$S/N = (W_T \epsilon_T - W_B \epsilon_B)(\tau) \left(\frac{D^*}{\sqrt{\Delta f}} \right) \left[\frac{\tau d'}{4(f/\#)^2} \right]$$

where ϵ = emissivity

W = radiant exitance (W/cm^2)

W_T = target exitance (W/cm^2)

ϵ_T = target emissivity

W_B = background exitance (W/cm^2)

ϵ_B = background emissivity

D^* = specific detector detectivity ($\text{cm} \cdot \text{Hz}^{1/2} \cdot \text{W}^{-1}$)

Δf = noise equivalent bandwidth (Hz)

τ = optical transmission

d' = detector size (cm) assuming square detector

$f/\#$ = f /number of optics

The first factor in the previous equation relates to the object we are imaging. This factor gives us the exitance difference between the primary object of interest or target and the area surrounding the object of interest or background. The second factor is the transmission of the atmosphere or other medium in which the system is immersed and the transmission through the optical elements. The third factor relates to the focal plane array and is the detectivity divided by the noise equivalent bandwidth. The fourth and final factor has the sensor size (the width of a pixel) and the optical transmittance in the numerator and the $(f/\#)^2$ in the denominator. This is where the optics becomes critical since the signal to noise is inversely proportional to the square of the $f/\#$. This drives many IR systems to extremely low $f/\#$ s in order to reach the desired signal-to-noise ratio. In addition, with some of the new uncooled microbolometers, the $f/\#$ often needs to be $f/0.8$ or even lower. The reader is referred to the reference for further information on this important relationship.

With respect to the optics, most glasses simply do not transmit above about $2.5 \mu\text{m}$. Certain special glasses do transmit up to $4.5 \mu\text{m}$, and fused

silica transmits up to about 4 μm . Infrared-transmitting materials are therefore essential, and as we will see there is only a limited selection available. These IR-transmitting materials are generally expensive and have other problems.

The Dewar, Cold Stop, and Cold Shield

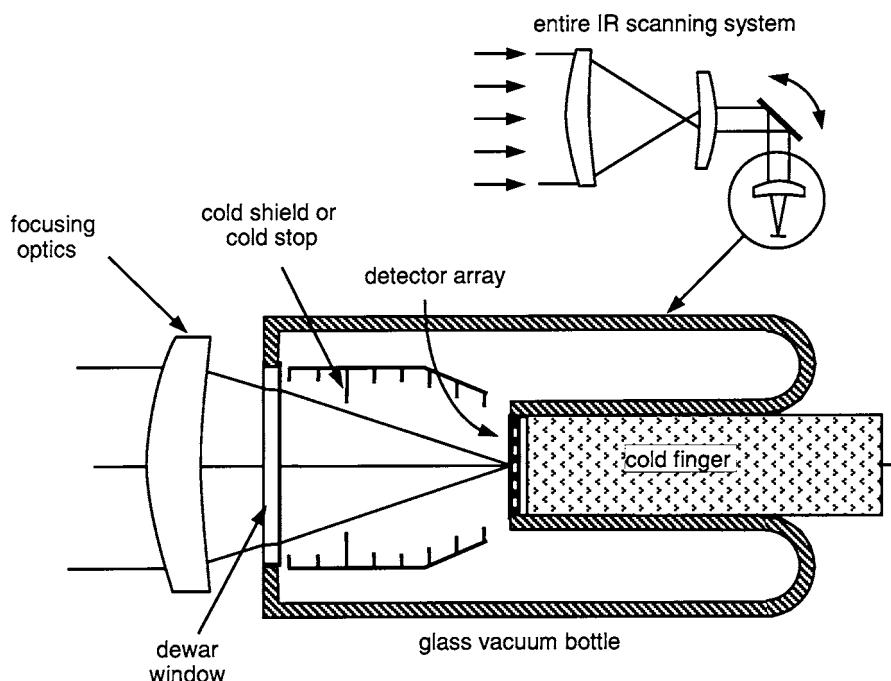
Since we are looking at thermal or heat sources with a thermal-imaging system, for maximum system sensitivity most thermal imaging systems use cryogenically cooled detectors which operate at the liquid nitrogen temperature of 77 K or even lower. If these detectors, or *focal plane arrays* (FPAs), are allowed to “see” any thermal energy other than the energy contained within the scene being viewed, then the sensitivity is reduced. In addition, if the magnitude of this nonscene energy changes or modulates over the field of view, then we often see cosmetically undesirable image anomalies. In order to achieve maximum sensitivity and avoid image anomalies, the IR FPA is cryogenically cooled and mounted into a thermally insulated “bottle,” or *dewar*, assembly.

Figure 12.3 shows a typical generic detector/dewar assembly intended for an IR-imaging application. Before we show how the dewar works, we need to see just how it interfaces with the rest of the optical system. The smaller figure on the upper right of Fig. 12.3 shows an entire scanning-imaging IR system. Light (actually infrared radiation) enters from the left into the larger lens generally called the *collecting optics*. After forming an intermediate image, the light is collimated by the second smaller lens. A further purpose of the second lens is to form an image of the larger collecting optics element, which is the system aperture stop, onto the scan mirror. After the light reflects from the scan mirror, it enters the region within the circle in Fig. 12.3. This area is shown enlarged and in more detail in the larger figure.

Collimated light from the scan mirror first enters the *focusing lens*, which is generally outside of the cryogenically cooled dewar and which focuses the light onto the FPA after passing through the dewar window. Note that for the example shown in Fig. 12.3 the detector array is a linear array extending in and out of the figure. The following attributes make up the dewar assembly:

Figure 12.3

Top: Entire IR Imaging System; Bottom: Typical Generic Detector Dewar Assembly



- The dewar is an evacuated bottle very similar to a classical Thermos bottle.
- The entrance window must, of course, transmit infrared radiation.
- A *cold finger* butts up against the aft end of the FPA to keep it cryogenically cold. The cold finger itself is a high-specific-heat metal rod of iron or steel, which is wrapped with a coil through which liquid nitrogen is pumped (or other similar operation). This, in turn, cools the aft end of the FPA.
- A baffle, called a *cold shield* or *cold stop*, is located as shown inside the dewar. We will define these later.

Now—an *extremely important point*—to evaluate what is happening with respect to the imaging light and potential stray light which may lead to undesirable image anomalies, you need to figuratively *put your eye at the detector and look out the front and ask yourself "what do you see?"* Let's do just that and put our eye figuratively at the center of the FPA where the light focuses.

- Within the solid angle forming the imaging cone of light, we “see” *scene energy*. We are, in effect, looking into the exit pupil of the optics, and we see a solid angle of radiation coming from the nominal scene. We do not see the image, but rather the exit pupil. This is much like looking toward a round porthole window in a ship from some distance away.
- At angles just outside the imaging cone, yet not quite hitting the cold shield baffle, we have radiation, which is not scene energy, nor is it cryogenically cold. This sliver of solid angle, which is a circular annulus, represents energy from the interior of the system, which literally reaches the FPA.
- Outside this sliver of interior system energy, we see the cold shield or baffle within the dewar. Since this component is cryogenically cold and finished with a nonreflective coating, it emits little or no radiation.

If, in the previous example, the cold shield had been the same diameter as the light cone, then the detector would only have been able to see scene energy. In effect, we have in this example made the cold shield slightly larger in diameter than it needed to be, and we will show why later. For now, we simply conclude that there is a sliver of solid angle, which is outside the scene energy and inside the cold shield, which can record energy from the system interior.

Cold Stop Efficiency

An IR system is said to be *100% cold stop efficient* if the detector can see or record energy only from the scene. What this really means is that with 100% cold stop efficiency, the detector records energy from both the cone of light representing scene energy and from the cryogenically cold thermal baffle, known as a cold stop (remember that being cryogenically cold, there is virtually no energy emitted from the cold stop itself). We have used the example in the previous section of putting your eye figuratively at the detector and looking out toward the front of the system and asking yourself “what do you see?” If, for every pixel on your FPA, you can convince yourself that your eye sees only the solid angle representing the imaging light (scene energy) and also portions of the thermal baffles representing the cold stop, then the system is indeed 100% cold stop efficient.

Note in Fig. 12.3 we have shown a series of small stray-light baffles within the cold stop or cold shield area. Without these stray-light suppression features, there may be stray radiation paths which will cause unwanted radiation to reach the FPA.

In Fig. 12.4 we show on the left a system which is not 100% cold stop efficient and one which is 100% cold stop efficient on the right. Casually looking at these figures shows little difference. The lower sets of figures are enlargements of the areas within the dashed circles of the upper figures. Note on the left figure how the aperture stop is both on the front element of the system as well as on the rear surface of the two-element reimaging lens group. Furthermore, if we place our eye at the lower end of the FPA as shown and look toward the scene, we see the solid angle representing the scene as well as a solid angle above this region and below the cold shield which is not coming from the scene but rather from some portion of the system interior. This nonscene energy is, in effect, analogous to stray light in a conventional

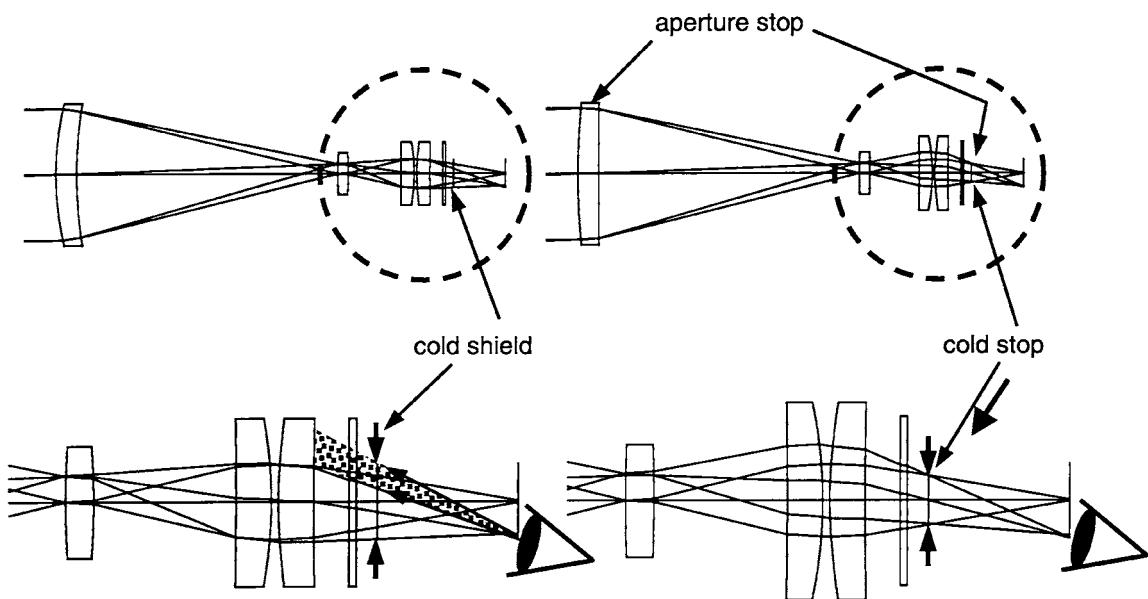


Figure 12.4

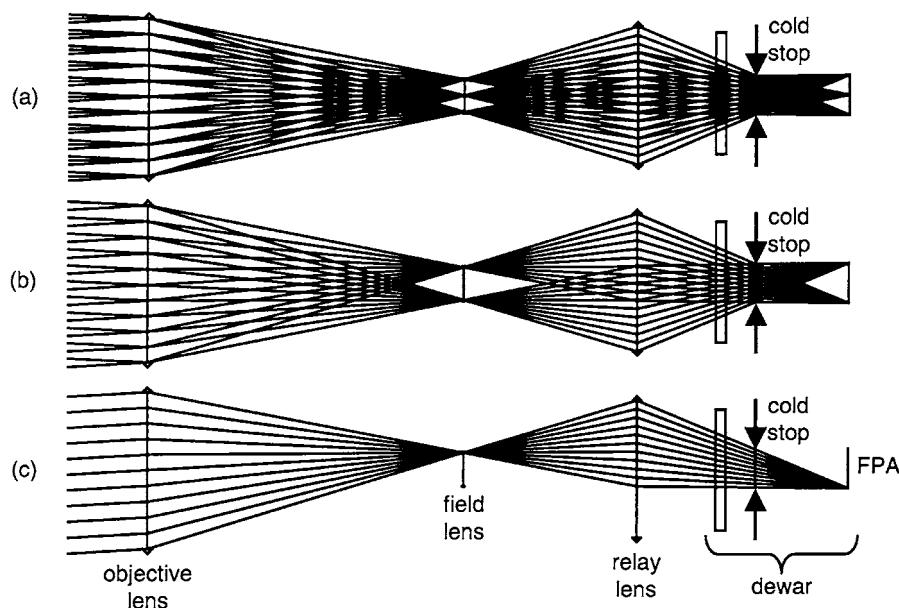
Left: System Which Is Not 100% Cold Stop Efficient; Right: System Which Is 100% Cold Stop Efficient

optical system operating in the visible portion of the spectrum. If this nonscene energy is “warm,” then the sensitivity of the detector will be reduced from its nominal value; however, if this nonscene energy changes in magnitude over the FPA or through scan, then you will have image anomalies similar to ghost images in a conventional visible system. This is bad, especially when the image anomalies are confused with the real scene which can sometimes happen.

On the right-hand figure in Fig. 12.4, we see how the aperture stop at the front element is reimaged into the cold stop plane inside the dewar assembly. Here if we look out from anywhere on the FPA, we see only scene energy and no system interior energy whatsoever. This system is said to be 100% cold stop efficient. The *cold stop efficiency* is the ratio of the total solid angle reaching a given pixel which comes from the scene to the total solid angle reaching the same pixel from the entire opening in the thermal baffle or cold shield. For example, we might find that the solid angle reaching a given pixel from the scene is 90% of the total solid angle which can possibly reach the same pixel within the thermal baffle. This system has 90% cold stop efficiency. The photons or thermal energy which come from some interior portion of the system or its housing is unwanted energy in the form of stray light. Its presence lowers the net sensitivity of the system, but even more importantly if the magnitude of this nonscene energy changes or modulates over the field of view, or through scan if we have a scanning system, then we likely will see image anomalies on the video display.

Figure 12.5 shows a paraxial IR lens system that is designed to be 100% cold stop efficient. In order to make it clear just how the ray paths proceed through the system, we show three permutations of the design. Note that the design itself is identical in all three cases. Figure 12.5a shows three field angles, on axis and $\pm 3^\circ$; Fig. 12.5b shows only the extreme fields of view; and Fig. 12.5c shows only the maximum field of view. Note that in all three cases the front objective lens is reimaged into the cold stop plane. Rays from various fields of view are shown all superimposed. It is clear how the ray cones all overlay on the front element and also on the cold stop surface. It should also be clear how, if you locate your eye anywhere along the FPA and look toward the front of the system, the only solid angle you can see is from the solid angle of image-forming energy. Anything outside these solid angles will view the aft side of the cold stop and/or the interior of the cryogenically cold dewar.

Figure 12.5
Paraxial IR Lens with
100% Cold Stop
Efficiency



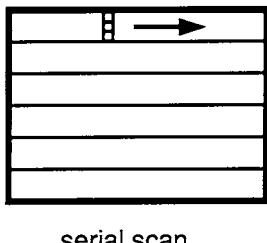
Scanning Methods

A typical imaging sensor, such as a CCD or a CMOS chip in a commercial camcorder or other video-imaging system, uses a full two-dimensional detector array or focal plane array (FPA). Very much analogous to film, these FPAs record the entire image essentially at once. In the thermal infrared these FPAs are called staring, mosaic, or two-dimensional (2-D) detectors.

Infrared detectors are still quite costly and difficult to manufacture, and for this reason detector arrays which are much smaller in extent than a full two-dimensional array are often used with appropriate scanning to allow imaging coverage over the full desired two-dimensional field of view. If we use a small detector array, we can create a full two-dimensional field of view by following the steps outlined here and illustrated in Figure 12.6.

- Scan the field of view in the azimuth or horizontal direction over the full width of the field. This is the upper swath shown in Fig. 12.6.
- Then simultaneously increment the field down by the vertical extent of the array while reinitializing to the original azimuth position on the left.

Figure 12.6
Serial Scanning



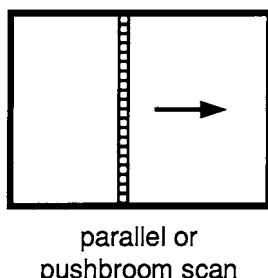
- Now we scan again in azimuth.
- This procedure is repeated until we cover the full vertical or elevation field of view, after which the entire process is repeated again and again to create the full two-dimensional field of view.

This process is known as *serial scanning*. It requires two scan motions, one for the azimuth or horizontal scan and another for the elevation or vertical scan. A good analogy to serial scanning is mowing your lawn. If you have a large rectangular lawn area, you likely use a lawn mower with a width of 0.5 m or so. And you likely mow in one direction, increment by the width of the mower at the end, and mow back. This is known as a *bidirectional scan* in optical terms. Needless to say, you can mow a large area lawn with a small lawnmower, just like covering a large field of view with a small detector array. The previous methodology can be used with very small arrays with as few as one element or pixel or with larger arrays, with up to 25% of the vertical field of view or more. Electronics called *scan-converting electronics* is used to combine or multiplex the data streams and create a virtually seamless standard video signal such as composite video, NTSC, etc.

Operationally, serial scanning requires two scan motions as discussed previously, horizontal and vertical. This can be implemented using two scanning or moving mirrors, one for the scan in the horizontal direction and the other in the orthogonal vertical direction. While, theoretically, both functions could be accomplished with one mirror scanning in both directions, this is rarely done due to the high bandwidths and other difficulties of the two required motions. And do keep in mind that the scan motion in the scan direction should be linear in its angular velocity.

A second form of scanning is known as *parallel* or *pushbroom* scanning, as shown in Fig. 12.7. Here we use a long detector array covering the entire vertical field of view, and the scanning motion is in the

Figure 12.7
Parallel or Push
broom Scanning



azimuth or horizontal direction (or conversely). For this form of scanning only one scan mirror is required. Comparing this with a serial scan system shows the parallel or pushbroom system requires one scan mirror motion and uses an FPA with more elements. The serial scan system uses two scan mirrors and has more complex scan-converting electronics but a far simpler detector. It is clear that there are many trade-offs to be made in selecting the optimum configuration for a given application.

It is important to note that while we are describing scanning over the field of view in the azimuth and elevation directions, what really happens in the optical systems is quite different. What literally happens is that the optical system is, in reality, quite stationary in space during this scanning operation (except of course for the scan mirror itself). The motion of the scan mirror or mirrors scans or translates the image of object space over or across the image plane or FPA. So if your eyes could see in the thermal infrared and you were able to view the focal plane area, what you would see is the image being translated or scanned across the FPA in azimuth and elevation (unless it is a parallel scan operation in which case the image is moved only in the one direction).

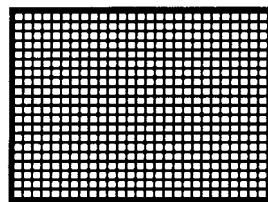
Staring or mosaic arrays are full two-dimensional focal plane arrays which sense the entire scene instantaneously with no mechanical scanning required. Figure 12.8 shows such an FPA. Here we are trading off a more complex sensor for simplified mechanics and scan-converting electronics. It is interesting to note that every commercial camcorder and digital camera uses staring arrays without scanning. The analogous form of sensors are, of course, far more costly in the thermal IR.

We will now show how we can implement a scanner in an IR system for a pushbroom scan. Consider Fig. 12.9 where we show what is, in effect, an astronomical telescope. The system takes collimated radiation of diameter D covering a field of view of $\pm\alpha$ and outputs collimated radiation of diameter d covering a full field of view of $\pm\theta$. For this system the

Basics of Thermal Infrared Imaging

Figure 12.8

Staring Array (No Scanning Required)



magnification, M , can be stated as D/d or equivalently θ/α . In order to scan over the $\pm\alpha$ full field of view, we could do either of the following:

- Place a large flat mirror in front of the system and scan it by $\pm\alpha/2$.
- Place a small flat mirror aft of the second collimating lens at the exit pupil and scan it by $\pm\theta/2 = \pm\alpha(D/d)/2 = \pm\alpha M/2$, where M is the magnification.

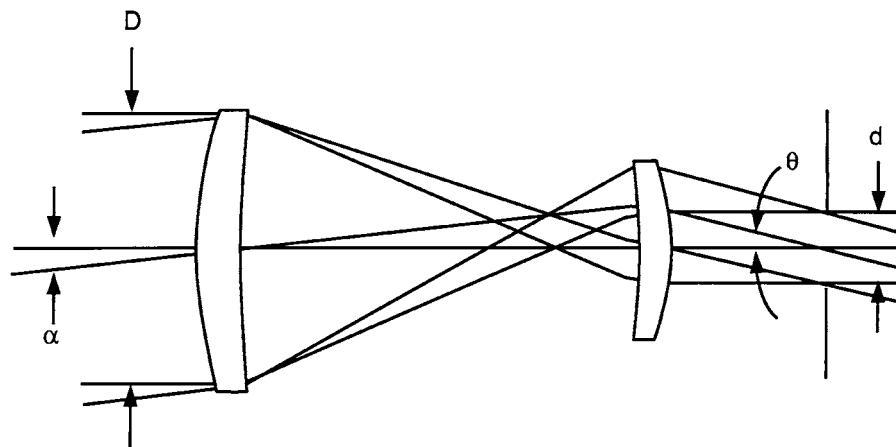
Consider the following example: Assume D is 100 mm in diameter and the magnification is $10\times$. Thus, $d = 10$ mm. And assume that $\alpha = \pm 2^\circ$, which means that $\theta = \pm 20^\circ$. We can now accomplish our scanning with either of the following scenarios:

- We can use a mirror approximately 100×140 mm and scan it by $\pm 1^\circ$.
- We can use a mirror approximately 10×14 mm and scan it by $\pm 10^\circ$.

While both of these methods will work, there are many reasons why scanning the smaller mirror by the larger angle is preferred. The smaller

Figure 12.9

Scanning Methodology



mirror is much easier to manufacture. If we require a given surface flatness on the mirror, this requirement will hold for the larger mirror as well as the smaller mirror, and the smaller mirror will be far less costly. Also, the smaller mirror system will be far better from a packaging point of view.

We show in Fig. 12.10 an oscillating mirror for a pushbroom scanning system as described previously. The scanning mirror is located at the exit pupil of the telescope as we described earlier. In Fig. 12.11 we show how a polygon mirror can also be used for a pushbroom scanner. A polygon mirror is a very interesting device. Each facet is, in effect, a separate scan mirror. And each facet, as it scans, rotates not about the vertex of the facet but rather about the center of rotation of the entire polygon. This creates a different motion in space and must be properly modeled to assure proper system performance. Note in Fig. 12.11 that we show the mirror in its "neutral" position as the top mirror. Once again, let's put our eye at the FPA and look outward. We reflect off of the prime facet to the left and ultimately view object space. Everything is fine. Now look at the bottom polygon mirror, which is shaded. The polygon has rotated around in a counterclockwise direction in the order of 20° , and we show it in a position where the next facet behind the prime facet is just allowing a sliver

Figure 12.10
Oscillating Mirror for
Pushbroom Scanning

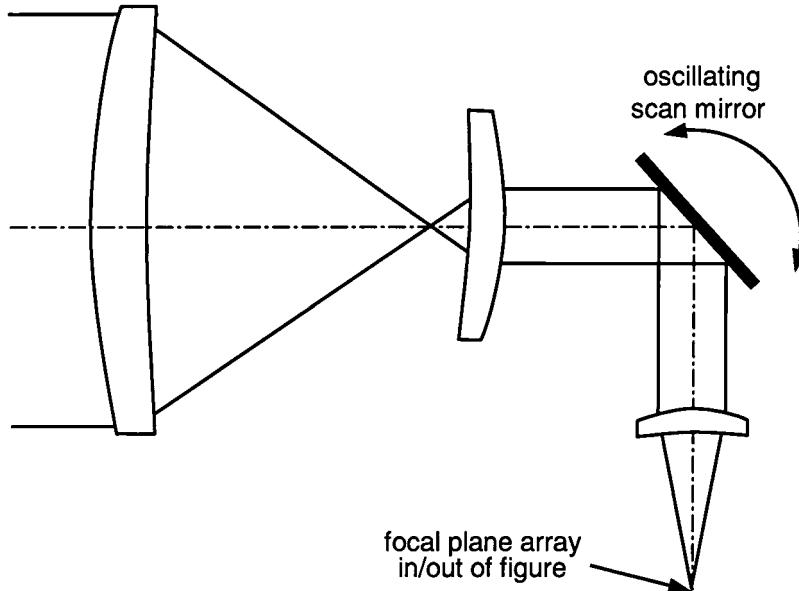
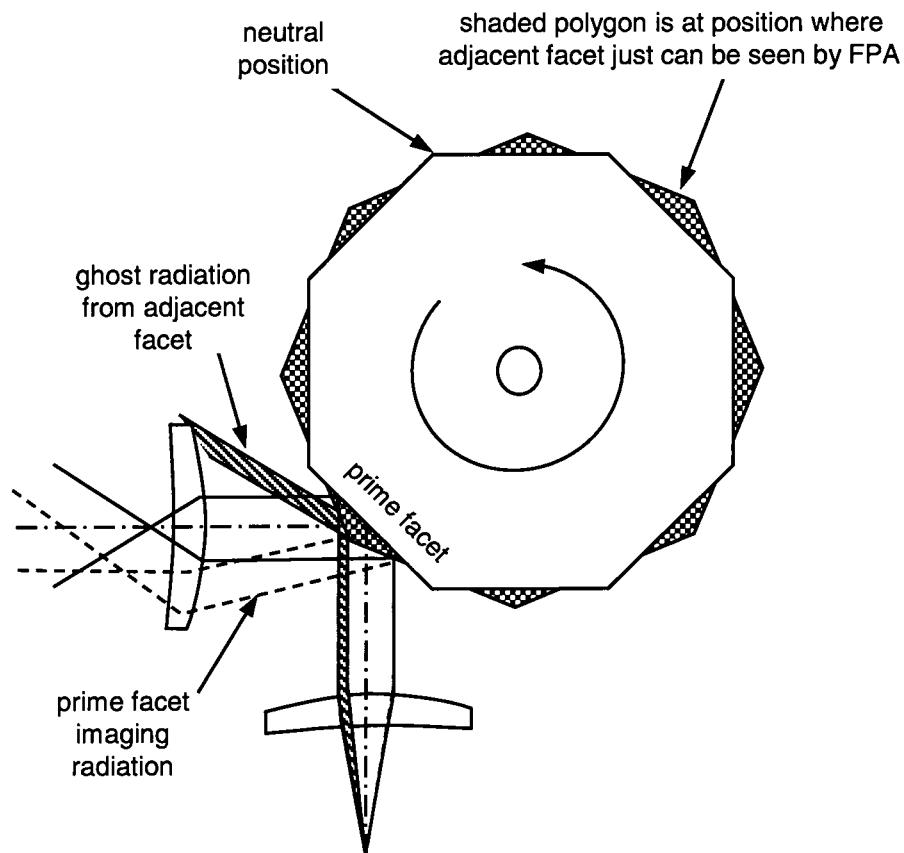


Figure 12.11
Polygon Mirror for
Pushbroom Scanning

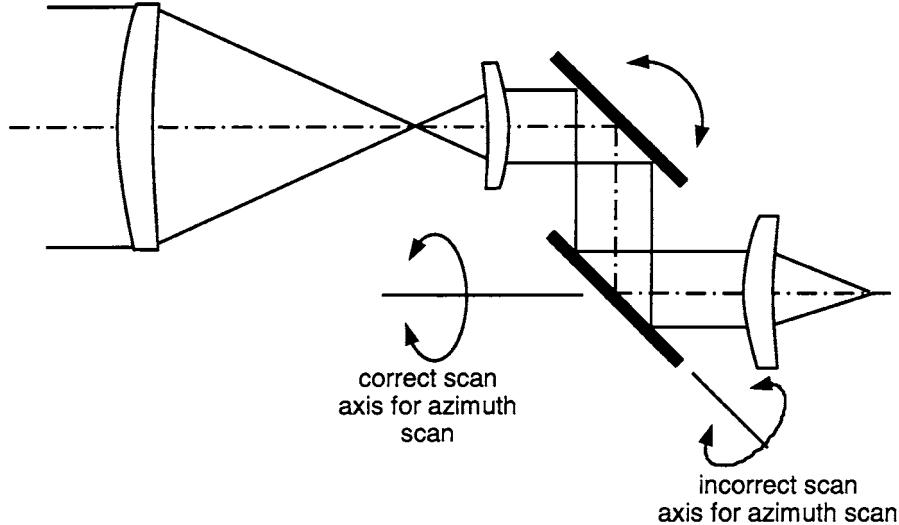


of radiation to be reflected from it. With the polygon in this position, the prime imaging facet reflects the light down to the left while the adjacent facet reflects the sliver of light up to the left. In a situation as shown, the angle of the radiation from the adjacent facet is likely larger than that from the prime facet, and for this reason the radiation may or may not actually reach object space. If it does not, it will strike an interior structure of the system and may create an image anomaly. This reflection from the adjacent facet is known as *ghosting*, and if the ghost radiation makes it to object space it is called *external ghosting* while if it strikes an interior portion of the system it is *internal ghosting*. In order to prevent ghosting, the electronics is often shut off just prior to the adjacent facet enters into the imaging beam of radiation. Occasionally, one can build a valid case to allow a small fraction of the pupil to be ghost

radiation, but it really does need to be small. The fraction of the time that the polygon is actually being used for imaging relative to the total amount of time it is running is called the *scan efficiency*. Scan efficiencies in the vicinity of 80% are not uncommon.

Figure 12.12 shows how we can create a serial scan motion. Recall that in a serial scanning system we scan in the azimuth direction while imaging onto a small detector array of only a few elements. We then increment in elevation and scan in azimuth once again. The process is repeated until we build up a full two-dimensional field of view. Such a system requires two scan-mirror motions, one for the azimuth scan and the other for the elevation direction. In Fig. 12.12 we show the upper mirror as the elevation mirror. For the azimuth scanning mirror we show two potentially viable mirror motions, one is incorrect and the other correct. As it turns out, if we rotate the mirror about the incorrect axis of rotation, we will be scanning an arc in the azimuth direction as opposed to a straight horizontal direction, which will result from the correct axis. It is extremely important that you fully understand just how the mirrors in your system work and how they scan in object space. How can we be sure that we do not have a problem? There are several ways. First, you really should have an accurate model of your scanning system in your computer lens design program. Take nothing for granted! You must spend the time to assure that your model is an accurate representation of the real world and show

Figure 12.12
Two Mirrors (Azimuth and Elevation) for Serial Scanning



how the scan motion works. One interesting thought is to set up your system in reverse from the FPA out toward object space. If you do this, you need only trace a central ray along the optical axis rather than a full cone of rays, as this ray will, in effect, be the chief ray at the given field and/or scan position. Now rotate your mirror or mirrors and monitor at what azimuth and elevation the ray leaves the system for object space. This will tell you directly whether you are scanning a straight line or not. One final hint: take some very simple “drugstore” mirrors and make a crude setup in the lab to model the mirror motions. If in such a setup you exaggerate the mirror motions, you can generally get a valid indication of what is going on. Regardless of how you approach the situation, we cannot overemphasize the importance of modeling properly your system with respect to its scanning motions and the resulting imagery into object space.

IR Materials

While there are many glass types available for visible systems, there are only a very limited number of materials that can be effectively used in the MWIR and LWIR spectral bands. Table 12.1 shows the more common materials and their most important properties. Figure 12.13 shows a plot of the transmittance of the more common IR transmitting materials. It is important to note that these data include surface reflection losses, and often a significantly higher transmittance results after applying high-efficiency antireflection coatings.

Figure 12.14 shows a “glass” map where we plot the refractive index in the ordinate versus the $V\#$ in the abscissa for common infrared-transmitting materials. Recall that the $V\#$ is inversely proportional to the material’s dispersion, and note how for germanium is nearly 1000 in the LWIR (very low dispersion) versus about 100 in the MWIR. You can use this glass map in much the same way as you would for visible systems.

We will now discuss each of the common IR materials:

GERMANIUM *Germanium* is perhaps the most common of infrared materials. It is used in both the LWIR where it is the crown or positive component of an achromatic doublet and in the MWIR where it is the flint or negative component of an achromatic doublet. This anomaly is due to the differences in its dispersion properties in the two spectral

TABLE 12.1

Properties of Common Optical Materials in the Thermal Infrared

Material	Refractive Index at 4 μm	Refractive Index at 10 μm	$dn/dt/\text{ }^{\circ}\text{C}$	Comments
Germanium	4.0243	4.0032	0.000396	Expensive, large dn/dt
Silicon	3.4255	3.4179*	0.000150	Large dn/dt
Zinc sulfide, CVD	2.2520	2.2005	0.0000433	
Zinc selenide, CVD	2.4331	2.4065	0.000060	Expensive, very low absorption
AMTIR I (Ge/As/SE:33/12/55)	2.5141	2.4976	0.000072	
Magnesium fluoride	1.3526	†	0.000020	Low cost, no ctg required
Sapphire	1.6753	†	0.000010	Very hard, low emissivity at high temperature
Arsenic trisulfide	2.4112	2.3816	‡	
Calcium fluoride	1.4097	†	0.000011	
Barium fluoride	1.4580	§	-0.000016	

*Not recommended.

†Does not transmit.

‡Not available.

§Transmits up to 10 μm but drops abruptly.

bands. In the MWIR germanium is approaching its lower absorption band and hence the refractive index is more rapidly changing, thus leading to a greater dispersion. This, in turn, makes it appropriate for the negatively powered element of an achromatic doublet.

With respect to germanium's optical properties, two parameters are of major significance. First, the refractive index of germanium is just over 4.0, which means that shallow curves (long radii) are feasible. As we will see later, along with the higher refractive index aberrations are easier to reduce, which is a significant benefit to the designer. Another parameter of significance is the dn/dt , which is the change in refractive index with respect to temperature. For germanium the dn/dt is $0.000396/\text{ }^{\circ}\text{C}$. This is a

Basics of Thermal Infrared Imaging

Figure 12.13

Spectral Transmittance of IR Materials, Including Surface Losses

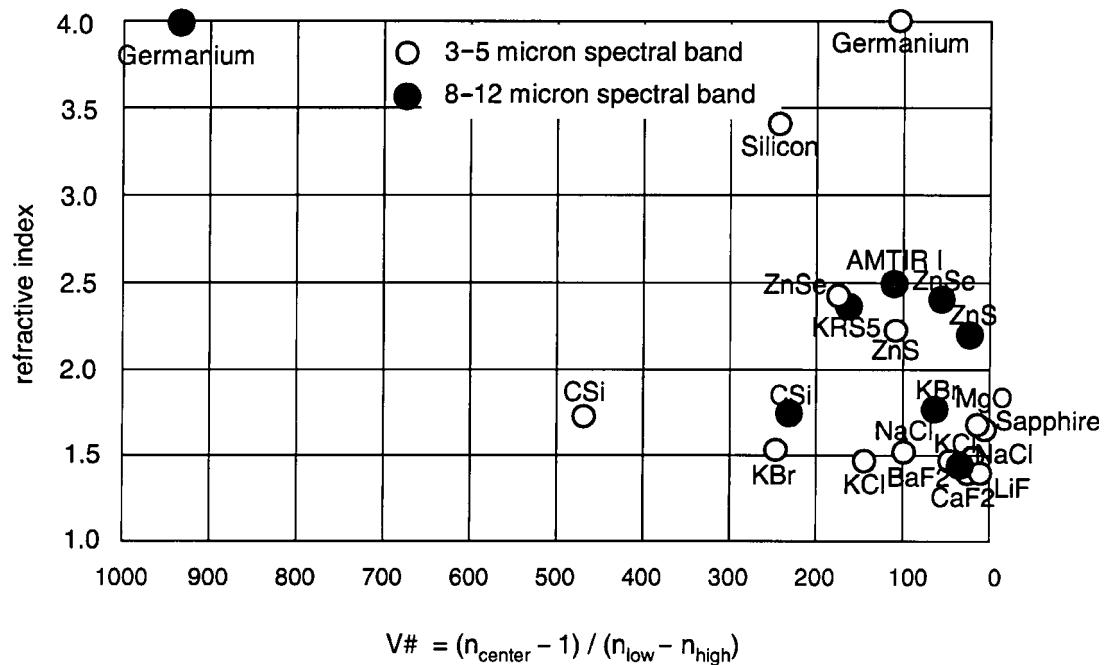
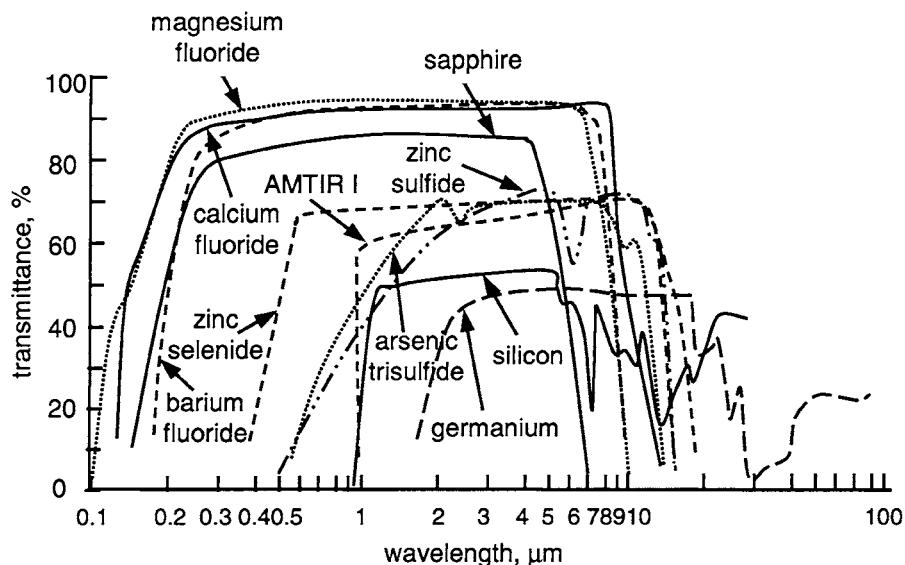


Figure 12.14

Glass Map for Common Materials in the MWIR and LWIR Spectral Bands

very large value, especially when compared with $0.00000360/\text{°C}$ for ordinary glasses such as BK7 glass. This can cause a large focus shift as a function of temperature. This defocus is generally so large that these systems often require some form of athermalization (focus compensation versus temperature).

Germanium is a crystalline material and, as such, can be grown in either polycrystalline form or monocrystalline form, which is also called single-crystal germanium. Depending on the manufacturing and refining processes, single-crystal germanium may be more costly than polycrystalline germanium. Throughout the 1970s and 1980s there was much confusion regarding the relative need for single-crystal germanium in high-performance thermal-imaging systems. For the most part, European designers specified single-crystal material, and in the United States polycrystalline material was generally called for. Studies in the mid- to late 1980s showed that indeed polycrystalline germanium had a larger refractive index inhomogeneity, and this was due primarily to impurities at the grain boundaries. Furthermore, the presence of these impurities could be imaged onto the FPA if the material were at or near an intermediate image plane. The single-crystal germanium is preferred. Fortunately, recent advances in material manufacturing have closed the cost differential gap, and, for the most part, the optics industry uses single-crystal germanium. At high temperatures germanium becomes absorptive, with near zero transmittance at 200°C .

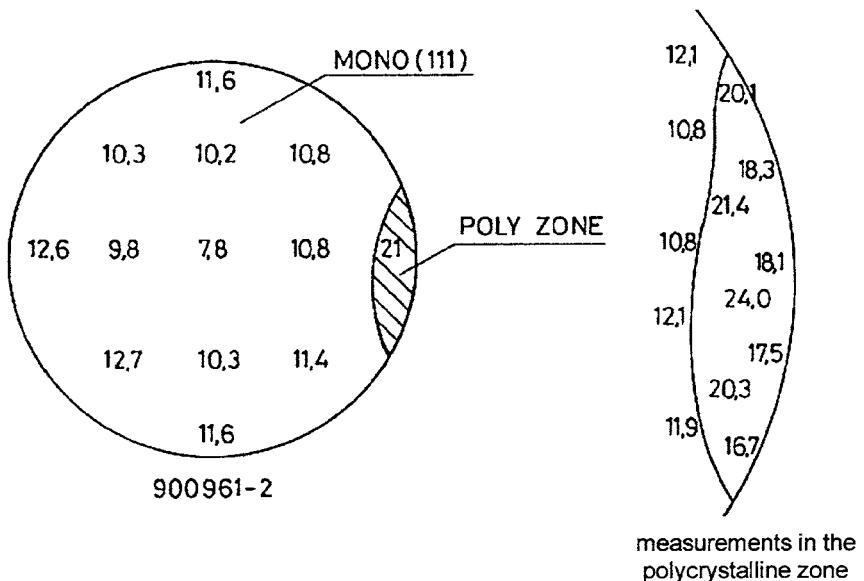
Single-crystalline germanium has a refractive index inhomogeneity of 0.00005 to 0.0001 , whereas polycrystalline germanium is 0.0001 to 0.00015 . For optical purposes germanium is generally specified as to its resistivity in ohm-centimeters, and the generally accepted value is 5 to $40 \Omega \cdot \text{cm}$ throughout the blank. Figure 12.15 shows a typical germanium blank with an area on the right which is polycrystalline. Note that the resistivity is well behaved and slowly changing radially in the single-crystal region, whereas it changes rapidly in the polycrystalline region. If you were to look into the material with a suitable infrared camera, you would see a somewhat bizarre convoluted image resembling cobwebs, with this appearance most accentuated at the grain boundaries. This is all due to the impurities induced at these boundaries.

There is one further comment regarding germanium—its susceptibility to chipping. You must be exceptionally careful during the optical manufacturing and coating processes as well as during assembly as a nearly inconsequential tap to the edge of a germanium element could result in a chip flaking off. For this reason, germanium is often bonded into its

Basics of Thermal Infrared Imaging

Figure 12.15

Resistivity Map, in Ohm-Centimeters, for a Polycrystalline Germanium Blank



housing using a semicompliant bonding material. Silicon and some of the other crystalline materials also have this problem, so be very careful.

SILICON *Silicon* is also a crystalline material much like germanium. It is used primarily in the 3 to 5 MWIR spectral band as there is absorption in the 8- to 14- μm LWIR spectral band. While the refractive index of silicon is somewhat lower than germanium (silicon is 3.4255 versus germanium which is 4.0243), it is still large enough to be advantageous with respect to aberration control. Further, the dispersion of silicon is still relatively low. Silicon can be diamond turned with great difficulty.

ZINC SULFIDE *Zinc sulfide* is a common material used in both the MWIR and the LWIR. While its visible appearance varies greatly, it is generally rust-yellow in color and translucent in the visible. The most common process for manufacturing zinc sulfide is known as chemical vapor deposition (CVD).

If zinc sulfide is "HIP'ed" (hot isostatic pressed), it can be made to be water-clear in the visible. While available from several manufacturers, Cleartran is the most common commercially available clear zinc sulfide. Cleartran can be used for multispectral windows and lenses from the visible through the LWIR.

ZINC SELENIDE *Zinc selenide* is in many ways similar to zinc sulfide. It has a slightly higher refractive index than zinc sulfide and is structurally weaker. Because of this, a thin layer of zinc sulfide is sometimes deposited onto a thicker zinc selenide substrate for environmental durability reasons. Perhaps the most significant advantage of zinc selenide over zinc sulfide is that it has a significantly lower absorption coefficient than zinc sulfide. For this reason zinc selenide is commonly used in high-energy CO₂ laser systems.

AMTIR I AND AMTIR III *AMTIR I* and *AMTIR III* are glassy materials manufactured of germanium, arsenic, and selenium in a ratio of approximately 33:12:55. The AMTIR family of materials begins transmitting in the near IR (NIR). For this reason you can often see a very deep and faint red color transmitted through AMTIR. The *dn/dt* of AMTIR I is about 25% that of germanium, making it attractive from a thermal defocusing standpoint.

MAGNESIUM FLUORIDE *Magnesium fluoride* is another crystalline material. In its crystalline form it transmits from the UV through the MWIR spectral bands. Magnesium fluoride is manufactured by either crystal growth or alternately by “hot pressing.” In this latter process a fine powder form of the material is subjected to very high temperature and pressure in a way similar to powdered metal technology. The result is a milky looking glassy material which transmits well in the MWIR. The caution, however, is that there may be undesirable scattering which could cause contrast degradation and off-axis stray-light problems (this can be avoided using crystalline-grown material). Fortunately, small particle scattering is inversely proportional to the fourth power of the wavelength, so the milky appearance in the visible is reduced by approximately 2⁴ power, or 16 times at 5 μm.

SAPPHIRE *Sapphire* is an extremely hard material (it has a 2000 Knoop hardness value as compared to 7000 for diamond). It transmits from the deep UV through the MWIR. One of the unique aspects of sapphire is that it has a very low thermal emissivity at high temperature. What this means is that the bulk material when at a high temperature will emit less thermal radiation than other materials. You might, for example, use sapphire for the window of a chamber which is subject to very high temperature, especially if you are viewing through the window in the IR. Another application for sapphire is for protective windows of supersonic

vehicles where window heating is a serious problem. The primary disadvantage of sapphire is that its hardness makes it difficult, time consuming, and expensive to optically manufacture. There is a related material whose general class is known as spinel. Spinel is, in effect, analogous to hot-pressed sapphire and can be used in place of sapphire. Spinel is also highly dispersive.

Sapphire is birefringent which means that its refractive index is a function of the plane of incident polarization.

ARSENIC TRISULFIDE *Arsenic trisulfide* is another material sometimes used in the MWIR and LWIR. It is deep red in the visible and is quite expensive.

OTHER VIABLE MATERIALS There are a number of other viable materials including *calcium fluoride*, *barium fluoride*, *sodium fluoride*, *lithium fluoride*, *potassium bromide*, and others. These materials can be used from the deep UV through the MWIR spectral bands. Their dispersion properties make them quite attractive for wide-spectral-band applications, especially from the NIR through the MWIR and on to the LWIR. Many of these materials have some undesirable properties, especially water absorption (they are hygroscopic).

Generally, optical manufacturing methods for IR materials, such as germanium, silicon, zinc sulfide, and zinc selenide, are similar to glass optics. While manufacturers clearly have their trade secrets in this area, to the designer, they can all be considered as manufacturable. Several of the crystalline materials are hygroscopic, which presents some challenges to the optical shop. Also, these materials need to be appropriately coated to prevent damage from moisture and their housings often need to be dry nitrogen purged. IR materials generally have an extremely high index of refraction, which requires antireflection coatings; otherwise the system transmission would be very low.

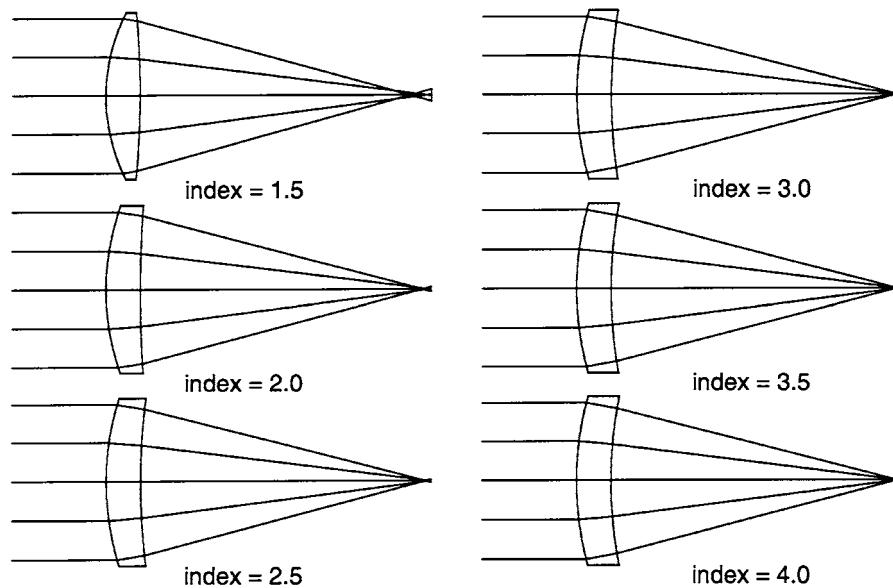
We have noted the hardness of sapphire previously. One further point of importance is that some IR materials can be single-point diamond turned. These include germanium, silicon, zinc sulfide, zinc selenide, AMTIR, and the fluorides. Sapphire cannot be diamond turned. Silicon can be diamond turned; however, the carbon in the silicon reacts with the carbon in the diamond and this results in a shorter tool lifetime and thus higher cost. Diamond turning can be extremely important if you need either aspheric surfaces and/or diffractive surfaces.

Reduced Aberrations with IR Materials

At the outset of this chapter, we indicated that the higher refractive indices of many of the infrared transmitting materials results in shallower and less steeply curved surfaces which, in turn, results in reduced aberrations. In order to illustrate this, consider Fig. 12.16, where we show six $f/2$ single-element lenses 25.4 mm in diameter, each bent for minimum spherical aberration. The refractive index of the lenses ranges from 1.5 to 4.0, where an index of 1.5 would be close to ordinary BK7 glass and 4.0 close to germanium. Note how the shape of the lenses is progressively changing. At index 1.5, the front is steeply convex and the rear is very slightly convex. At an index of approximately 1.62, the rear surface becomes flat. As the index keeps increasing, the lens becomes more concentric looking. While this is indeed very interesting, unfortunately it does not tell us anything about the aberrations.

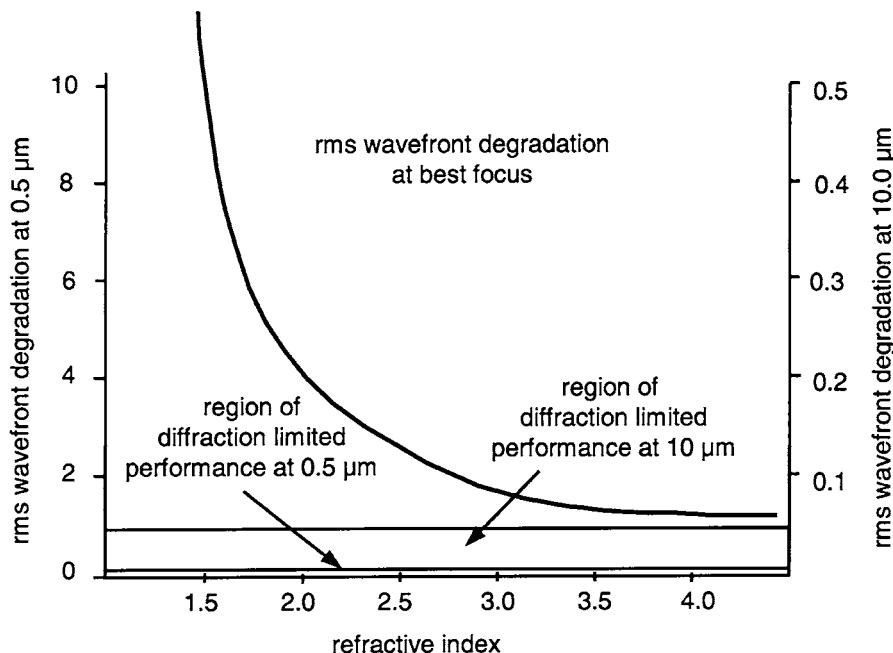
Figure 12.17 shows a plot of the rms wavefront error in waves at the wavelength of $0.5 \mu\text{m}$ versus refractive index for lenses bent for minimum spherical aberration. At index 1.5, we have over 10 waves rms

Figure 12.16
Lens Bending As a Function of Refractive Index for Minimum Spherical Aberration



Basics of Thermal Infrared Imaging

Figure 12.17
Spherical Aberration
As a Function of
Refractive Index for
 $f/2$ Lens 25 mm in
Diameter



which equates to approximately 50 waves peak to valley! This is an enormous amount of spherical aberration. Note how the aberration rapidly decreases with increasing refractive index. At an index of 2.0, which is about as high as we can find for visible glass, we have about 3 waves rms, or approximately 15 waves P-V. Note how at an index of 4.0, we have about 1 wave rms, or about 5 waves P-V. While this decrease in aberration is noteworthy, the most important point is that at an index of 4.0 we really must be thinking of the thermal infrared wavelengths, as glasses are not available with this refractive index. Let's therefore change the scale of the ordinate of the plot to indicate rms wavefront error at 10 μm . Since 10 μm is 20 times the visible wavelength of 0.5 μm , we need to reduce the values in the ordinate by 20 times. Thus, our 1-wave rms becomes 0.05 waves rms, which is approximately 0.25 wave P-V. This now meets the Rayleigh criteria and is diffraction limited! To summarize this extremely noteworthy finding: We have shown that a single $f/2$ element of glass 25.4 mm in diameter which is bent for minimum spherical aberration and has a refractive index of 1.5 similar to BK7 has approximately 50 waves P-V in the visible (200 times the diffraction limit). An equivalent single element of germanium with refractive index 4.0 when bent

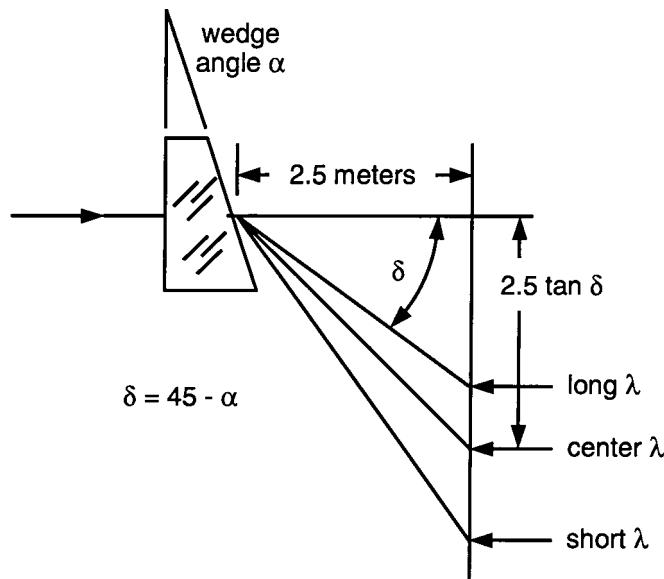
for minimum spherical aberration and referenced to a wavelength of $10 \mu\text{m}$ in the LWIR just meets the Rayleigh criteria with approximately 0.25 wave P-V!

It is for this reason that infrared optical designs are generally simple in form as compared with their visible counterparts. While a single germanium element may indeed suffice for an $f/2$ LWIR lens 25.4 mm in diameter, in the visible we would require three separate elements to achieve a diffraction-limited performance.

It was noted earlier that some infrared materials have very low dispersion, and for this reason color correction may not be required in some scenarios. In order to demonstrate this, we have come up with a virtual lab experiment right here in the book. Figure 12.18 shows the setup. We have a prism located 2.5 m from a vertical wall. We will now manufacture prisms similar to what is shown of various materials so that the center of the respective spectral band will be descending at a $(45-\alpha)^\circ$ angle, where α is the prism angle. We will then measure the length of the resulting spectrum for the visible, MWIR, and LWIR spectral bands.

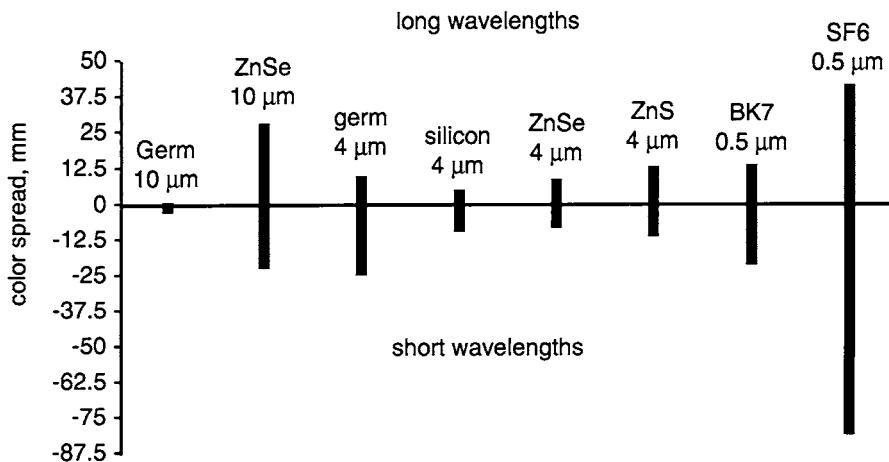
Figure 12.19 shows the results. Highly dispersive SF6 glass used in the visible from 0.4 to 0.7 μm has a large spectrum measuring about 120 mm. Less dispersive BK7 glass would have a spectrum measuring about 30 mm. In the 3- to 5- μm spectral band zinc sulfide would have a spectrum about

Figure 12.18
Hypothetical Experiment to Show Length of Spectrum for Different Materials



Basics of Thermal Infrared Imaging

Figure 12.19
Length of Spectrum
on Wall 2.5 m from
Prism



35 mm long, zinc selenide about 17 mm long and silicon about 12 mm. Germanium, as we know, is quite dispersive in the MWIR spectral band, and its spectrum would be about 35 mm. In the LWIR we find that the spectrum for zinc selenide is about 50 mm long and finally germanium in the LWIR is about 4 mm long. This is significantly less than any of the other materials, and this is why we can often use germanium alone in the LWIR spectral band without the need for color correction.

Image Anomalies

Thermal infrared systems often show cosmetically undesirable image anomalies which are not seen in visible optics. These effects include *narcissus*, *scan noise*, *beam wander*, *ghosting*, and *shading*. The effects are similar to what we generally think of as ghost images, and the resulting imagery can vary from slight brightness variations over the format to sharp bright or dark areas. While the mechanisms differ, all of these effects are due to the detector seeing more (or less) thermal energy over the field of view or through scan than dictated by the scene energy itself.

As we discussed in the section “Cold Stop Efficiency” earlier in this chapter, one of the most important methods for evaluating the properties of thermal infrared systems is to put your eye “figuratively” at the detector and look forward (into the exit pupil) and ask yourself “what do

you see." This is sometimes called the *detector's eye view*. For an IR system with 100% cold stop efficiency you should see a solid angle containing only scene energy, and everything outside this solid angle should be cryogenically cold. If this is the case, then you will indeed be accurately recording or imaging the thermal radiance from the scene. However, if you can see any thermal energy outside the solid angle representing scene energy and inside the cold stop solid angle, this represents extra energy which will behave in a similar fashion to stray light in visible systems. We will now review the primary causes of image anomalies in thermal imaging systems.

NARCISSUS *Narcissus* occurs because of a change in the magnitude of radiation reflected back into the dewar from lens surfaces within the system. Consider Fig. 12.20 where we show a scanning IR system at the center of scan (Fig. 12.20a) and at the end of scan (Fig. 12.20b). The focusing optics and detector are shown as rotated about the scan pivot point in Fig. 12.20b which is an optically valid representation of what is happening. We also show an enlarged view of the aft end of the system in Fig. 12.20c. We will be discussing only what happens from the rear surface of the last lens element. The total Narcissus effect is the radiometric summation from all lens surfaces.

If at the center of scan (Fig. 12.20a) we look out from the center FPA pixel, this ray will travel along the optical axis and reflect right back on itself ultimately returning through the cold stop into the dewar. Thus this ray "sees" only cold radiation from inside the dewar, or perhaps better said, no radiation at all. Now let's consider the ray from the center of the FPA which just passes by the edge of the cold stop. This ray, after reflection from the lens surface, diverges on its way back and misses entirely the cold stop aperture. In fact, this ray will strike some portion of the system interior, which could be ambient temperature or it could be a hot electronics board. The total solid angle, which can return into the cold stop is shown as the shaded solid angle in Fig. 12.20.

Now consider Fig. 12.20b, which represents the edge of scan. Here all of the energy within the entire solid angle within the cold stop reflects from the lens surface into the system interior. None of this energy will return into the cold stop and "see" cryogenically cold temperature. We thus can see how there is more "cold energy" (or lack of warm energy) at the center of scan than the edge of scan. If we now sum up the radiometric effect from each lens surface, and if we take into account the difference in solid angle returning into the dewar at the center and the

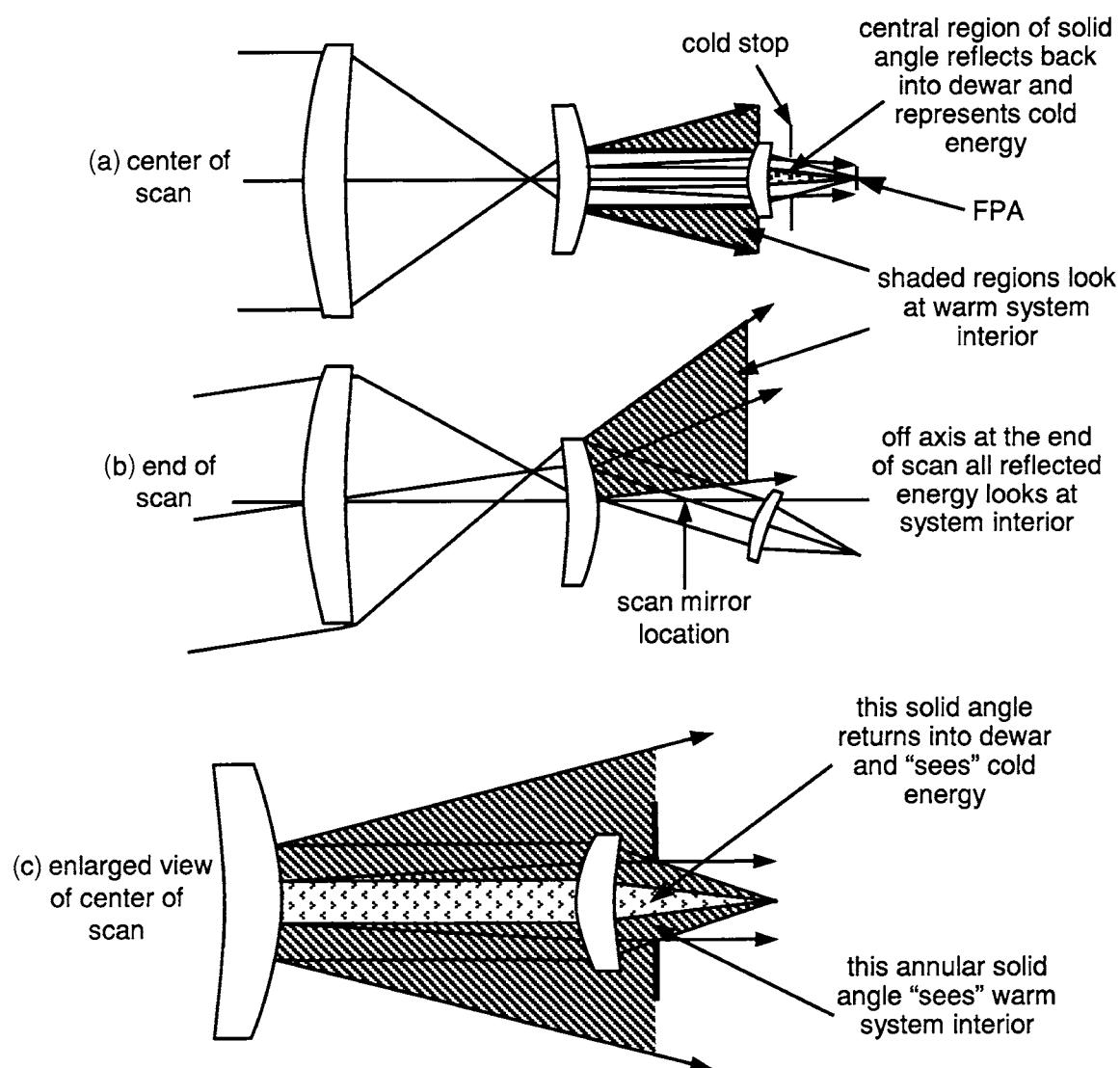


Figure 12.20

Illustration of How Narcissus Is Formed in a Scanning System

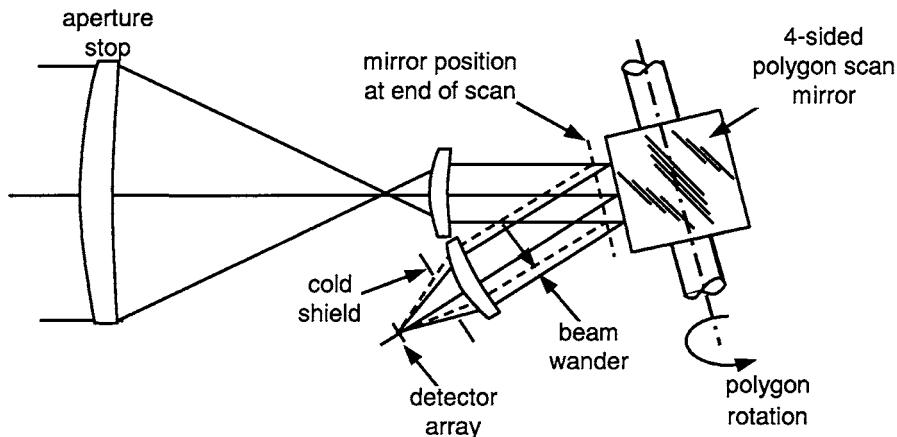
edge of scan, we will find that there is more “warmth” seen at the edge of scan than at the center of scan. This is, of course, due to the summation of the dark solid angles of energy returning into the dewar at the center of scan from each lens surface. The bottom line is that the video monitor will show less thermal energy at the center of scan than the edge of scan, and this may appear as a dark porthole or disk on the monitor. The diameter of the dark disk and the gradient from the center of the monitor to the edge will depend on the specific geometry and how rapidly the cold energy solid angle reduces with scan.

Note that in nonscanning staring array systems using two-dimensional FPAs you can still have Narcissus. There is one important difference in this regard between scanning and nonscanning or staring systems—the practice in staring systems of performing a “nonuniformity correction.” What this means is that the system is periodically aimed toward a uniform thermal source, and then each pixel is adjusted in its offset to yield a constant gray level over the entire format. For many thermal imaging applications where we are looking at near-ambient temperatures, simply draping a black cloth over the front of the system and then performing the nonuniformity correction will suffice. However, if the temperature of the system interior where the reflected radiation strikes at the edges of scan changes, there may be a need to perform a new nonuniformity correction.

How can we minimize the effect of Narcissus? There are two basic methods to minimize Narcissus. First we can use so-called anti-Narcissus coatings. These antireflection coatings typically have from 0.2 to 0.3% average reflectivity from 3 to 5 μm or alternatively 8 to 14 μm . The second approach is to change the relative bendings of the offending lens surfaces so as to minimize the cold solid angle of reflected radiation. This is a common technique, and to keep the desired level of optical performance, often requires the use of aspheric surfaces. Fortunately, this is not a major problem with the use of single-point diamond turning, now a mature technology. Optical systems often have a flat protective window in front of the telescope. In order to avoid the Narcissus generated by the window, the window is tilted, so that the reflected radiation falls out of the sensor’s field of view.

GEOMETRICAL SCAN EFFECTS *Geometrical scan effect (beam wander)* is illustrated in Fig. 12.21, where we show a generic four-sided polygon scanning IR system. We will assume that the system aperture stop is on the front lens element, a likely situation. At the center of scan, the

Figure 12.21
Geometrical Scan
Effects and Beam
Wander



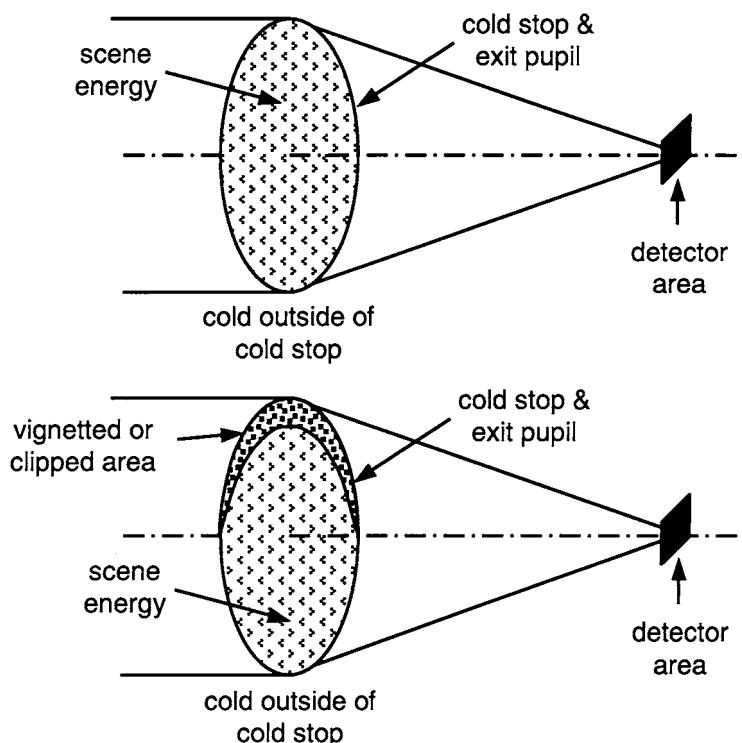
Polygon will be at its nominal or neutral position as shown. As the polygon rotates to the end of scan, its geometrical position in space, where the radiation reflects from its surface, translates to the dashed position (to the left). This represents the location where the intersection of the polygon facet with the plane of the figure has translated for end-of-scan imaging. Since we have stated previously that the aperture stop is on the front element, the beam of radiation reflected from the facet must translate or shift laterally by an amount called the *beam wander*.

We now have a dilemma... should the cold stop be sized for the center of scan and thus clip energy from the beam-wandered edge of scan radiation or should we increase the diameter of the cold shield (along with a lateral translation of the cold shield)? The "lesser of the two evils" is generally to increase and laterally shift the cold shield so as to eliminate clipping of the imaging radiation from any scan position.

SCAN NOISE *Scan noise* is often taken to be any undesirable change in nonscene thermal energy reaching the FPA through scan, thus causing anomalies such as bright bands and other image defects. The appearance of scan noise can often resemble flare and stray light in visible systems. One of the more common causes of scan noise is clipping or vignetting which can result if a portion of the housing infringes on the radiation bundle at the ends of scan.

Consider Fig. 12.22 where we show a representation of the solid angle of radiation reaching the detector at the center of scan (top) and at the edge of scan (bottom). At the center of scan, the solid angle is totally

Figure 12.22
Effect of Clipping or
Vignetting



scene energy within the cold stop and cryogenically cold outside the cold stop. Thus, the only energy the detector will "see" or record is from the scene. At the edge of scan, the solid angle within the cold stop comes primarily from the scene; however, there is a crescent-moon-shaped shaded area which is caused by clipping or vignetting from a housing interior. The solid angle outside the cold stop is cryogenically cold as for the center of scan case. Thus the only difference from the center to the edge of scan is the clipping from the housing interior.

It can be shown that the perceived change in temperature is approximately given by

$$dt \sim \frac{\Delta A}{A} (\text{temperature clipped} - \text{temperature scene})$$

The quantity $\Delta A/A$, is the percent of area of the exit pupil within the cold stop which is clipped. Let us take some representative values for the preceding quantities. Assume that $\Delta A/A = 1\% = 0.01$, the temperature

of the clipped housing is 50°C , and the scene temperature is 0°C . This gives us a perceived temperature difference from the center of scan to the edge of scan of $0.01 \times 50^{\circ}\text{C} = 0.5^{\circ}\text{C}$. If we are scanning in the azimuth direction with a polygon or alternatively with an oscillating mirror, the net result will likely be bright bands at the left and right of the monitor, which get brighter toward the corners of the field of view.

With today's detector technology, temperature differences far below 0.1°C can easily be seen. It is thus clear that even 0.1% clipping can often be seen and may be a problem. In fact, a piece of dirt on a lens element measuring only $0.4 \text{ mm} \times 0.4 \text{ mm} = 0.16 \text{ mm}^2$ is equivalent to 0.1% clipping for a 25.4-mm-diameter ray bundle.

GHOSTING *Ghosting* is a term applied to an effect unique to polygon scanners. Consider Fig. 12.11 where we showed radiation incident onto a polygon scanner facet. If the polygon is in its neutral position with the facet tilted at 45° to the incoming radiation, the radiation simply reflects downward toward the FPA. Recall that to best evaluate IR systems relative to image anomalies, you should put your eye at the detector and look out and ask yourself "what do you see?" Let's do this for the facet in the central position. All we see is scene energy, as expected. However, now rotate the facet around in the counterclockwise direction until the next adjacent facet has just entered the beam of radiation. If we look out from the FPA, we will see energy reflected from the prime facet heading down to the left and a small sliver of energy from the adjacent facet heading up to the right. This sliver of radiation will be at a greater angle numerically than from the prime facet, and thus it may miss the lenses and hit an interior portion of the housing. Alternatively, this energy may make it out into object space in which case it is viewing a portion of the scene at the opposite side of the field and outside the nominal field of view. When it is hitting a housing interior, this is known as "internal ghosting" and when it views object space it is known as "external ghosting." Thus, a person holding a bright flare just outside the field of view on the right side may appear inside the field of view on the left due to external ghosting.

What is required to best understand effects such as ghosting is to perform an accurate three-dimensional modeling of your system. Every aspect of the system must be properly modeled, including the polygon axis of rotation, facet clear aperture, and the clear aperture of other lenses within the system. Only then will there be some hope of accurately predicting what is happening.

SHADING *Shading* is very different from any of the previous phenomena, and it is more difficult to explain. Consider the exit pupil diameter at the center of scan reaching the focusing optics. If we have a 100-mm-diameter entrance pupil and a magnification of 20 \times , then the radiation bundle diameter at the exit pupil is 5 mm. Now as we proceed off axis in field of view and scan, what happens to our 5-mm beam diameter? For example, let's assume that we have positive distortion off axis, which means that the magnification increases with field of view. Thus, the 5-mm exit pupil will be smaller in scan mirror space and as it enters the focusing optics. If the aperture stop is at the front element, this will result in some area of the housing being seen at off-axis field and scan positions. The net result is a brightening of the display monitor away from the center of the field of view due to less scene energy and more housing energy.

Athermalization

As any optical system is subjected to higher or lower temperatures several things happen: the lens elements expand or contract, the housing expands or contracts, and the refractive index of the lens materials increases or decreases. While the lens and housing relative expansion can sometimes be a problem, the primary problem with infrared systems is the very large change in refractive index as a function of temperature, or dn/dt . Germanium has a dn/dt of 0.000396/ $^{\circ}\text{C}$. For comparison, BK7 glass has a dn/dt of 3.6 0E-6. It can be shown for a simple lens that

$$df = \text{change in focal length} = \frac{f}{n - 1} \frac{dn}{dt} \Delta t$$

where df is the change in focal length and Δt is the change in temperature in degrees Celsius.

Consider the following example: Assume that we have a 75-mm-diameter f/1.5 germanium lens with a focal length of 112.5 mm. Applying the previous relationship yields a change in focal length of 0.599 mm for a 40 $^{\circ}\text{C}$ Δt thermal soak. For reference, the Rayleigh criteria for one-quarter wave of defocus is ± 0.046 mm, so the preceding defocus value equates to 3.3 waves of defocus, which is 13.1 Rayleigh criteria depths of focus, a huge amount! This issue can, and often is, a very serious problem in thermal infrared systems. In this example, if we were to control the

Basics of Thermal Infrared Imaging

temperature so as to stay within a one-quarter wave of defocus, we would need to control the temperature to within $\pm 3^\circ\text{C}$.

It is thus apparent that unless the user can actively refocus the lens, athermalization is imperative. We could translate the FPA or the prime collecting lens; however, this is generally impractical. We could translate another lens element by a greater or lesser amount depending on the magnification, and this is often the approach. Two other approaches are also viable: negative elements with a high dn/dt can be used and use of reflective optics.

To show how effective reflective optics can be in achieving athermalization, consider Fig. 12.23. We show here two very similar systems, a fully refractive system above the optical axis and a system below the axis where a Cassegrain reflective system takes on the role of the two elements following the curved window of the refractive system. The system has a 75-mm entrance pupil diameter and is used in the IWIR at 10 μm central wavelength. Table 12.2 shows the defocus contributions for the refractive and its reflective counterpart for a 50°C thermal soak (uniform temperature change). First, look at the refractive system as in Fig. 12.23. The front

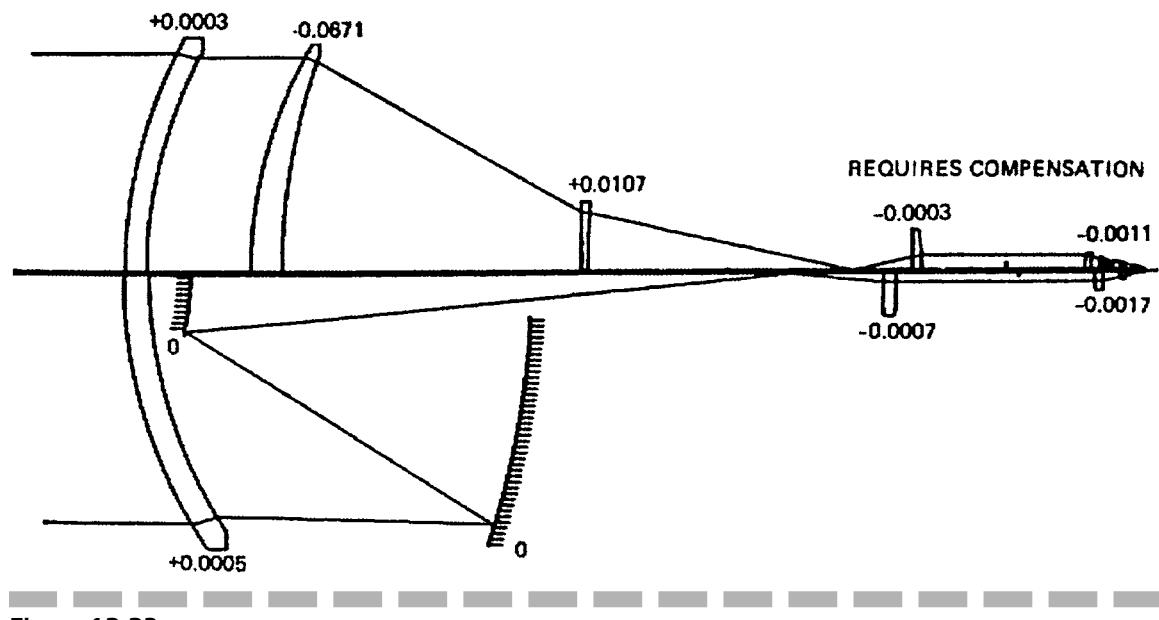


Figure 12.23

Thermal Sensitivities for Refractive System (Top) and Reflective System (Bottom)

TABLE 12.2

Thermally Induced
Focus Shifts for
Refractive and
Reflective IR Optical
Systems

Parameter	Refractive System (mm)	Reflective System (mm)
Curved dome	+0.0076	+0.0127
Large lens or mirror	-1.704	0.0
Next lens or mirror	+0.272	0.0
First small lens (collimating)	+0.0076	-0.018
Second small lens (focusing)	-0.028	-0.028
Entire system	-1.712	-0.048

curved dome has very little optical power (the power is slightly negative) and its contribution is to defocus the image outward 7.6 μm . The large first element accounts for a 1.7-mm inward image defocus, and the negative next element moves the image outward by 0.27 mm. The last two elements cause only a small defocus. The net system defocus due to the 50°C thermal soak is about the same as the large powered element at 1.71 mm inward, which is the same as the first element alone.

Looking now at the reflective Cassegrain system, we see that the two mirrors cause no thermal refocusing at all, and the total system defocus is 0.048 mm inward which just happens to be extremely close to the quarter wave Rayleigh criteria.

These data are summarized in Table 12.2.

It is important to understand just why the thermal defocus of the reflective Cassegrain is zero. In fact, there is a subtlety here—each mirror does cause a significant defocus with temperature because it changes shape with temperature. However, the total reflective optics system (the Cassegrain) has virtually zero thermal effect. Perhaps the best way to convince yourself that this is true is to consider a Cassegrain system with only a primary and a secondary mirror along with its support structure, and draw the rays coming from infinity, converging from the primary to the secondary, and finally focusing onto the image plane. Assume that the mirrors and all of the support structure supporting the mirrors and the FPA are all made of the same material such as aluminum. Under a positive thermal soak condition, the entire system will expand uniformly, and the system will therefore stay in perfect focus. To be fully convinced of this conclusion, simply think about taking the drawing you have just made and enlarge it on your copy machine; then look at the enlarged drawing and notice that it is still in focus! So when

we assign a zero sensitivity to the mirrors, we are assuming that this is in the context of a total reflective system of the same material.

We have shown in the preceding example how a nearly all-reflective system is athermalized to within the Rayleigh criteria. How do we restore focus in the refractive system? We could axially translate the detector or alternately translate the front element aft of the curved window. Both of these approaches are unattractive. Moving the detector/dewar assembly is just not a good idea with all of the electronic as well as cryogenic connections, and translating the massive first element is also a difficult task. The best approach is to translate another element within the system by an appropriate amount, which will be dependent on its magnification within the system. In the system shown, the best candidate is to move axially the negative element aft of the large front lens element. To first order, we will need to translate this element about the same amount as the image shifts from the large front element, and this is about ± 1 mm rough order of magnitude. Thus, as the system heats up by 50°C , the image from the large front element moves forward by about 1 mm, which means that the negative element must also move forward by approximately 1 mm in order to restore focus on the final image sensor. The specific magnitudes of motion need to be fully verified using your computer model. If this were the method to be employed, you could locate two to three thermistors or other temperature sensors at the outer periphery of the first element and drive the compensating element (element 2) a distance stored in a look-up table based on the temperature. The temperature sensors would be at this position in the system since the large first element is by far the most sensitive to thermal defocus, as shown earlier. Alternately, an active system approach similar to that used in 35-mm and other camera systems can be used, whereby the algorithm is based on maximizing scene contrast or some other similar criteria is used.

One of the most important points to be made here is that any element motions for athermalization reasons must be sufficiently accurate to locate the image to within approximately one Rayleigh criteria depth of focus from the nominal. In the examination of which element or elements to move, you must examine the total range of motion required. Within this range, you need to determine the finest focus that needs to be made. Make sure that this fine focus adjustment is achievable with the envisioned motion mechanism; otherwise, there will be conditions where a sharp focus cannot be achieved.

System Design Examples

To illustrate how lens designs for the thermal infrared look and work, we show in Figs. 12.24 through 12.31 designs for an $f/2$ LWIR lens 25 mm in diameter covering a field of view of $\pm 2.5^\circ$. The materials include germanium, AMTIR 1, zinc sulfide, and zinc selenide. We show for each design a layout, the transverse ray aberration curves on axis and at 2.5° off axis. Also tabulated is the design prescription data along with the rms wavefront error, which is the average of the axis, 1.25° off axis and 2.5° off axis. With these data you can set up these designs and work with them as you wish. Do note that there is no consideration given to cold stop efficiency in these examples.

The resulting performance is shown in Table 12.3.

In Fig. 12.32 we show an $f/3$ lens with 100% cold stop efficiency along with its design data. Note that this design form is different from that shown in Fig. 12.4 where we achieved 100% cold stop efficiency by reimaging the first element, which also was the aperture stop, into the cold stop. Here we are not using any reimaging at all, and in order to make the cold stop in the dewar the aperture stop it was simply set up that way. Notice the extreme beam motion over the front elements and

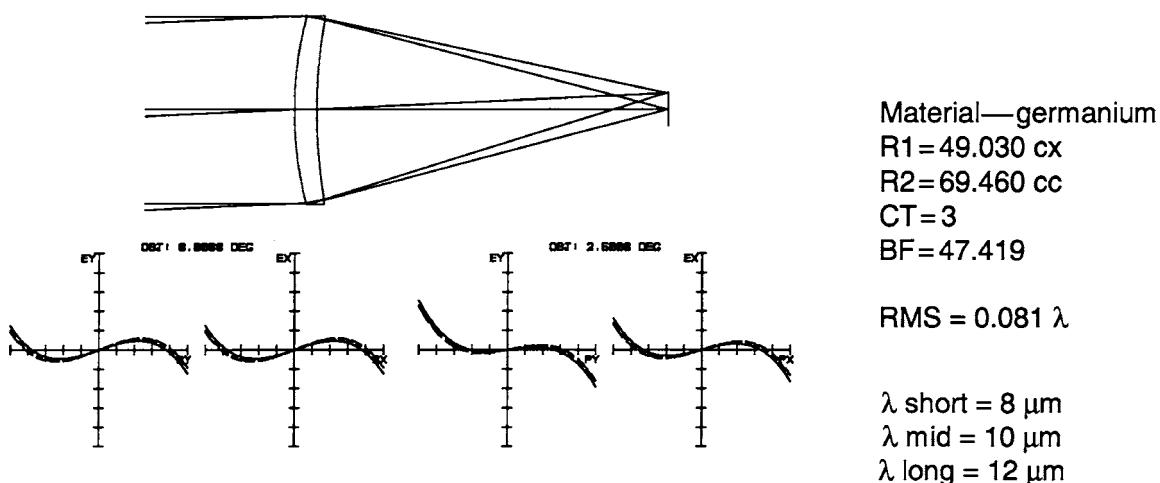


Figure 12.24
Single-Element Germanium Lens, $f/2$ 50-mm Focal Length

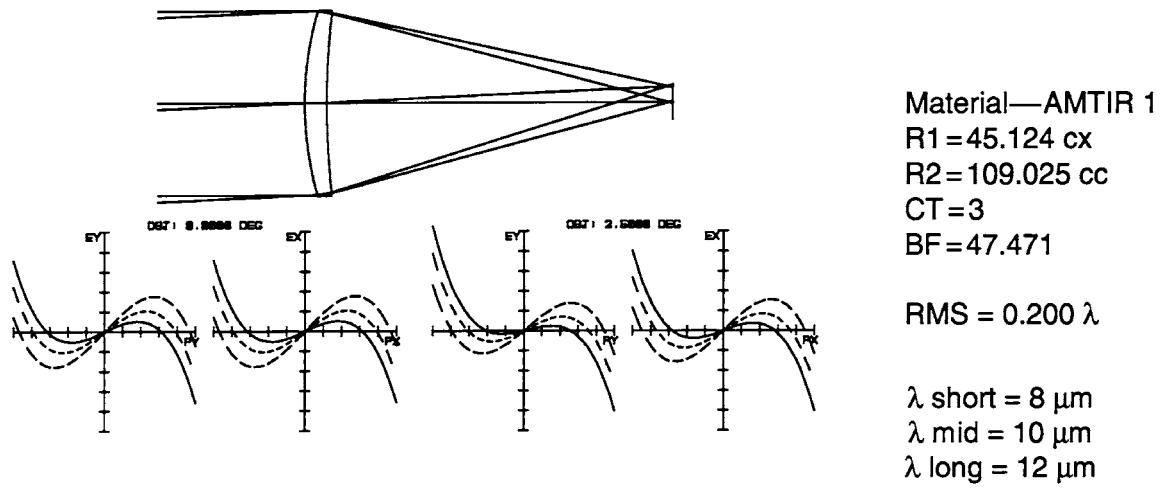


Figure 12.25

Single-Element AMTIR 1 Lens, f/2 50-mm Focal Length

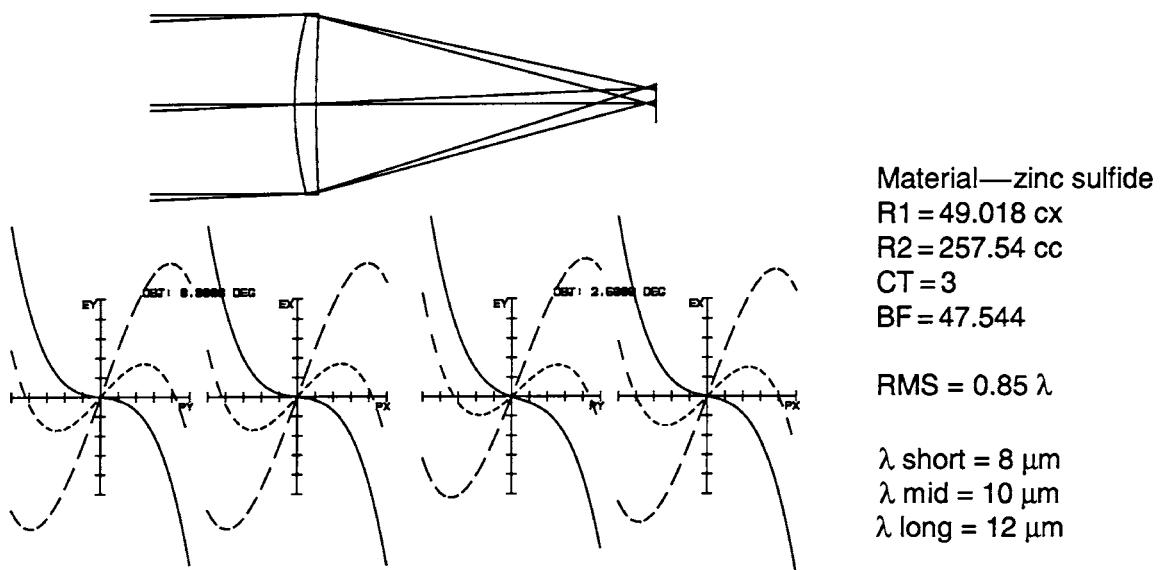


Figure 12.26

Single-Element Zinc Sulfide Lens, f/2 50-mm Focal Length

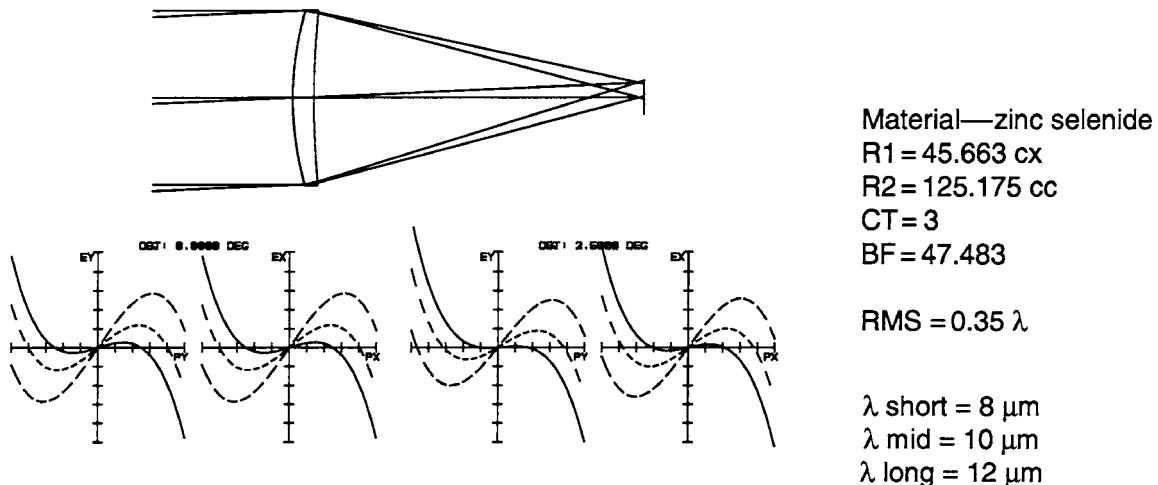


Figure 12.27
Single-Element Zinc Selenide Lens, $f/2$ 50-mm Focal Length

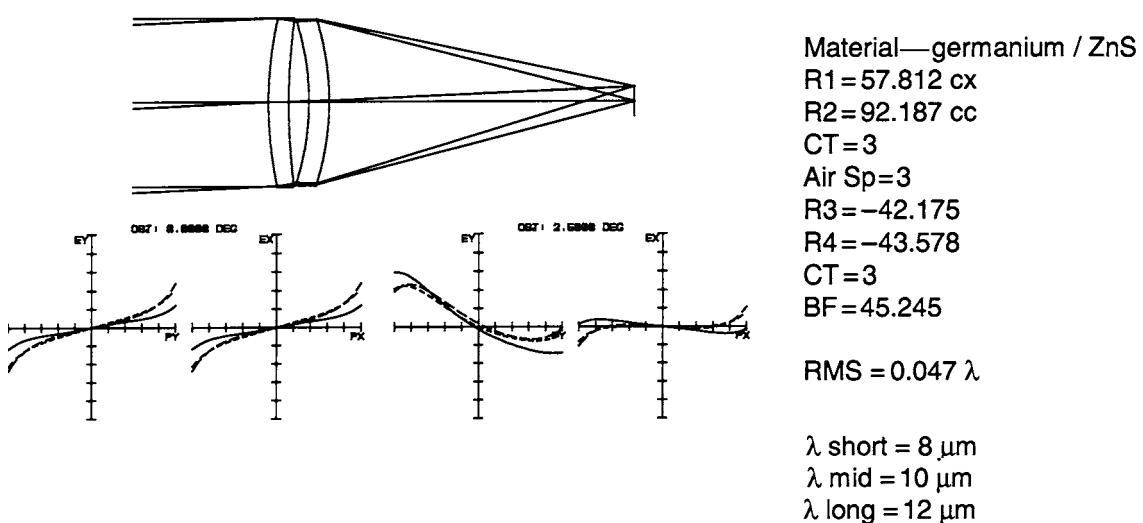


Figure 12.28
Germanium/Zinc Sulfide Lens, $f/2$ 50-mm Focal Length

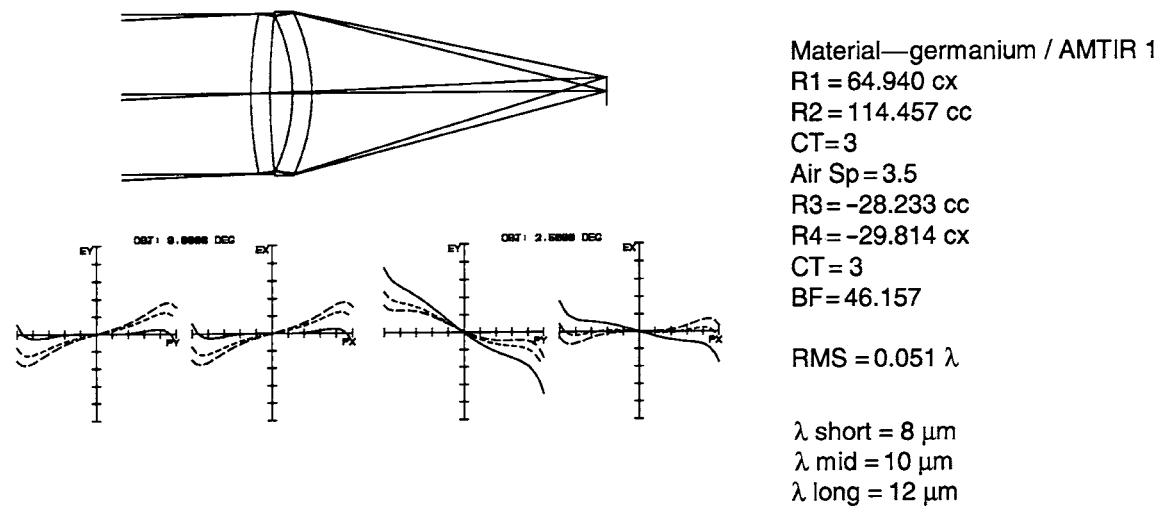


Figure 12.29

Germanium/AMTIR 1 Lens, f/2 50-mm Focal Length

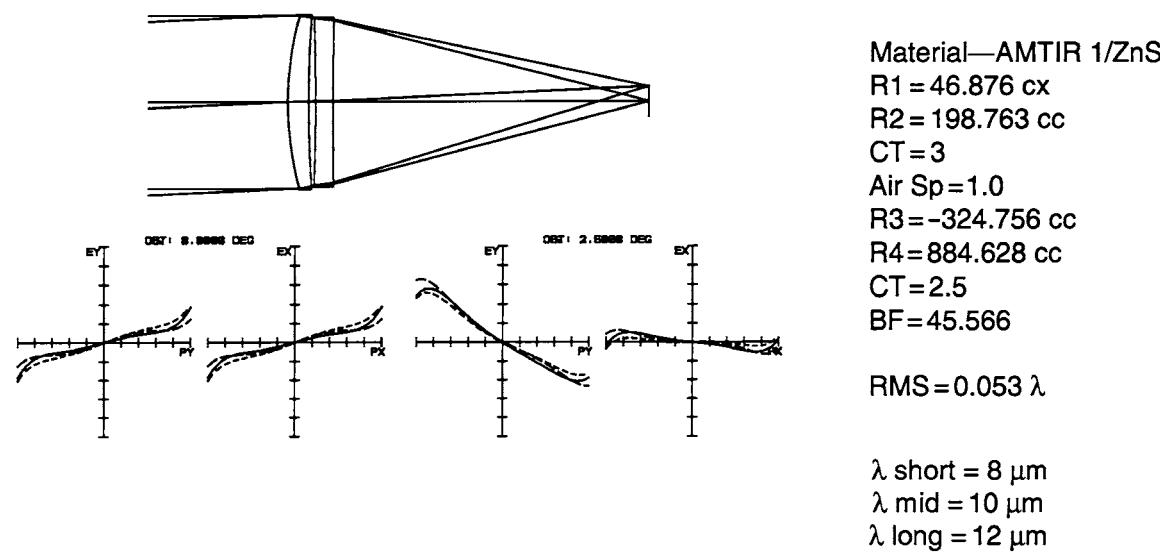
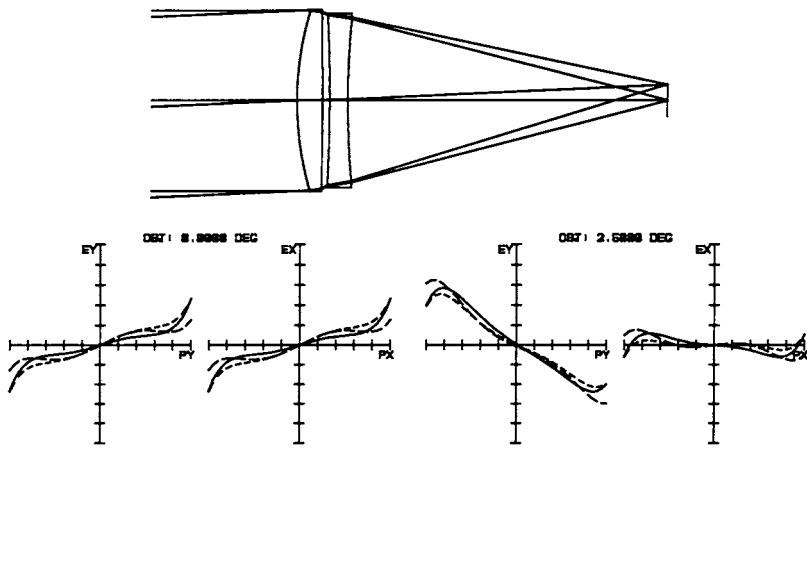


Figure 12.30

AMTIR 1/Zinc Sulfide Lens, f/2 50-mm Focal Length



Material—ZnSe/ZnS
 $R_1 = 43.667$ cx
 $R_2 = -1267.634$ cx
 $CT = 3.5$
 $Air Sp = 1.0$
 $R_3 = -231.778$ cc
 $R_4 = 125.291$ cc
 $CT = 2.5$
 $BF = 44.285$
 $RMS = 0.057 \lambda$

λ short = 8 μm
 λ mid = 10 μm
 λ long = 12 μm

Figure 12.31
 Zinc Selenide/Zinc Sulfide Lens, $f/2$ 50-mm Focal Length

the large size of these elements with respect to the entrance pupil diameter. When you can use this nonreimaging form of design, you should do so as the design is simpler and generally performs well; however, in some situations you must use a reimaging configuration. This might be the case, for example, if you have a very tight packaging requirement.

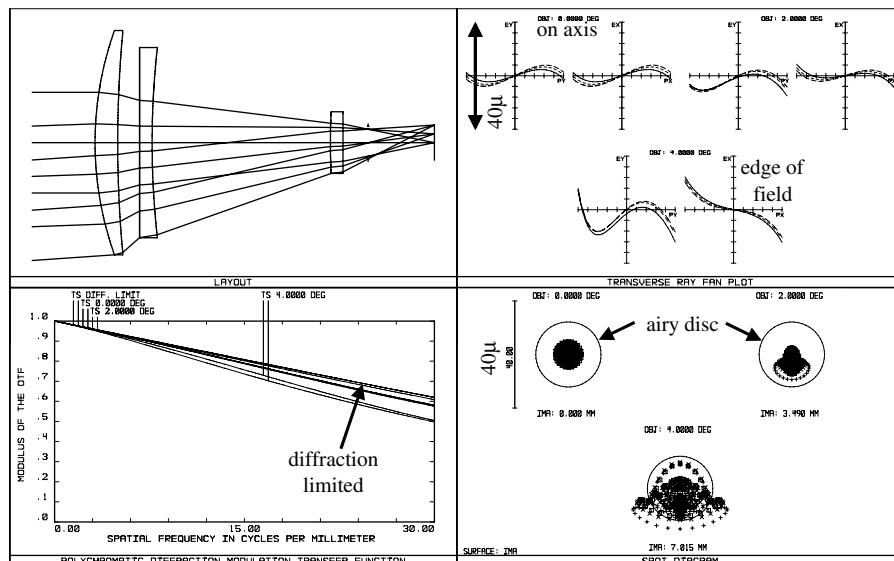
TABLE 12.3

Relative Performance of LWIR Design Examples

Lens Construction	RMS Wavefront Error	Diffraction Limited
Germanium singlet	0.08	Nearly
AMTIR 1 singlet	0.20	No
Zinc sulfide singlet	0.85	No
Zinc selenide singlet	0.35	No
Germanium/zinc sulfide doublet	0.047	Yes
Germanium/AMTIR 1 doublet	0.051	Yes
AMTIR 1/zinc sulfide doublet	0.053	Yes
Zinc selenide/zinc sulfide	0.057	Yes

Basics of Thermal Infrared Imaging

Figure 12.32
100-mm Focal
Length MWIR Lens
with 100% Cold Stop
Efficiency



Surface	Radius	Thickness	Material
1	132.661 cx	9	silicon
2	480.403 cc	8.671	
3	Infinity	5	germanium
4	340.888 cc	70.846	
5	242.243 cx	5	silicon
6	Infinity	10	
7 stop	Infinity	26.3	
8	Image, infinity		

Optical Systems for the UV

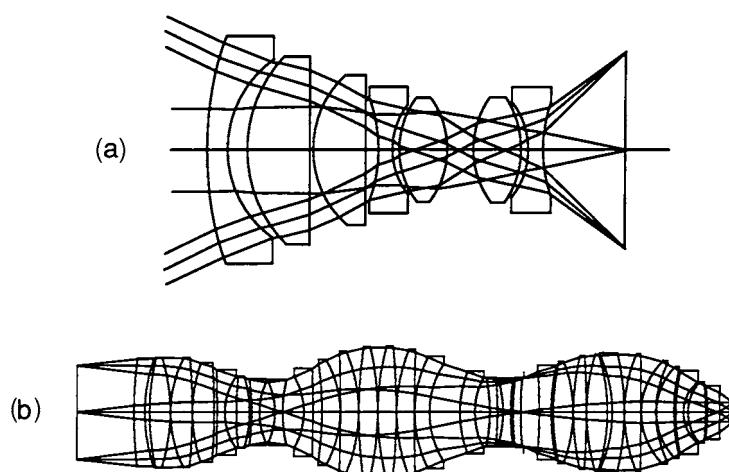
Working with systems in the UV is extremely challenging and demanding. In the thermal infrared, specifically the MWIR and the LWIR spectral bands, we found that the wavelength was 8 and 20 times the visible wavelength, respectively, which in some ways made IR lens systems more forgiving. In the UV we find that we now have a wavelength about one-half that of a visible system. In addition to the Airy disk expressed in micrometers being about one-half the *f*/number, we also

find a limited number of viable optical materials are available. These include fused silica, several of the fluorides (barium fluoride, calcium fluoride, and lithium fluoride), UBK7 glass, and sapphire. The index of refraction of these materials is generally not very high. Many of these materials (especially the fluorides) are very difficult to work with and have other problems such as being hygroscopic. This leads to extreme care in manufacturing and assembly, and you may need to nitrogen purge your system to prevent moisture damage. Even sodium chloride can be used in the UV, but we recommend that you “take it with a grain of salt.”

Figure 12.33 shows two deep UV lens systems, the first of which is a relatively wide-angle lens using calcium fluoride and Ultran 30 materials (the latter is no longer available). Note the relatively steep radii which is due to the inherently lower refractive indices of the materials. The second lens shown is a wafer stepper lens from a patent. This lens is similar to the Glatzel lens in Fig. 5.15, except here many more lens elements are used. Having searched the patent files extensively, this indeed is one of the more complex multielement lenses we found.

Mirror systems offer a unique advantage in the UV. Since there are no chromatic aberrations with mirrors, these aberrations are by definition zero. And since the dispersion of refractive materials is larger at lower wavelengths, using mirrors makes good sense when you can. Two very clever reflective systems are the Schwarzschild reflective microscope

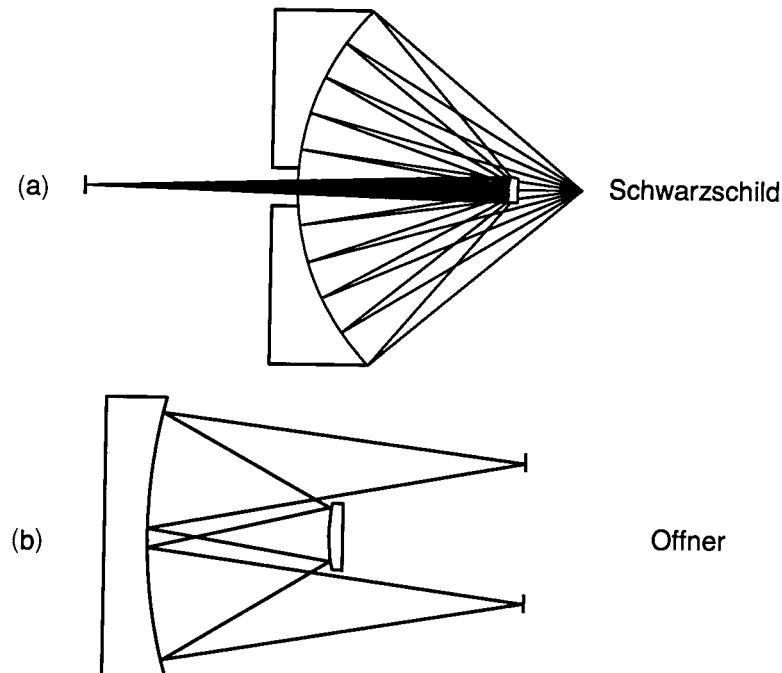
Figure 12.33
Refractive Designs for
the Deep Ultraviolet



objective and the Offner $1\times$ relay for lithography, both of which are shown in Fig. 12.34. The Offner design is especially clever in that it produces a ring field of view where virtually all orders of aberrations are corrected to zero. The original patent for this system was granted in 1973 (USP 3,748,015), and the abstract reads as follows:

A catoptric [all reflective] system for forming in accurate micro detail an image of an object at unit magnification with high resolution is provided by convex and concave spherical mirrors arranged with their centers of curvature coinciding at a single point. The mirrors are arranged to produce at least three reflections within the system and they are used in the system with their axial conjugates at said point and to provide two off axis conjugate areas at unit magnification in a plane which contains the center of curvature, the axis of the system being an axis normal to the latter plane and through said point. This combination is free from spherical aberration, coma, and distortion and, when the algebraic sum of the powers of the mirror reflecting surfaces utilized is zero, the image produced is free from third order astigmatism and field curvature.

Figure 12.34
Reflective Designs for
the Deep Ultraviolet



Note that UV systems are often prone to scattering problems. The total integrated scatter (TIS) is

$$\text{TIS} \approx \left[\frac{4 \pi \delta (\cos \theta)}{\lambda} \right]^2$$

where δ is the rms surface roughness, λ is the wavelength, and θ is the angle of incidence. This means that surface imperfections and finish, as well as bulk material scattering, can introduce unwanted stray light.

CHAPTER

13

Diffractive Optics

Introduction

What Is Diffraction?

A chapter on diffractive optics within a classical optical design book is the perfect way to introduce the dual nature of light.¹ Light can be studied either through its corpuscular nature, the photon, (the basis of ray tracing and classical optical design of lenses and mirrors) or through its wave nature, an electromagnetic wave (the basis of physical optics used to model diffractive optics and other micro- or nano-optical elements, like integrated waveguides, and even photonic crystals).

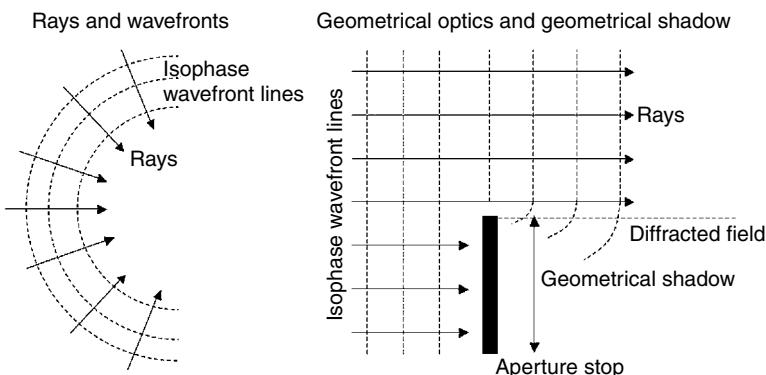
In the simple knife edge example presented in Fig. 13.1, the corpuscular nature (ray tracing) accounts for geometrical optics whereas the wave nature (physical optics) nature accounts not only for the light in the optical path, but also for the light appearing inside the geometrical shadow—through diffraction at the edges of the knife edge—where no light should appear according to ray tracing.

From Refraction to Diffraction

Many diffractive elements have their counterparts in the classical realm of optical elements. However, the similarity is only superficial since their behavior under various operating configurations may be very different.

Figure 13.1

The Dual Nature of Light



As we will see later in this chapter, in many cases, diffractives are best used in conjunction with refractive or reflective optics in order to provide new and/or extended optical functionality (such as hybrid achromat or athermal singlets).

The Blazed Grating

Consider a linear blazed diffraction grating and its prism refractive counterpart, which would bend an incoming light beam into a specific—and same—direction, for a specific wavelength and a specific launch angle. The diffractive and refractive behaviors differ rapidly when one drifts away from these exact configurations. The same is true for a diffractive Fresnel lens and a refractive lens, where the dispersion occurs with opposite signs (the sign of their respective Abbe V numbers) although their focusing power and phase profiles might be exactly the same. Figure 13.2 shows both Snell's equation (ray tracing) and the grating equation (physical optics) which accounts for the amount of light bending, for a small prism and a linear blazed grating.

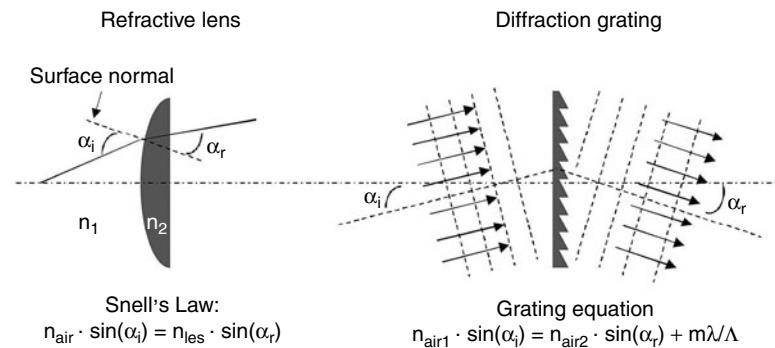
As one looks closer at a blazed grating, one can consider the various periods of this grating as many individual refractive microprisms, and therefore apply not only the grating equation to the entire blazed grating (array of microprisms), but also Snell's law of refraction to each individual microprism, as depicted in Fig. 13.3.

It is interesting to note that the bending angle predicted by refraction through the local microprism structures and the bending angle predicted by diffraction through the blazed grating are not necessarily the same. In effect, they are equal only in one very specific case: when the geometry of

Diffractive Optics

Figure 13.2

Snell's Law of Refraction and the Grating Equation Ruling the Bending of Light



the microprism is carefully chosen (height, length, and refractive index) as shown in Fig. 13.4. Maximum diffraction efficiency is then reached for the blazed grating (which can reach 100% efficiency when both previously described effects are yielding the same angle bending).

Snell's law predicts the amount of refraction at a given optical interface, and gives thus also the expression of the refracted beam angle through the microprism (Eq. 13.1).

$$n_1 \sin(\alpha_i) = n_2 \sin(\alpha_r) \Leftrightarrow \sin(\alpha + \gamma) = n_1 \sin(\alpha) \quad (13.1)$$

Physical optics, or the grating equation, gives the expression of the diffracted angle through the entire array of microprisms (Eq. 13.2).

$$\sin(\beta) = m \frac{\lambda}{\Lambda} \quad (13.2)$$

Figure 13.3
The Blazed Grating
and its Microprism
Array Structure

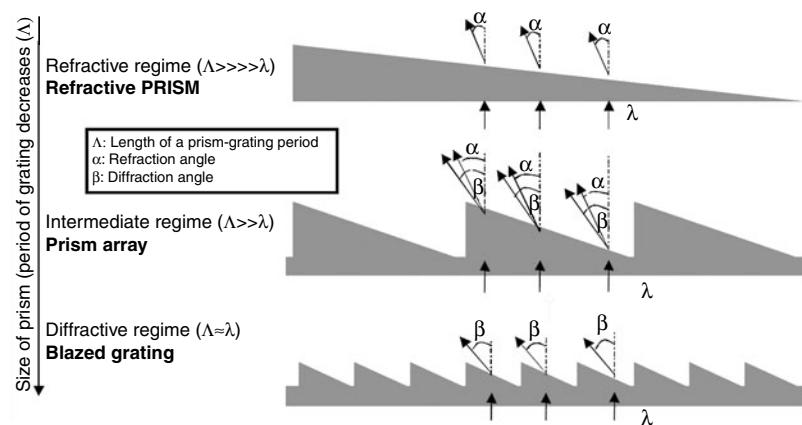
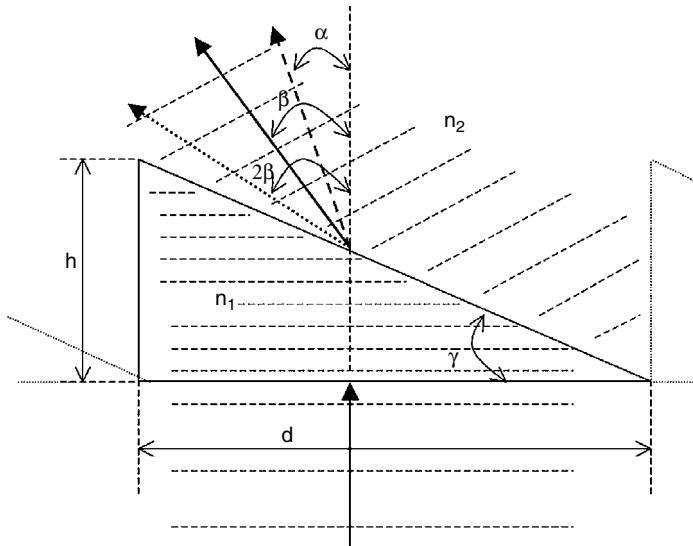


Figure 13.4
The Local Micropism Effect and the Global Grating Effect



Intuitively, the maximum efficiency will thus occur when $\alpha = \beta$ (Eq. 13.3)

$$\alpha = \beta \Rightarrow h = \frac{\lambda}{n_1 - \sqrt{1 - \frac{\lambda}{\Lambda}}} \Rightarrow h = \frac{\lambda}{n_1 - 1} \quad (13.3)$$

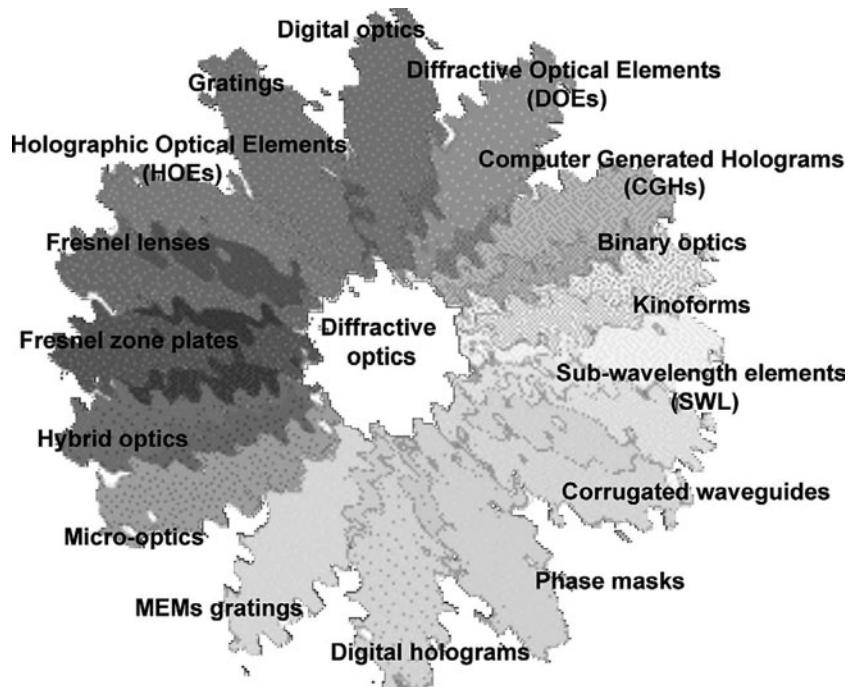
Therefore, by using the same concepts to increase the light bending efficiency, and by carefully shaping the overall grating geometry (grating period, groove height, groove angles, and refractive index), one can engineer any type of diffractive grating or diffractive lens for specific applications (aspheric lenses, circular gratings, etc. listed in the following section).

The Many Faces of Diffractive Optics

Marketing and sales managers, venture capitalists, engineers, academics as well as technical writers have given numerous names to diffractive optics in the last three decades. Some most commonly used names are binary optics or digital optics, diffractive optical elements (DOEs), computer generated holograms (CGHs), holographic optical elements (HOEs) and kinoforms. Figure 13.5 shows a compilation of those various names.

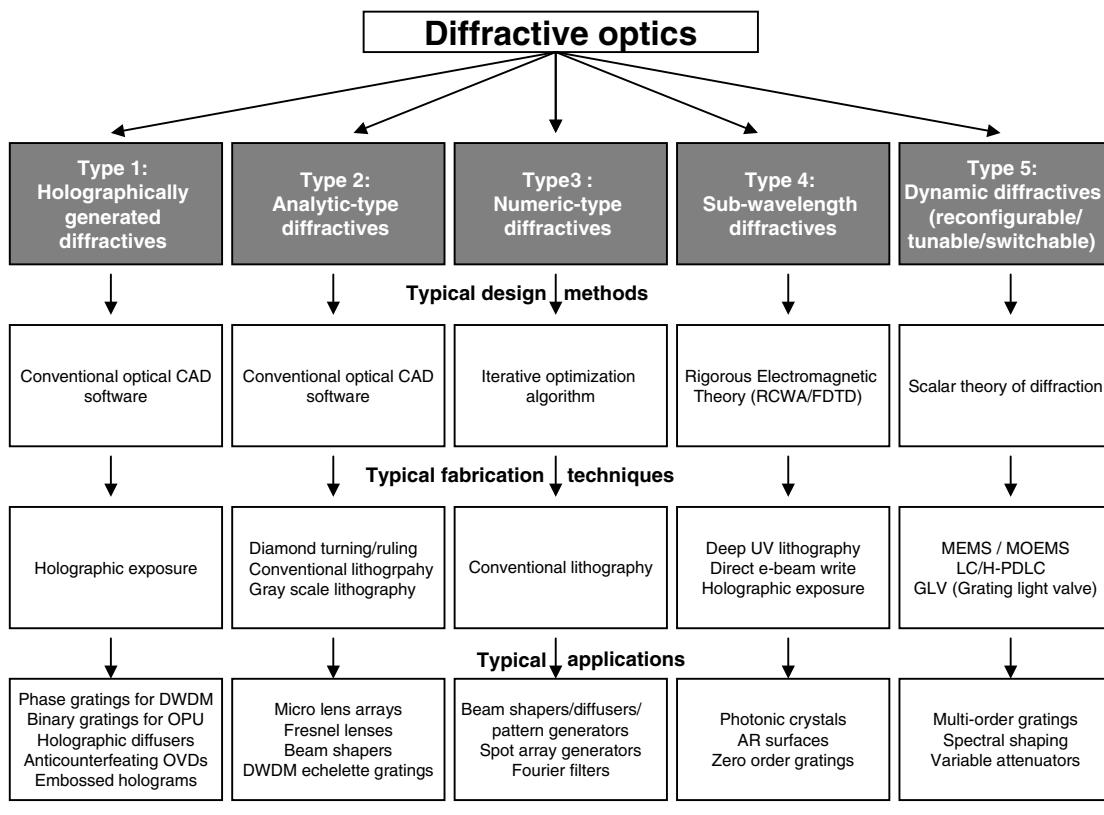
Diffractive Optics

Figure 13.5
The Many Names of Diffractives



There are roughly five different groups of diffractive optics that have been reported in literature since 1967, (when Prof. Adolph Lohman first introduced the concept of “Synthetic Hologram”²), which are categorized not so much along their optical functionalities, but rather along the design techniques and the physical implementations used to produce them. Note that many different techniques can be used to design a same diffractive optical functionality. Figure 13.6 summarizes these five different types³:

1. Holographic optical elements—are referring to traditional optical holographic recording of volume phase holograms (in phase modulation materials) or surface relief holograms (in photo resist materials).
2. Analytic type diffractives—refer mostly to elements that can be designed or optimized by the means of analytical methods, like ray tracing (as it is done in most of the optical computer-animated design [CAD] tools) or by solving an analytical equation (as it is done for Fresnel lenses or gratings). These are the most common diffractives.

**Figure 13.6**

The Five Different Diffractive Elements Types

3. Numeric type diffractives—refer mostly to elements that cannot be designed or optimized by analytical methods, and require stochastic iterative optimization procedures and algorithms. These elements can implement more complex optical functions than analytic-type diffractives, but have their limitations (amount of CPU power required, need to rasterize the element in the design process, etc...). They are increasingly used in industry (see also the section “Where Are Diffractives Used?”).
4. Subwavelength diffractive elements—(SWG or subwavelength gratings) refer to elements which basic structures are smaller than the reconstruction wavelength, and are thus highly polarization sensitive and act very differently from the previous two diffractive types.

5. Dynamic diffractives—refer to all the technologies used to implement reconfigurable, tunable, or switchable diffractive optical functionalities. Note that these elements can actually incorporate any of the four previous diffractive types (as for their design and fabrication techniques). This last type of diffractive has recently gained much interest for emerging optical market and applications (especially in laser display), as we will see later in the section “Where Are Diffractives Used”.

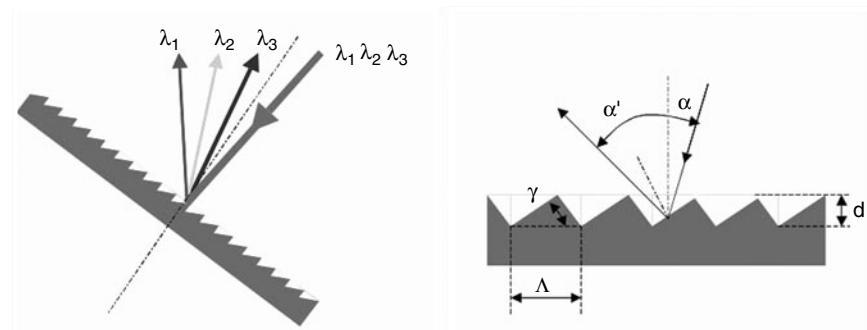
Diffraction Gratings

Linear diffraction gratings have historically been the first type of diffractive element that has been studied, fabricated and used successfully in industrial applications. They still account for most of the diffractives used in industry today (for example, in spectroscopy, dense wavelength division multiplexing [DWDM] telecom applications, optical security devices, optical data storage, optical sensors, etc...).

Linear diffraction gratings are designed with the grating equation (Eq. 13.2), and thus are analytic-type elements (Fig. 13.6). As they are rather simple in geometry, there is no need to use a special CAD tool to design such elements. They can be fabricated by a wide variety of techniques and technologies, from diamond ruling to holographic exposure and microlithography.

Figure 13.7 shows a reflective linear sawtooth grating (or échelette grating) which is used for its unique spectral dispersion characteristics (as in a DWDM telecom Demux).

Figure 13.7
A Reflective Sawtooth Linear Grating



The spectral dispersion (or strength of the wavelength demultiplexing functionality) of such a grating can be written as:

$$\frac{\partial \alpha'}{\partial \lambda} = \frac{m}{\Lambda \cos(\alpha')} \quad (13.4)$$

The resolution of such a grating can be written as:

$$R = mN_o \quad (13.5)$$

where Λ is the period of the échelette grating, λ the incident wavelength, α the incident angle, α' the diffracted angle, N_o the number of grating grooves, and m the target diffraction order considered in the application.

Note that in many cases, the optimal diffraction order is not the fundamental order, but a much higher order (diffraction orders up to 15 are commonly used in DWDM gratings).

Diffractive Optical Elements

DOEs—analytic type elements—are the most popular diffractives used today (although numeric type elements—CGHs—are getting used more and more in industry, see section “Where are Diffractives Used?” later). These elements have usually optical power (that is, lenses, unlike gratings or Fourier elements which have no optical power). The Fresnel lens is the most straightforward example of a DOE.⁴ DOEs are usually calculated in an analytical way, and in most cases one would use a classical optical CAD tool based on ray tracing to optimize an aspherical phase profile within a plane or curve (holographic or diffractive plane).

It is interesting to note that unlike its refractive lens counterpart, fabricating a highly aspherical diffractive lens bears the same price tag as fabricating a simple spherical diffractive lens.

The phase profile used to describe a diffractive element can take on many forms depending on the optical software used to design and optimize it. There are mainly three different analytical descriptions used in today’s CAD tools to describe diffractive phase profiles:

1. The traditional sag equation (Eq. 13.6):

$$\theta = \frac{2\pi}{\lambda} \left(\frac{Cr^2}{1 + \sqrt{1 - (A + 1)r^2C^2}} \right) + \sum_{i=1}^n C_i r^{2i} \quad (13.6)$$

2. Rotationally symmetric elements (simple polynomial in r —Eq. 13.7):

$$\theta = \sum_{i=2}^n C_i r^i \quad (13.7)$$

3. Nonrotationally symmetric elements (general polynomial in x and y —Eq. 13.8):

$$\theta = \sum_{j=1}^m \sum_{i=1}^n C_j x^i \cdot C_j y^i \quad (13.8)$$

Once the aspheric phase profile is defined in this infinitely thin surface (which can be planar or mapped on a refractive aspherical curvature surface), the phase profile is sliced into 2π phase shift slices for the considered wavelength for maximum diffraction efficiency in the fundamental positive order, a process shown in Fig. 13.8.

An alternative and simpler way to compute the position and widths of the zones of a spherical Fresnel lens, is to calculate directly integer numbers of waves departing from the desired focal point and aiming at the DOE plane (see Eq. 13.9 and Fig. 13.9):

$$R_n = z + i \cdot \frac{\lambda}{2} \quad (13.9)$$

USEFUL PARAMETERS OF DIFFRACTIVE LENSES The numerical aperture (NA) is an important parameter of a diffractive lens and can be expressed as in Eq. 13.10 (where D is the diameter of the lens, α_{\max} the

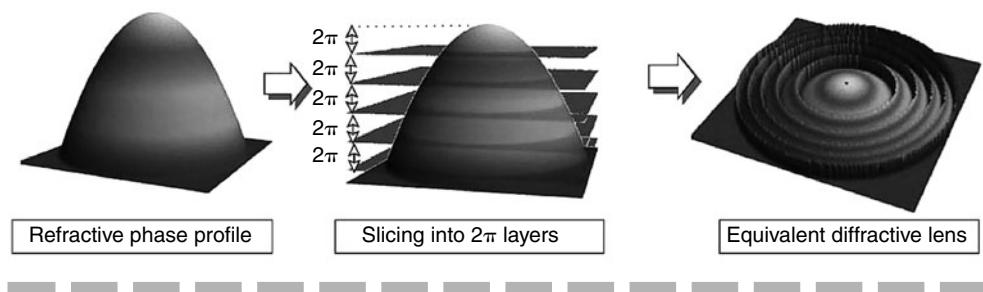
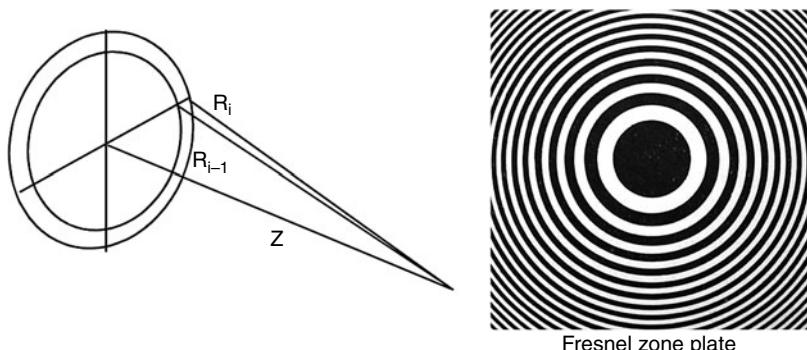


Figure 13.8
From the Initial Aspherical Phase Profile to the DOE Fringes

Figure 13.9
The Fresnel Lens
(Fresnel Zone Plate)



maximum diffraction angle at the edges of the lens, λ the reconstruction wavelength and f the focal length).

$$\text{NA} = \sin(\alpha_{\max}) = \frac{\lambda}{\Lambda_{\min}} = \sin\left(\arctan\left(\frac{D}{2f}\right)\right) \approx \frac{D}{2f} \quad (13.10)$$

The minimum fringe width δ of such a lens (or the minimal local grating period Δ_{\min} at the edges of that lens) becomes then (Eq. 13.11):

$$\delta = \Lambda_{\min} = \frac{\lambda}{\text{NA}} \quad (13.11)$$

This smallest feature is also a very important parameter, since this usually is the “go or no go” red flag when designing a lens to be fabricated by a specific technology, since *a lens is only useful if it can be actually fabricated*. For example, a 16-phase level Fresnel lens would require a $\delta/16$ minimum feature size to be fabricated. So if the minimum dimension printable on the wafer is $1.0 \mu\text{m}$, the minimum printable period of the lens will be $16 \mu\text{m}$. This parameter is also handy, as we will see later in the section on diffraction efficiency, where the efficiency of a lens is actually also a function of the local period. If the period gets too small (although printable), the efficiency can drop considerable and/or parasitic polarization effects can occur.

BROADBAND DIFFRACTIVE OPTICAL ELEMENTS Diffractive elements are very sensitive to wavelength changes (strong spectral dispersion), both in their focal plane (chromatic aberrations) and in their efficiency (see also fabrication section). Nevertheless it is, possible to optimize diffractives (for example, Fresnel lenses) to function over a wider range of wavelengths or over a set of predetermined individual

wavelengths. These lenses are commonly referred to as multiorder or harmonic diffractive lenses. The etch depth of such lenses is optimized so that its phase difference would yield an integer number of 2π phase shifts for each wavelength (see Fig. 13.10). This has the advantage of widening the fringe width, but also the inconvenient of deepening the groove depth. So precaution has to be taken when using this technique.

INTERFEROGRAM-TYPE DIFFRACTIVE OPTICAL ELEMENTS

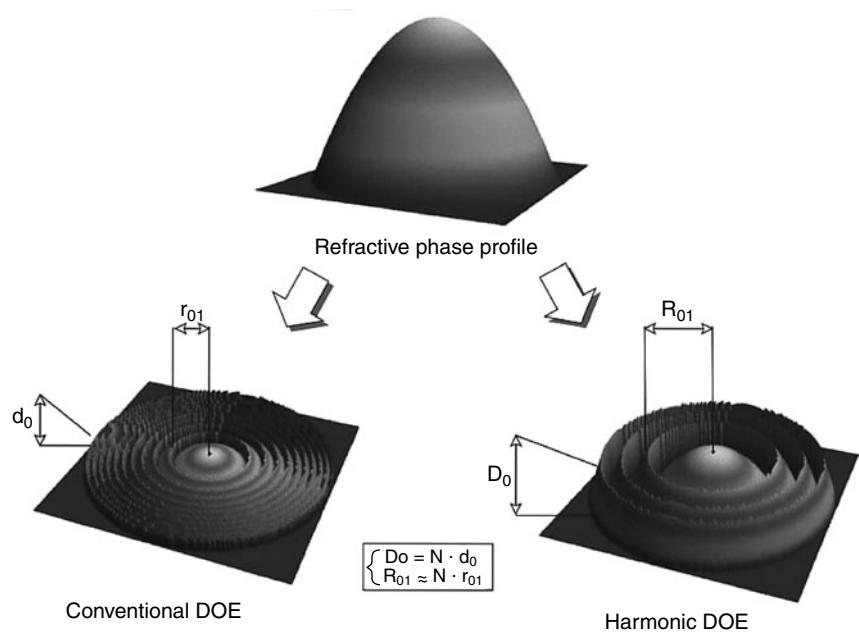
Another type of analytical element used in literature is the interferogram-type DOE⁵, in which phase function is calculated as the interference pattern of a given object wave and a reference wave (similar to holographic exposure).

As an example, consider a tilted planar reference wavefront which makes an angle α with the DOE plane, and interfering with a normal planar wavefront. The resulting interference pattern can be expressed in Eq. 13.12:

$$t(x,y) = c + 2A_0(x,y) \cdot \cos \left[\varphi_0(x,y) + 2\pi \frac{\sin(\alpha)}{\lambda} \cdot x \right] \quad (13.12)$$

Although the physical interference pattern is an intensity distribution, the resulting DOE interferogram can be recorded as a pure phase

Figure 13.10
Narrowband and
Broadband Diffrac-
tive Lens Genera-
tion from a Same
Phase Profile



element (in phase or surface relief modulation), or as an amplitude element. These elements have very remarkable properties when used in off-axis configurations as complex lenses (like toroidal, conical or helicoïdal lenses), but lack in diffraction efficiency due to their typical sinusoidal phase profile (which limit their efficiency to 33% in sinusoidal mode or 40% in binary mode—see fabrication section). In order to increase their efficiency as a surface relief element, one can fabricate the calculated fringes as blazed structures instead of sinusoidal structures.

Hybrid Optical Elements

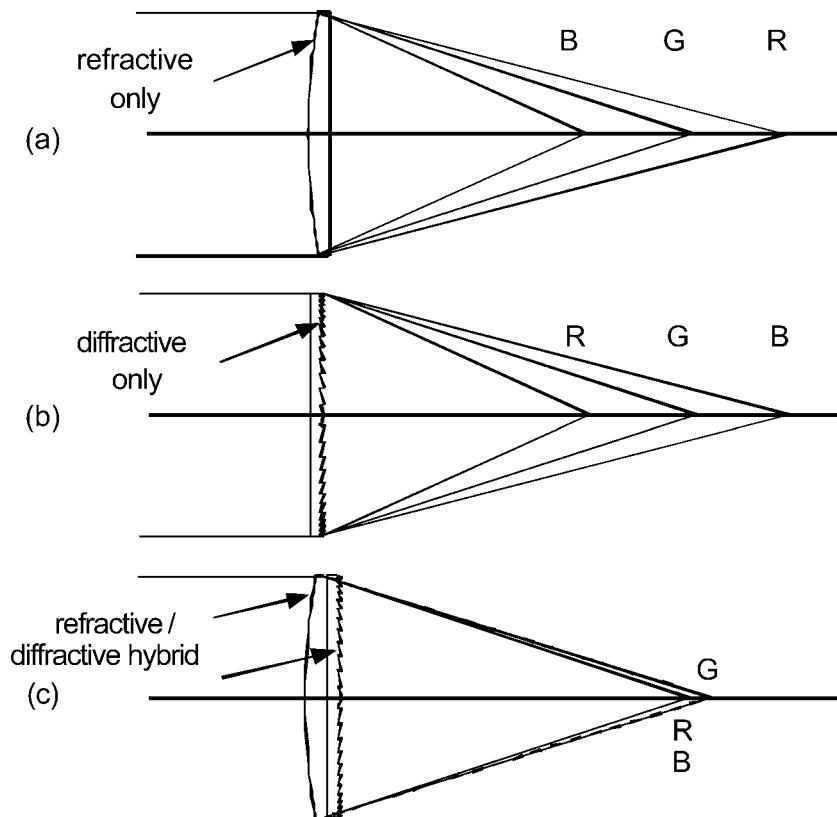
As discussed in the introduction lines of the chapter, diffractives are best used in combination with conventional optical elements, as hybrid optical elements⁶ in order to extend existing optical functionality, or introduce additional functionality. Traditional ways of achromatizing⁷ or athermalizing⁸ refractive lenses is to insert a diffractive structure carefully designed, so to balance the spectral dispersion as the Abbe V numbers of refractives and diffractives have opposite signs. As the spectral dispersion of diffractives is much stronger than for refractives, an achromatic singlet has typically a much stronger refractive power than diffractive power—about 10 times or more (see Fig. 13.11).

The section below, “What Design and Modeling Tools Should I Use?”, shows examples of parametric design of such hybrid optical elements.

Another very promising application of hybrid refractive/diffractive optics is the dual focus optical pickup (OPU) lens, which is used in most of the CD/DVD readers available in the market today. Such a lens has a convex/convex refractive, with one bearing a diffractive profile. The diffractive profile is intentionally detuned so that it produces only 50% efficiency in the fundamental positive order, leaving the rest of the light in the zeroth order. Such a lens therefore creates two wavefronts to be processed by the refractive profiles (which are always 100% efficient). When the various profiles are carefully optimized so that the zero-order combined with the refractive profiles would compensate for spherical aberrations at 780 nm wavelength through a CD disk overcoat, and the diffracted fundamental order combined with the same refractive profiles would compensate for spherical aberrations at 650 nm wavelength through a DVD overcoat, this lens is ready to pick up either CD track and DVD tracks.

Diffractive Optics

Figure 13.11
Achromatization
with Hybrid Diffrac-
tive/Refractive
Singlet



Hybrid optics and hybrid optical compound lenses are also very suitable for the design of triple focus lenses to be used in the next generation BD OPUs (blue ray disks), which feature three different wavelengths (780 nm, 650 nm, and 405 nm), three different focal lengths, and three different spherical aberrations, to compensate for three different disk media overcoat thicknesses.

Computer Generated Holograms

CGHs are numeric type elements (see Fig. 13.6), which are often used as single standing optical elements, unlike analytic-type elements. The versatility of the numerical iterative optimization processes of CGHs allows the design of complex optical functionalities which could not have been implemented by classical refractive or reflective optics earlier.

CGHs are mainly used as complex optical wavefront processors, and are the tools for the development of unconventional optical applications, especially in displays, as we will show in the last section of this chapter.

There are two types of CGHs, the Fourier type and the Fresnel type.⁹ The first one projects the desired pattern in the far field whereas the second one reconstructs the pattern in the near field, and thus can be considered as a complex diffractive lens (bearing optical power as well as optical shaping/splitting).

Figure 13.12 depicts a Fourier and Fresnel CGH, which implement nearly the same optical functionality, namely splitting the incoming beam into an array of 3×3 beams. While the first one would create nine collimated beams from a single collimated beam, the second one would create nine converging beams from the same collimated beam, and therefore could be considered as a multifocus diffractive lens. Note the fringe-like pattern in the Fresnel type which does not appear in typical Fourier-type elements.

The notable difference between a standard array of 3×3 diffractive microlenses and a 3×3 multifocus diffractive lens is that the NA of the multifocus lens is much larger than for the standard microlens array (three times larger). We can therefore call such a multifocus lens a phase multiplexed microlens array, in opposition to the conventional space multiplexed microlens array.

Many different iterative optimization algorithms have been successfully developed and applied in literature to the CGH design process^{10–12}, some of them listed in the following table (Table 13.1), along with their specifications:

Figure 13.12
Fourier and Fresnel
Type CGH

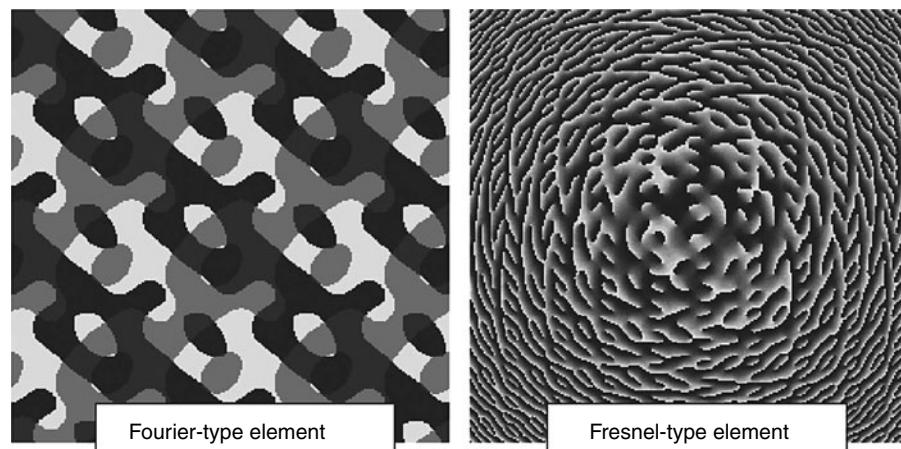


TABLE 13.1 The Various CGH Optimization Algorithms Used in Literature

Algorithm	Convergence Speed	Typical Application	Advantage	Inconvenient
Inverse FFT	Very fast	Fast display	Very simple to program	Low efficiency low uniformity, low SNR
DBS (Direct binary search)	Relatively fast	Spot array generators	Easy to program	Gets stuck into local cost function minima
Simulated annealing	Slow	Spot array generators	Very good uniformity	Slow
IFTA (GS)	Very fast	Beam shapers/displays	Very fast	Lacks uniformity
Genetic algorithms	Very slow	Spot array generators	Very versatile	Very slow, limited to small sizes
Direct analytical	Very fast	Fourier filters (band pass, edge detection,...)	Very simple to implement	Only available for Fourier filtering

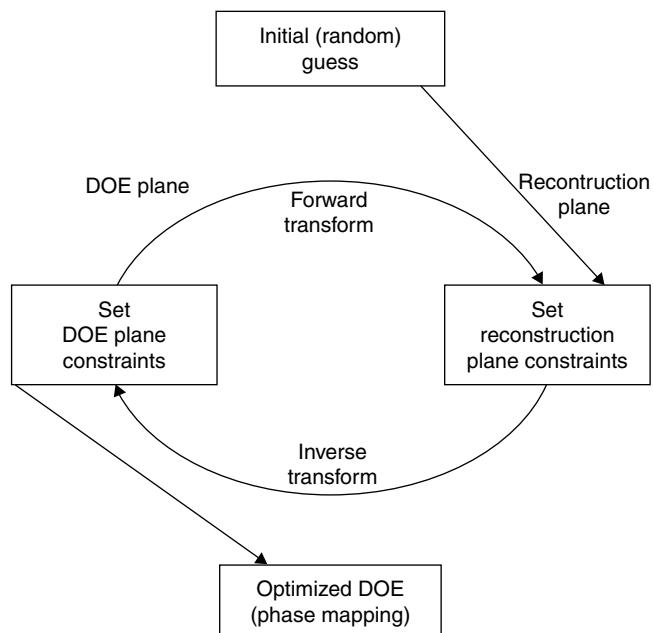
The most commonly used algorithm in industry to design both Fresnel or Fourier CGHs is an iterative Fourier transform algorithm (IFTA) called the Gerchberg-Saxton (G-S) algorithm¹³, originally developed for e-beam microscopy phase retrieval. It is both fast and very reliable (that is, it converges in most of the cases), and very easy to implement and use. However, although diffraction efficiency is optimized pretty well with the G-S algorithm, it is not the case of uniformity. Figure 13.13 shows the flow chart of the G-S algorithm.

Once the complex data of the CGH have been calculated (optimized), the optical design engineer has to choose a specific encoding technique in order to encode the phase, amplitude or complex information into a substrate.¹⁴

We listed below some of the most commonly used CGH optical functionalities¹⁵:

- Optical image processing and Fourier filtering (edge detection, low- and high-pass filters, etc.)
- Optical interconnections via spot array generators
- Far field pattern generation for two-dimensional display or focus control

Figure 13.13
The G-S Iterative CGH Optimization Algorithm (IFTA Type)



- Beam shaping (for example, Gaussian to top hat)
- Beam splitting
- Beam sampling
- Structured illumination generation for machine vision

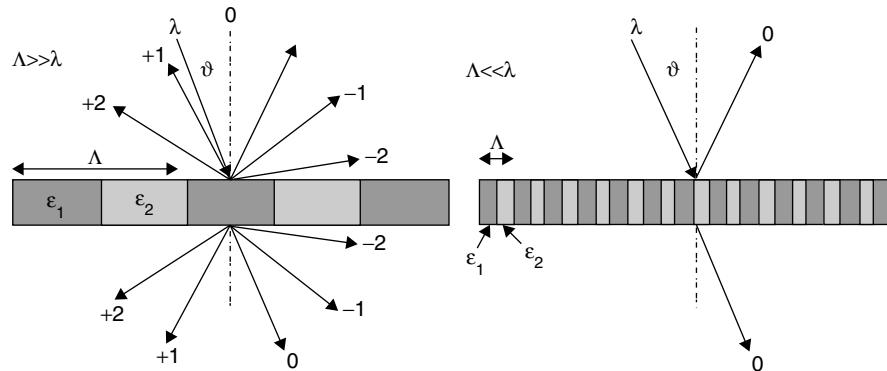
Subwavelength Diffractives and Photonic Crystals

Subwavelength (SW) diffractives (type 4 elements in Fig. 13.6 classification) are microstructured diffractive elements in which the smallest features (or local grating periods) are smaller than the reconstruction wavelength ($\Lambda/\lambda \ll 1$), such as only the zero—backward and forward—orders are propagating, all other higher diffraction orders are evanescent.

As the incident wave cannot resolve the SW structures, it sees only the spatial average of its material properties (effective refractive index). Figure 13.14 shows the differences in diffracted and propagating orders between a standard multiorder diffractive and a SW diffractive.¹⁶

Diffractive Optics

Figure 13.14
Multiorder Diffractives
and SW Diffractives
Operations



A zero-order grating—or SWG—(where no other diffracted order is propagating) is defined when Eq. 13.13 holds true:

$$\frac{\Lambda}{l} < \frac{1}{\text{Max}\{\sqrt{\varepsilon_1}, \sqrt{\varepsilon_{III}}\} + \sqrt{\varepsilon_1} \sin \theta_{\max}} \quad (13.13)$$

Note that SW diffractives can be implemented either as thin surface relief phase elements or as index modulated holographic elements. These elements also show very strong polarization dependence, unlike standard diffractive with much larger feature sizes.

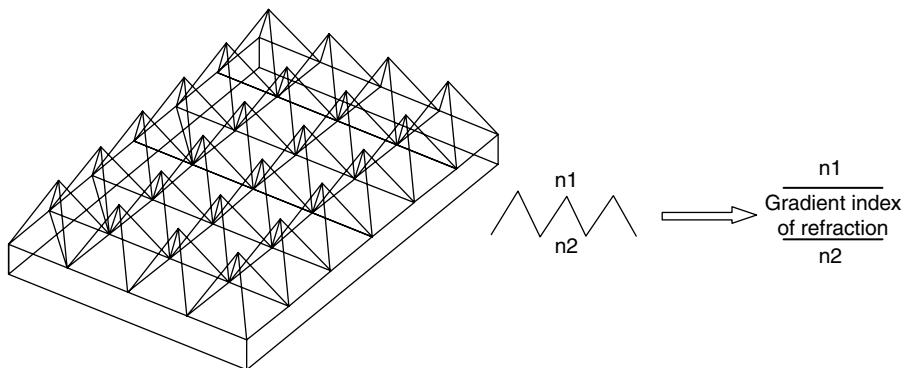
Applications requiring SW diffractives include mainly:

- Polarization sensitive elements (polarization splitting, polarization combining)
- Antireflection (AR) surfaces
- High-resolution resonant filters (in reflection or transmission)
- Integrated waveguide gratings (with Bragg reflection or coupling effects)
- Phase plates (in reflection and transmission)

Figure 13.15 shows such an antireflection surface composed by a two-dimensional array of pyramidal structures, with SW periods.

Another application where SW diffractives show their full potential is the encoding of smooth phase profiles as small binary SW structures.¹⁷ For example, the design and fabrication of SW blazed Fresnel lenses or sawtooth gratings have been reported. The incident wave does not

Figure 13.15
Example of a Two-Dimensional SW Period Grating Producing an Antireflection Surface



resolve anymore the local binary phase modulation, since its wavelength is much larger than these structures, but sees the overall effective medium picture. Thus, the incident wave considers the material as a smoothly varying phase element. The effective medium theory (EMT) is applied to such elements in order to model them.

Holographic Optical Elements

HOEs—type 1 elements in the Fig. 13.6 classification—are considered as digital diffractive elements in the sense that the recording setup is usually designed by a CAD tool, although the fabrication remains mostly analog (holographic recording). Figure 13.16 shows the holographic recording set-up of a simple linear HOE grating.

A holographic grating is considered as thick or thin in Bragg incidence when its quality factor Q is respectively larger than 10 or lower than unity. For values of Q in between 1 and 10, the grating behaves in an intermediate state. Q is defined in Eq. 13.14 as:

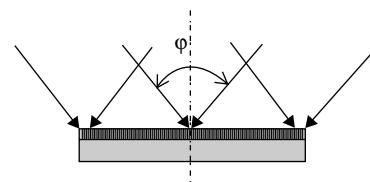
$$Q = \frac{2\pi\lambda d}{n\Delta^2 \cos\alpha} \quad (13.14)$$

Figure 13.16
The Recording of a Linear Grating HOE as a Surface Relief Element in Resist

Holographic grating example:

The period Λ of the interference fringes generated when two beams of wavelength λ intersect at an angle of φ is given by:

$$\Lambda = \frac{\lambda}{2 \sin \frac{\varphi}{2}}$$



where d is the thickness of the grating, Λ its period, n its average refractive index and α the incident angle.

See also the section “How Are Diffractives Fabricated?”, later in this chapter for further information about holographic recording of HOEs.

What Design and Modeling Tools Should I Use?

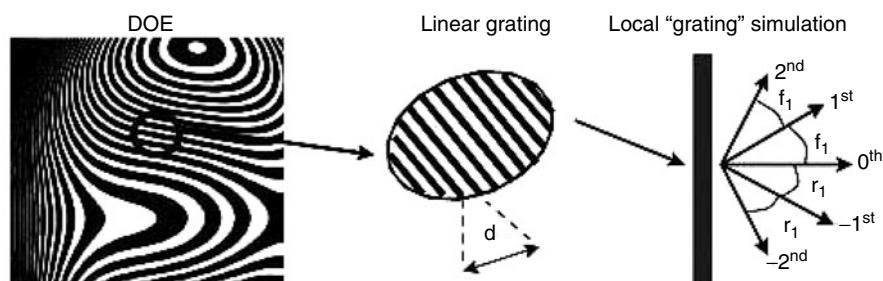
In the previous section we have described the various design techniques used today to calculate and optimize analytic as well as numeric type diffractives. In this section, we will focus on the various tools available to the optical engineer to accurately model the behavior of these diffractives (CAD tools used for modeling effects of illumination and/or opto-mechanical tolerancing, effects of various fabrication techniques and technologies and related systematic fabrication errors—see also next section on fabrication).

Ray-Tracing-Based Tools

Since diffractives are very often used in conjunction with other optical elements (refractives, reflectives, graded index, etc.), these simulation tools have to be able to interface with standard optical modeling tools, for the most based on ray tracing. Therefore, many optical CAD tools available on the market use simple ray tracing algorithms through diffractives as is the diffractive it would be special refractive element. The ray tracing uses the local grating approximation (LGA) for any diffractive element constituted by smooth fringes. Note that this is possible only if the element is constituted of fringes like a DOE or Fresnel CGH, a Fourier CGH could not be modeled this way. Figure 13.17 shows how the local grating approximation uses the grating equation in order to predict the direction of the ray passing through a particular area of the DOE.

The local angle of diffraction is given by the grating equation at the equivalent local grating location, and the efficiency at that location is computed as a function of the diffraction order considered, the groove depth, the wavelength, and the number of phase levels.

Figure 13.17
Ray Tracing by LGA
of an Aspherical Diffractive Element



Note that this technique does not inform about crucial aspects like real diffraction efficiency, multiorder diffraction, and resulting multi-order interferences. Therefore, this technique can only be used effectively for 100% efficient elements, like blazed Fresnel lenses or gratings.

This technique is thus best suited for the modeling of hybrid refractive/diffractive elements, where the diffractive element is usually fabricated by diamond turning, and thus yields very smooth fringes and high diffraction efficiency. This method constitutes the vast majority of diffractive optics modeling tools available in optical CADs on the market today.

Other similar techniques compare the diffractive to a refractive element which would have an infinite refractive index, and thus being infinitely thin, the Sweatt model.¹⁸

Scalar-Diffraction-Based Tools

In order to simulate the effects of multiorder diffraction in the scalar domain⁹, as well as the effects of fabrication techniques and systematic fabrication errors, it is best to use a physical optics approach, and implement scalar-diffraction-based propagators, which consider in parallel all propagating orders through the diffractive. Such scalar diffraction propagators can model any type of diffractive structures, composed of fringes or not (DOEs and CGHs), as analog, binary or multilevel surface relief elements. This is true only in the realm of validation of scalar diffraction theory, that is, as long one is in the paraxial regime (low NAs [see Eq. 13.10], low angles, and smallest structures much larger than the wavelength, without any polarization effects).

Helmholtz's wave equation, with Huyghen's principle of secondary sources over the wavefront's envelope, injected in Green's function is the major foundation of scalar theory, and gives rise to the Helmholtz and

Kirchhoff integral theorem. The Rayleigh-Sommerfeld diffraction formulation for monochromatic waves follows, and gives rise to the Fresnel and then Fourier approximations of the diffraction through a thin planar screen in the far and near fields.

These two formulations are the basis of most of the physical optics modeling tools used today. Equation 13.15 summarize these two formulations:

$$\left. \begin{array}{l} \text{Fourier approximation (far field)} \\ \text{Fresnel approximation (near field)} \end{array} \right\} \begin{aligned} U_D(u, v) &= \iint_{\infty} U(x_1, y_1) \cdot e^{-j2\pi(ux+vy)} \cdot dx \cdot dy \\ U_D(x_0, y_0) &= \left(\frac{e^{jkf}}{j\lambda f} \right) \iint_{\infty} U(x_1, y_1) \cdot e^{j\frac{\pi}{\lambda f}((x_0-x_1)^2+(y_0-y_1)^2)} dx_1 \cdot dy_1 \end{aligned} \quad (13.15)$$

Note that the Fresnel approximation can be described in two different ways, as a direct integral here or as a convolution. When using complex two-dimensional fast Fourier transform (FFT) algorithm, one can rewrite the previous equations and implement them in a numerical tool.

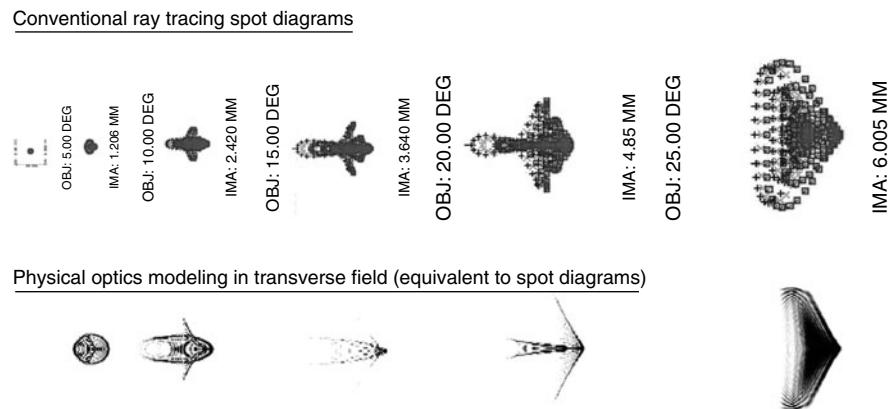
As an application example, we consider here a hybrid refractive/diffractive lens used in an infrared (IR) digital camera. The lens has a spherical/aspherical convex/convex surface with an aspherical diffractive surface on the aspherical refractive surface (second surface). The lens is optimized and modeled in a standard optical CAD tool from the market as far as ray tracing is concerned.¹⁹

We will perform a classical lens field of view analysis, which means that we will launch a collimated wavefront on this lens under increasing angles and compute the resulting focus plane, both with standard ray tracing tools and with a numerical Fresnel propagator based on the previously described analytical formulation of diffraction in the near field.

The resulting lateral spot diagrams for incident angles from 0° to 25° are shown in Fig. 13.18.

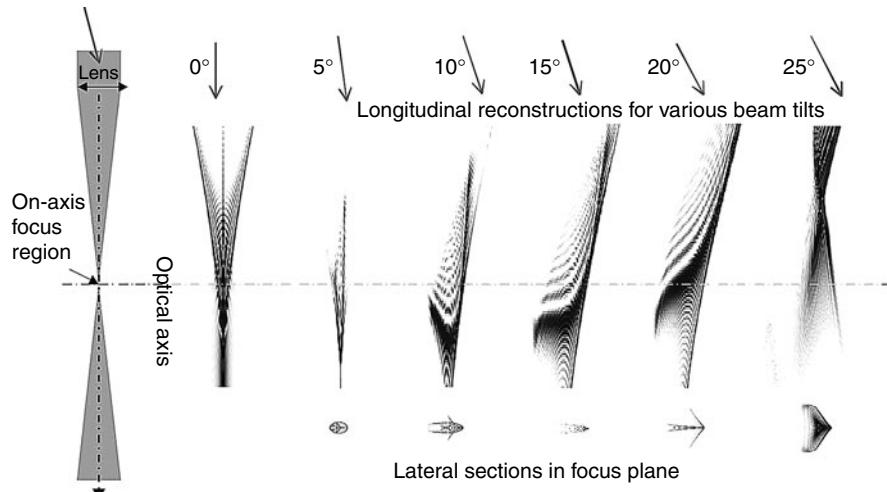
Both modeling tools (ray trace and scalar numerical Fresnel propagators) agree well on all spot diagrams, even with angles up to 25°. However, the resolution in the numerical analysis tool when using the scalar model is much greater than when using only geometrical ray trace. Also, effects of multiple diffraction orders interferences (see the fringes created in the reconstructions) can be observed, which cannot be demonstrated with conventional ray trace algorithms.

Figure 13.18
Ray Traced Spot Diagrams and Corresponding Physical Optics Modeling of a Hybrid Refractive/Diffractive Lens—a Comparison in the Lateral Plane



Scalar propagation of diffracted/refracted/reflected wavefronts gives actually much more flexibility in modeling diffractives (and any other microoptical elements), in the sense that not only lateral spot diagrams can be computed, but also longitudinal intensity profiles, as shown in Fig. 13.19. Actually, the reconstruction can be computed on any surface, planar or curved, even in a three-dimensional volume. This gives a much deeper insight on the modeling aspects, which was not possible with only ray tracing.

Figure 13.19
Longitudinal Reconstructions via Numerical DFT-Based Scalar Propagators



FFT- VERSUS DFT-BASED NUMERICAL PROPAGATORS FFT based propagators, however have yield some severe drawbacks, which are namely limitations on the size of the sampled field, location of the reconstruction plane as well as amount of off-axis of that reconstruction plane. Discrete Fourier transforms (DFT) are thus used in order to implement the Fresnel and Fourier approximations of the Rayleigh-Sommerfeld integral (Eq. 13.15). Actually, the exact Rayleigh-Sommerfeld integral can also be implemented by using DFT algorithms, which is not the case with FFT algorithms.

The main disadvantage of DFT based propagators is the CPU required to compute the numerical reconstruction. There is actually an exponential factor between the speed of FFT based and DFT based numerical propagators. The numerical reconstructions in Fig. 13.19 have been computed by using a DFT based propagators.

Rigorous Electro-Magnetic Modeling Techniques

Scalar theory of diffraction as expressed previously as a thin phase approximation is an accurate tool for elements which have smallest feature dimensions on the order of the reconstruction wavelength.^{20,21}

In order to account for the coupling effects along the boundary of the microstructures, rigorous electromagnetic modeling techniques should be used. Many different techniques have been reported in literature. The modal method (MM), the finite difference time domain (FDTD), and the effective medium (EM) theories are amongst them.

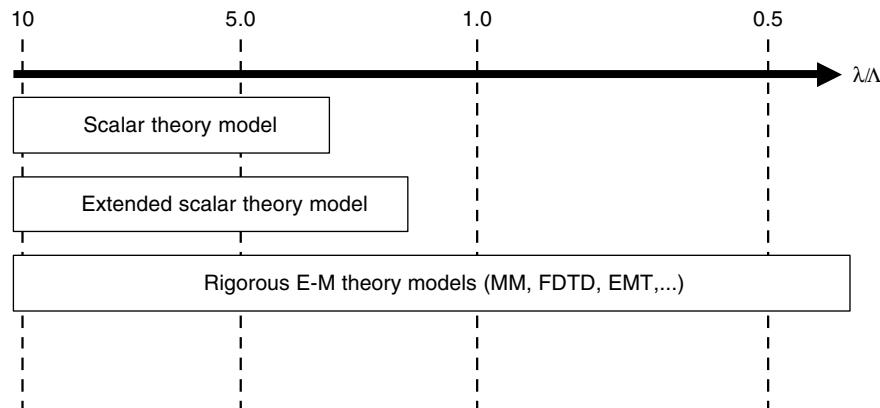
In between the realm of validity of scalar theory of diffraction and the more rigorous EM theories, extended scalar theories have also been reported, which deal mainly with the shadowing effect for high angles in surface relief elements. Figure 13.20 shows the three modeling regions and which technique should be applied.

When contemplating the possibility to use a rigorous modeling technique to design and model a diffractive element, it is important for the optical engineer to remember that due to high CPU time and complex implementation of the numerical tools, anything more complex than a linear grating is nearly impossible to model. Therefore, rigorous theories are not often used to design DOEs or CGHs.

It is also interesting to remember that the diffraction angles (and thus the position and geometry of the optical reconstruction) predicted by

Figure 13.20

Realm of Validity for Scalar and Rigorous Diffraction Theories



scalar theory still hold true even though if outside of the realm of scalar theory. Only the prediction of the diffraction efficiency is different from those predicted by the more accurate rigorous theories.

Parametric Design Example of Hybrid Optics via Ray Trace Techniques

In order to illustrate hybrid diffractive/refractive optics design through ray trace techniques with a conventional optical CAD tool and demonstrate the relative merits of these designs, we will show several representative examples. The specifications for the design example are as follows:

Entrance pupil: diameter 25 mm

Field of view on axis only

Wavelengths C (656.3 nm), D (587.6 nm), F (486.1 nm)

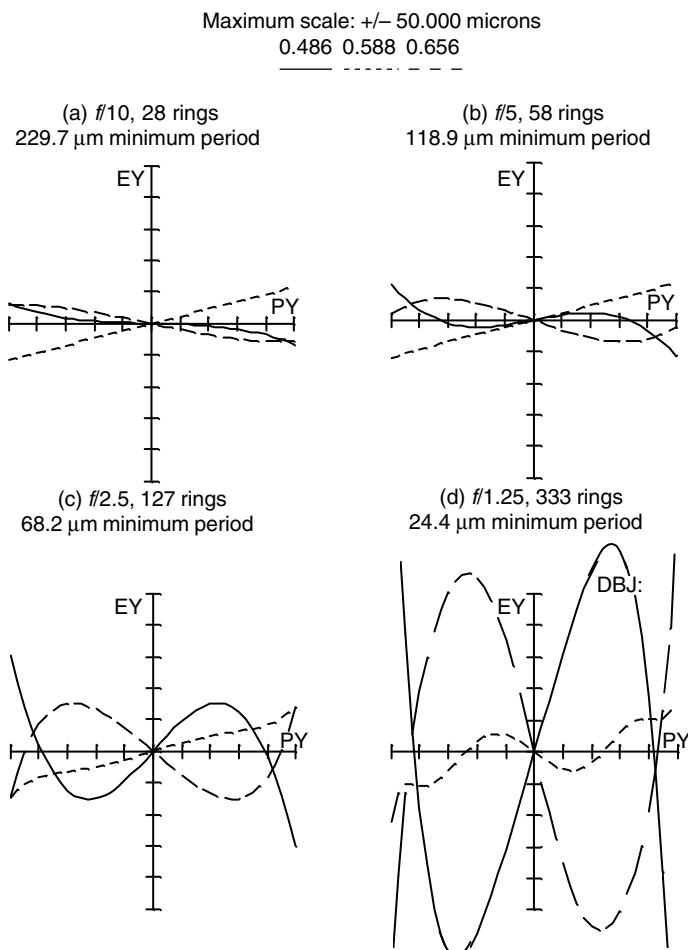
f#: f/10, f/5, f/2.5, f/1.25

Figure 13.21 shows the transverse ray aberration curves for a f/10 hybrid singlet per the specifications above. The substrate material is BK7 glass, and very similar results would arise for acrylic, which would be a fine material choice if the element were to be mass produced.

The diffractive surface is located on the second surface of the lens. The spacing or separation between adjacent fringes was allowed to vary with respect to the square as well as the fourth power of the aperture radius, or y^2 and y^4 , where y is the vertical distance from the vertex of

Diffractive Optics

Figure 13.21
Hybrid Refractive/Diffractive Achromatic Singlet as a Function of $f\#$

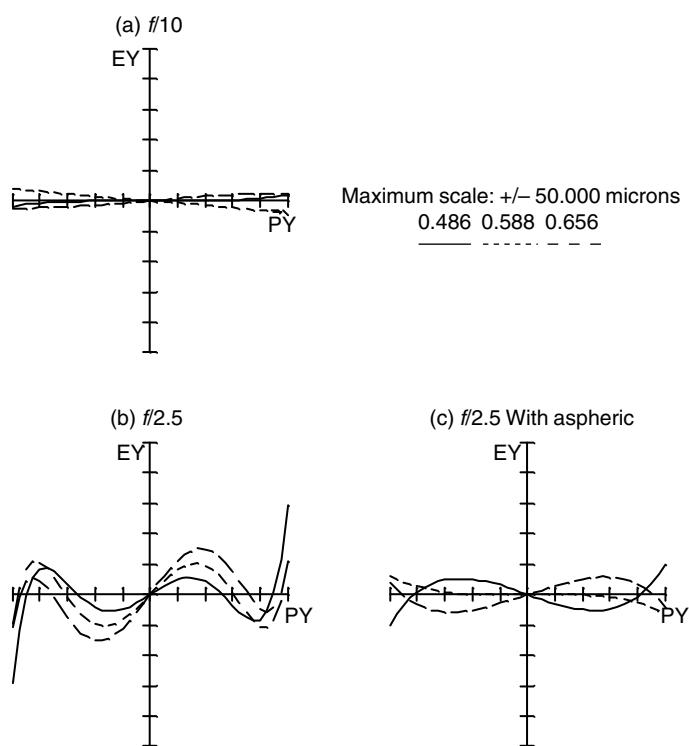


the surface perpendicular to the optical axis. The quadratic term allows for the correction of the primary axial color whereby the F and C light (blue and red) are brought to a common focus. The fourth order term allows correction of the third order spherical aberration as well. The resulting ray trace curves show the classical performance typical of an achromatic doublet. Since the lens is of relatively high $f\#$, the spherical aberration at the center wavelength is fully corrected. There is a residual of spherochromatism which is the variation of spherical aberration with wavelength. This residual aberration is due to the fact that the dispersion of the diffractive is linear with wavelength whereas the material dispersion

of BK7 glass is nonlinear. The resulting surface has 28 rings with a minimum period of $229.7 \mu\text{m}$. The insert for injection molding this surface could be easily diamond turned. Figure 13.22 shows similar results for hybrid singlets that are $f/5$, $f/2.5$, and $f/1.25$, respectively. Note that as the $f\#$ gets lower and lower the higher order spherical aberration increases, and the spherochromatism increases as well, to a point where the spherochromatism is the predominant aberration. The number of rings and the minimum fringe period in the diffractive are listed.

Note that these data do not scale directly, as we can do with conventional optical designs. As a diffractive optical element is scaled down in focal length while maintaining its $f\#$, a true linear scaling of all parameters (except of course the refractive index which is unitless, and thus does not scale), is not correct. This is because we need to maintain the fringe depths to create a total of 2π phase shifts for the target wavelength, and a linear scaling would result in only 1π phase shift. Thus, we find that for a 0.5 scaling, we end up with one half of the number of fringes with

Figure 13.22
 $f/10$ and $f/2.5$ Classical Achromatic Doublets



approximately the same minimum fringe period. For example, if we were interested in a 12.5 mm diameter $f/2.5$ hybrid, we would scale the radii, thickness, and diameter by $0.5\times$ from the 25 mm starting design. However, the number of fringes will decrease by a factor of 2, with essentially the same minimum fringe period. It is highly recommended to reoptimize any lens or lens system containing one or more diffractive surfaces after scaling in order to assure that the surface prescription is correct.

It is feasible to manufacture diffractive surfaces as well as binary surfaces with minimum periods of several microns or less, however it is best to discuss specific requirements with the foundry prior to finalizing the design (see also the section "How Are Diffractives Fabricated?"). Figure 13.22 shows, for comparison, the performance of classical $f/10$ and $f/2.5$ achromatic doublets using BK7 glass.

The results are somewhat improved over the hybrid solution. Note that at the lower $f\#$ of $f/2.5$ (b) the spherical aberration of the achromatic doublet is becoming a problem, and we show an aspherical design in (c).

Figure 13.23 shows graphically the relationship between the lens $f\#$, the number of fringes, and the minimum fringe period for the hybrid designs in Figure 13.21. As we would expect, the lower the $f\#$, the larger the number of fringes and the smaller the minimum period. Figure 13.24 shows several design scenarios, all for a constant $f/5$ single element lens.

Figure 13.23
Number of Fringes
and Minimum Period
as a Function of $f\#$
for a Hybrid Singlet

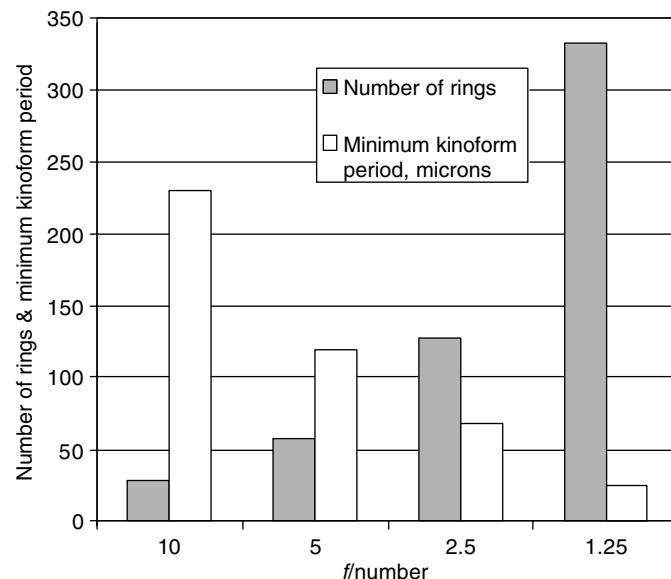
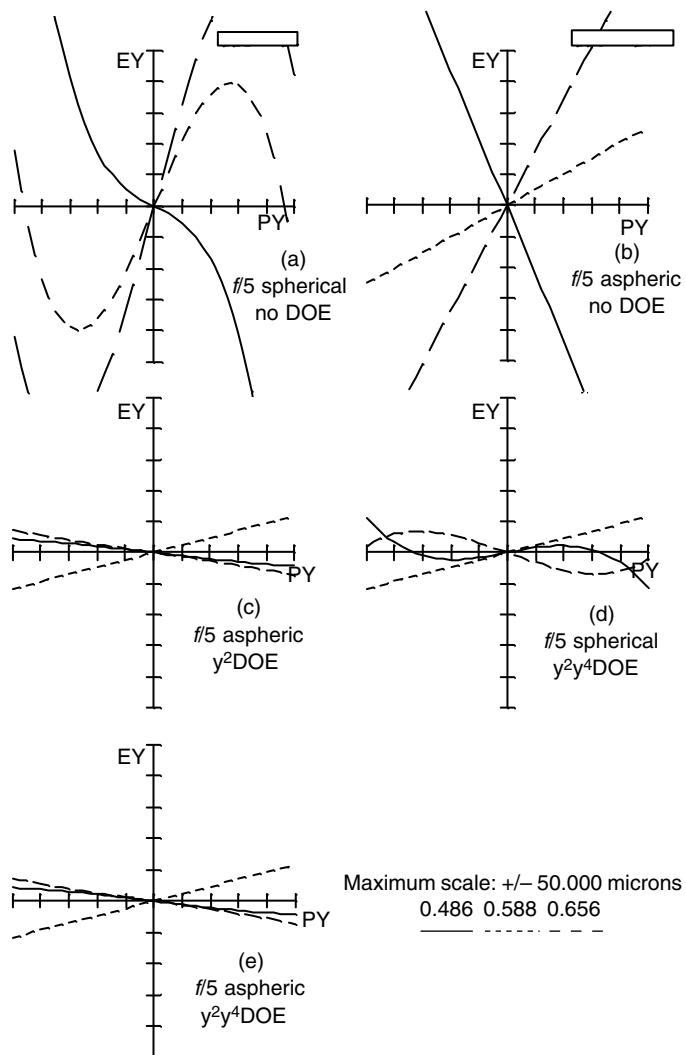


Figure 13.24
Performance of f/5
Hybrid Singlets with
Different Surface
Descriptions



For reference, (a) and (b) both have no diffractive surface, however (b) does have an aspheric surface for correction of spherical aberration. The primary axial color is the same in both lenses, and quite large as expected. Figure 13.24c has an aspheric surface for spherical aberration correction and a quadratic diffractive surface for correction of the primary axial color. Figure 13.24d is all spherical with a quadratic and a fourth order diffractive fringe width variation. It is interesting that this solution is very similar to (c), except that this solution has more spherochromatism than

the solution with the aspheric surface. Finally, (e) allows both an asphere as well as a quadratic and fourth order diffractive fringe width variation. The aspheric along with the quadratic diffractive fringe width variation of (c) was so well corrected that no further improvement is possible here.

One of the more interesting observations is that the aspheric surface along with the diffractive surface allows for the correction of both the spherical aberration as well as the spherochromatism. The all-diffractive surface with the quadratic and the fourth order fringe width variation has a residual of spherochromatism. The reason for this subtle difference is that in the aspheric case the spherical aberration correction and the chromatic aberration correction are totally separate from one another thereby allowing better performance. In the all-diffractive design we are more constrained and do not have sufficient variables to eliminate the spherochromatism as well as the spherical aberration and the primary axial color.

How Are Diffractives Fabricated?

Depending on the target diffraction efficiency and smallest feature size in the diffractive element, the optical engineer has the choice of a wide variety of fabrication technologies, ranging from simple grating ruling to complex gray scale optical lithography.

The flowchart in Fig. 13.25 summarizes the various fabrication technologies as they have appeared chronologically.²²

For a single optical functionality, for example, a spherical lens, the optical engineer can decide to use many different fabrication technologies which have their specific advantages and limitations, mainly in terms of diffraction efficiency. Figure 13.26 shows six different physical implementations of the same diffractive lens with their respective fabrication technology and their respective diffraction efficiencies (both theoretical and practical).

This invites us directly to the definition of diffraction efficiency, one of the most critical criteria when it comes to fabricate a diffractive element.

Holographic Exposure

As discussed in the previous section, holographic exposure of HOEs can yield a wide variety of optical functionalities, either as a thin surface relief HOEs (by the use of photo resist spun over a substrate) or as a

Figure 13.25
Chronological Fabrication Techniques and Technologies for Diffractives

<u>Historical overview</u>		
1785	Rittenhouse	- Grating made of hairs
1882	Rowland	- Ruling engines
1910	Wood	
1948	Gabor	- Holography
1962	Leith/Upatnieks	- Holography Optical Elements (HOEs)
1969	Lohmann/Brown	- Computer Generated Holography (photoreduction)
1972	D'Auria/Huignard	- Photo-refractive materials
1975	-	- CNC precision diamond ruling/turning
1983	Gale/Knop	- Direct analog laser write in resist
1985	Arnold	- Direct binary e-beam write
1989	Swanson/Weldkamp	- Multilevel optical lithography
1992	Lee/Daschner	- Direct analog e-beam write
1993	Potomac	- Direct analog Excimer laser ablation
1996	Canyon materials	- Gray scale masking lithography
2000	Digilens, SBG Labs	- Dynamic holography
2001	DOC, DS,	- Wafer scale hybrid integration
2003	-	- Dynamic digital diffractives (MEMs and others)

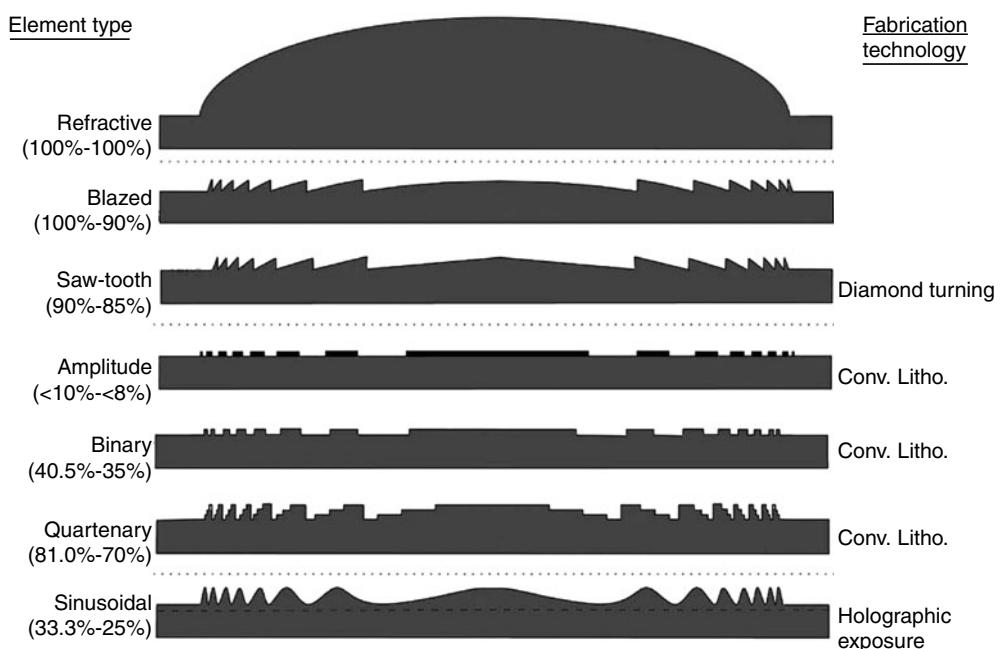


Figure 13.26
Six Different Physical Encoding Schemes for a Same Fresnel Lens

Diffractive Optics

more complex Bragg grating structures in a volume hologram emulsion. Figure 13.27 shows an example of holographic recording of a simple transmission HOE lens by using two source points at a specific wavelength required by the material and playing back the hologram at another wavelength, dictated by the application.

More complex optical functionalities can be recorded by using a master diffractive element like a DOE or a binary or multilevel CGH. In this double recording process, a single diffraction order is used from the CGH (which diffracts multiple propagating orders) to generate the object beam, all other diffraction orders present are blanked out. Figure 13.28 shows such a CGH/HOE recording process. Although the CGH generates many orders, the resulting HOE will generate only a single diffraction order without altering the initial optical functionality (if the recording has been done properly in the Bragg regime angles).

Table 13.2 shows the various holographic materials used in industry and their specifications.

Although a very versatile and efficient method, holographic recording of volume HOEs cannot be used in many cases, where materials with long lifetime are necessary (due to temperature swings, vibrations, humidity, or daylight/UV exposure), or simply when mass replication of cheap diffractive elements are necessary.

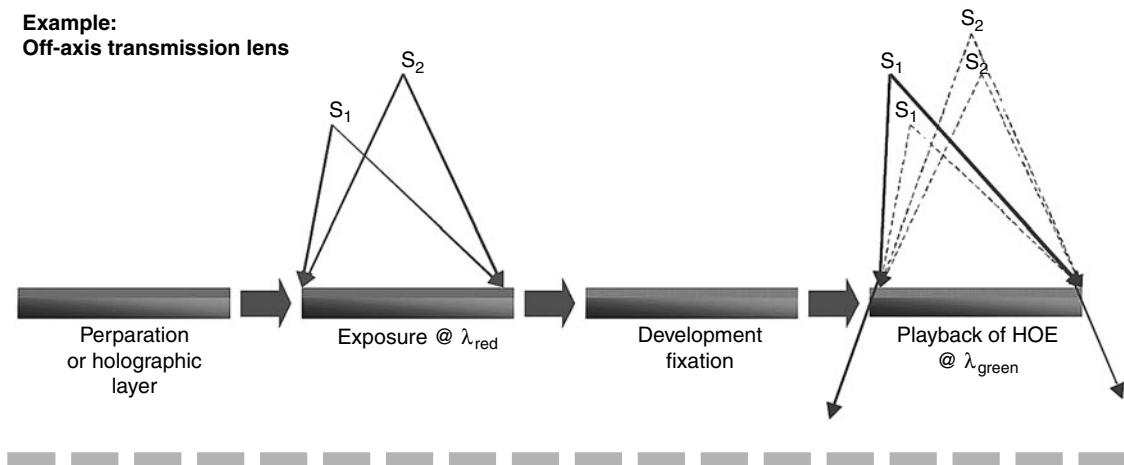


Figure 13.27
Standard Holographic Recording of an HOE Lens in a Holographic Emulsion

Figure 13.28
Surface Relief CGH
Recording Process
into a Volume HOE

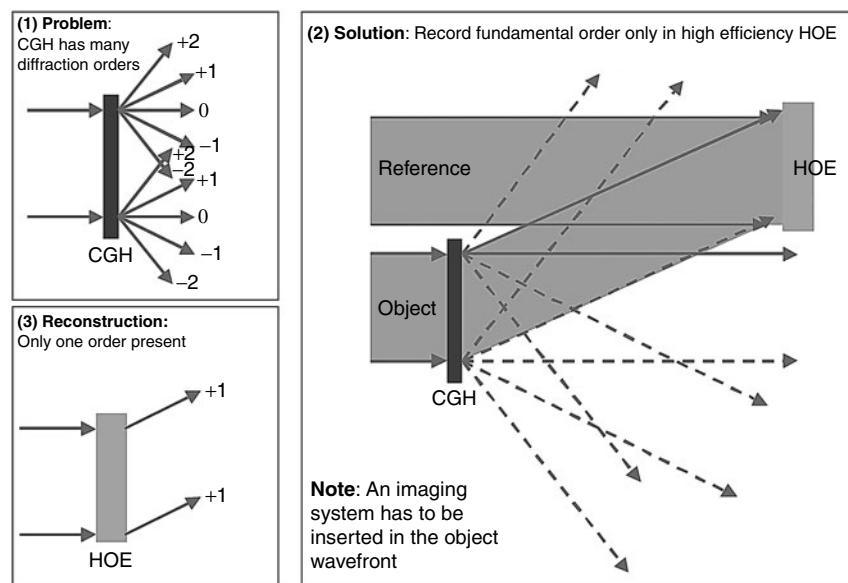


TABLE 13.2 Holographic Materials for the Recording of HOEs

Material	Hologram Type	Application	Advantage	Inconvenient
Photo-resist	Thin (surface relief)	Diffractive optics / lithographic patterning	Simple process, resist pattern can be etched by RIE in substrate	Low efficiency due to resist modulation.
Silver halide	Volume	First holograms	Cheap	Thin, low selectivity
DCG (Dichromated gelatin)	Thick volume hologram	Selectivity applications (optical storage, DWDM demux)	Cheap relatively thick for strong angular and spectral selectivity	Long term stability
Photopolymer (DuPont type)	Thin hologram	Replication in mass of anticounterfeiting holograms	Replicable in mass	Low resolution unavailable outside DuPont
Photorefractive crystal	Very thick volume hologram	Laboratory experiments	Erasable/rewritable Very high angular and spectral selectivity	Expensive—polarization issues
AO (acousto-optic) modules	Volume Bragg grating	Display	Fast reconfigurable	Only linear gratings, need piezoelectric transducers
PDLC (polymer dispersed liquid crystals)	Volume hologram	Display, telecom, datacom, storage	High index, high Δn , switchable through gray levels, fast	Long-term stability—polarization issues

Diamond Ruling/Turning

Diamond turning or diamond ruling via a computer controlled high-resolution CNC ruler or lathe can be a very effective way to produce high quality-efficient diffractives in a wide range of materials (plastics, glass, metals, ZnSe, Ge, etc.), see Fig. 13.29. However, there are limited to either one-dimensional (linear gratings) or circularly symmetric elements like on-axis symmetric lenses, with smallest fringes several times greater than the diamond tool. In the triangular or sawtooth grating example, the diamond tool can have the exact shape of the grating groove.

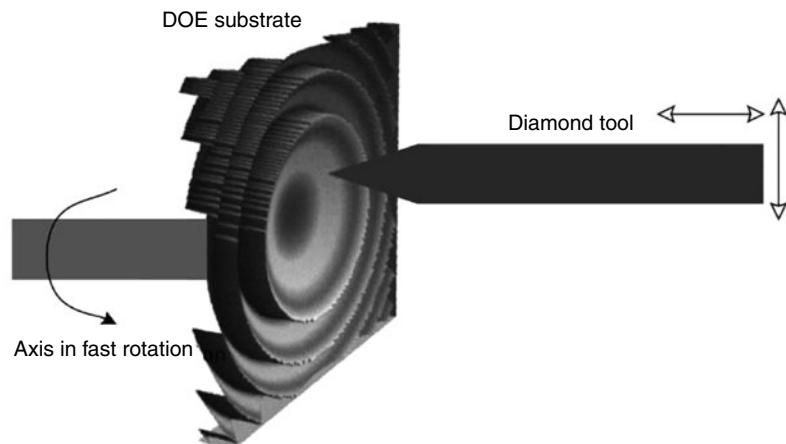
Note that more complex 6-axis CNC lathes can generate complex anamorphic fringes and profiles, as it is also done for their refractive counterparts.

Optical Microlithography

Optical microlithography fabrication technologies and techniques are derived from the standard integrated circuit (IC) industry, and have been applied to the fabrication of diffractives since the mid-1980s.²³

These techniques are used to fabricate diffractive optics when either the following conditions cannot be met by the previously described fabrication technologies (diamond turning and holographic exposure)²⁴:

Figure 13.29
Diamond Turning of
a Spherical Blazed
Fresnel Lens

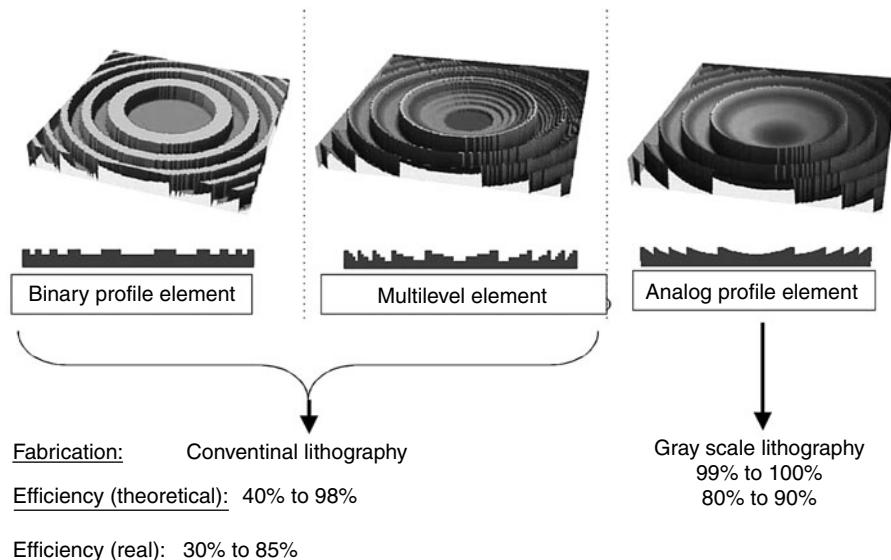


1. Fabrication of nonsymmetric or nonlinear features (CGH-like features)
2. Complex three-dimensional structures which cannot result from holographic interference
3. Precise spatial multiplexing of various diffractives (planar optical bench, arrays)
4. Integration of precise digital micro structures with additional alignment or integration features (fiducials, marks, targets, etc.)
5. Use of durable materials like quartz, glass, silicon, sapphire, ZnSe, Ge, etc.
6. Need for mass replication of the master element in these same materials

Microlithographic fabrication techniques can result in a wide variety of microstructured phase relief elements as depicted in Fig. 13.30 (from binary to multilevel to quasi analog surface relief).

In many cases, the diffraction efficiency is the most important criteria to consider, and therefore a multilevel or quasi-analog surface relief is often desired rather than a simple binary element, but this would also require a high fabrication budget.

Figure 13.30
The Various Microlithographic Techniques used for Diffractives Fabrication



Diffractive Optics

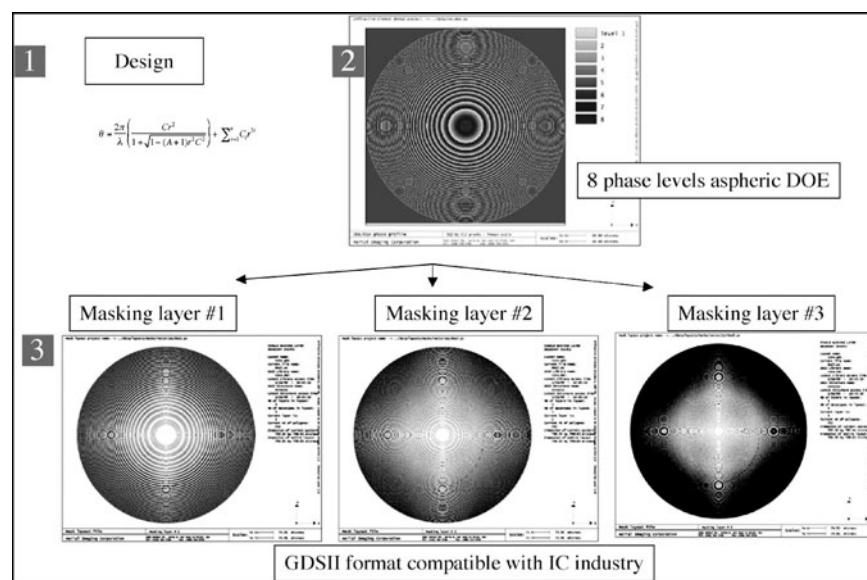
The microlithography fabrication process can be quite cumbersome when the optical designer does not have the adequate CAD tools to generate the required mask layouts.

The lithographic fabrication process can be split into five major tasks:

1. Generation of the phase profile (via analytic or numeric techniques)
2. Fracture of the resulting fringes into GDS2 polygons (IC industry standard format for mask layout)
3. Photomask patterning via laser or e-beam patterning system
4. Optical lithography, etching, and dicing
5. Potential replication of master elements by embossing or injection molding

Figure 13.31 shows the first two processes (generation of the layout) for an eight-phase levels diffractive lens for which three successive binary masks are generated and which fringes are fractures into polygons. These three masks are then used in multilevel lithography to generate the eight-phase levels.

Figure 13.31
Generation of the
GDS2 Layouts (Suc-
cessive Masks) from
the Initial Lens
Description



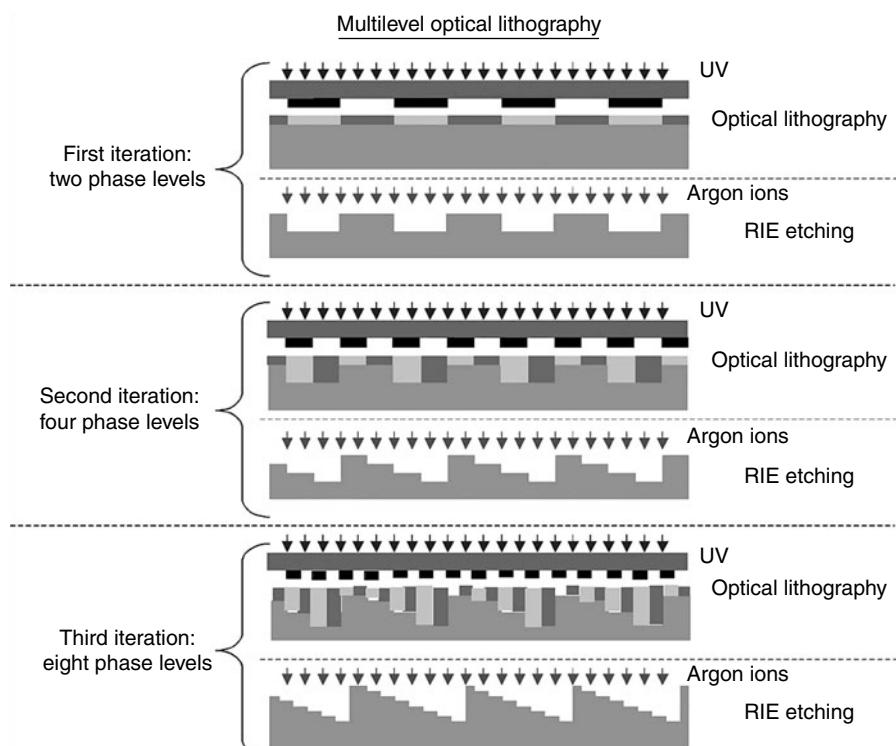
Multilevel lithographic fabrication is now becoming the most widely used diffractive fabrication technique. Although it is not the ideal one, it is the best understood and a widely available technique.^{22,23} Figure 13.32 shows this process, where 2^N phase relief levels can be fabricated with a set of N binary amplitude photomasks.

There are two different ways to use optical lithography: by contact or projection. Table 13.3 shows the specifications of each lithographic technique, their advantages and limitations. Both are now widely used for diffractive optics fabrication.

EXAMPLE OF MULTILEVEL LITHOGRAPHY FABRICATION

Figure 13.33 show the two successive steps involved in the fabrication of a four-phase levels Fresnel lens (first photo after the first lithography/etching step and second photo after the second lithography/etching step).

Figure 13.32
Conventional Multi-level Lithographic Fabrication Technique



Diffractive Optics

TABLE 13.3

Contact and Projection Lithography for Multilevel Diffractives Fabrication

Parameter	Contact Lithography	Projection Lithography
Lithography tool	Mask aligner	Stepper—Stepper/scanner
Demagnification factor	1×	1× to 10× (mostly 4×)
Minimum feature size printed on wafer	1.0 μm	0.85 μm (h-line) 0.35 μm (i-line) 0.18 μm (deep UV) <0.1 μm (extreme UV or X ray lithography)
Max field size on wafer (for individual element)	Size of the wafer — 0.5"	<30 × 30mm
Minimum lateral field misalignment	Approx 0.5 μm	<0.25 μm
Cost of machine	Medium (250K)	Very high (>5 M)
Costs per batch	Low	High
Minimum batch	Individual wafers	Cassettes batch (25 wafers)
Wafer size versus mask (reticle) size	2"/3" 3"/4" 4"/5" 5"/6"	4"/5" 6"/7" 8"/9" 12"/13"

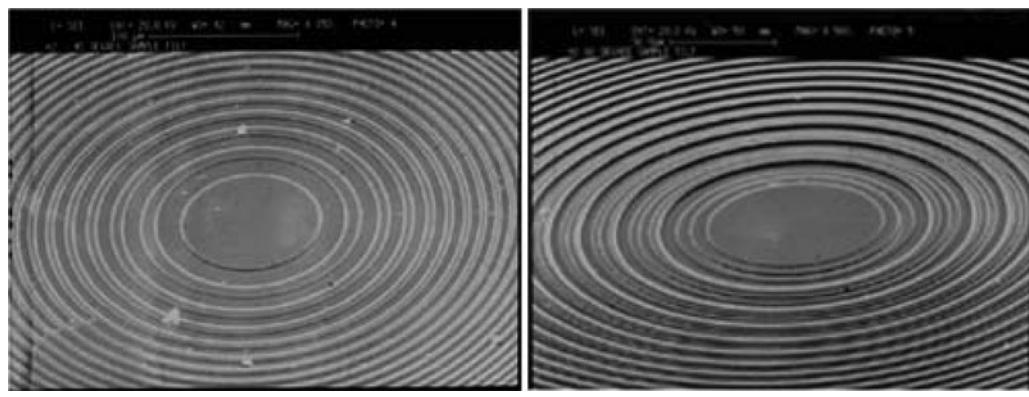


Figure 13.33

Fabrication of a Four-Phase Levels Fresnel Lens with Two Successive Masking Layers

SUCCESSIVE MASK ALIGNMENTS IN MULTILEVEL OPTICAL LITHOGRAPHY FABRICATION In order to align properly the two fields (photomask layers), alignment marks, and alignment fiducials are necessary. The accuracy of the alignment is crucial since any misalignment would decrease the efficiency and would create parasitic diffraction effects²⁵ due to high frequency structures arising from the misregistrations between successive layers (see Fig. 13.34).

Note the spikes generated by the successive field misalignments, which are typically less than $0.5 \mu\text{m}$ with a contact masking lithography process. The lateral structure sizes are of about $2 \mu\text{m}$. The alignment

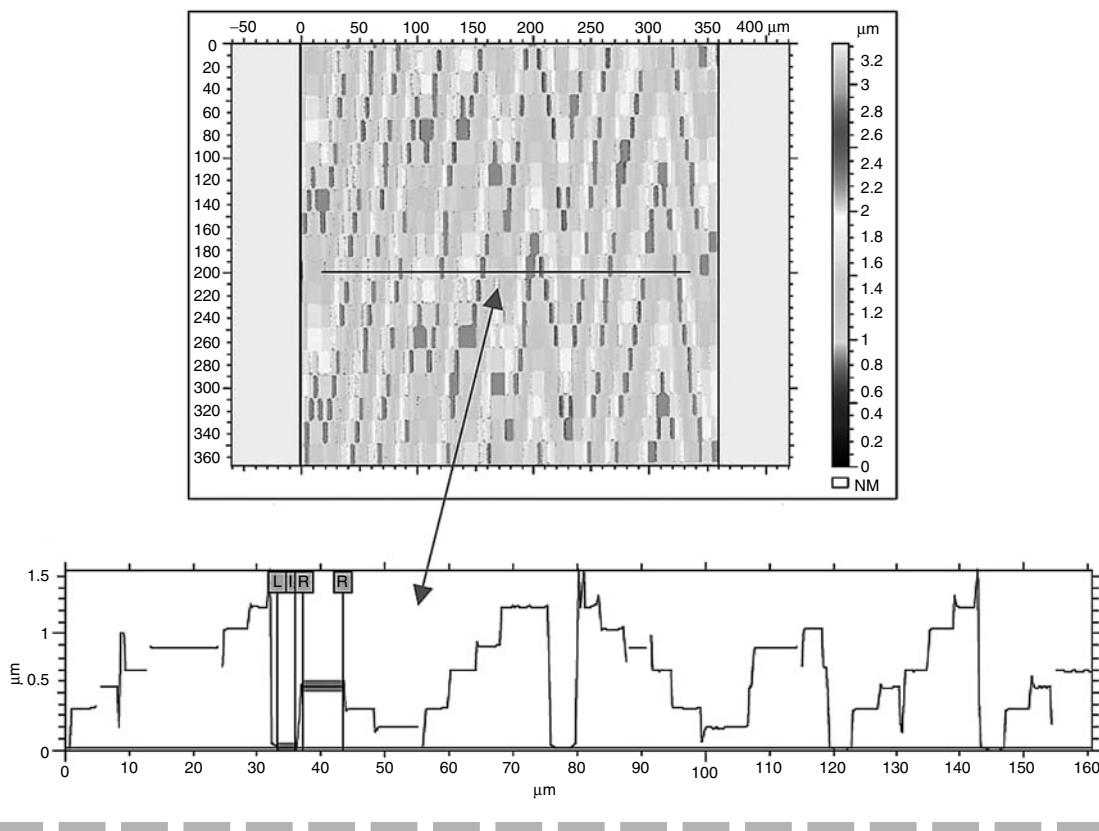


Figure 13.34

Lateral Misregistrations Detected via White Light Interferometry over a Eight-Phase Levels Beam Shaper CGH (Three Successive Masking Layers)

errors here are thus smaller than $0.25 \mu\text{m}$, which nevertheless create high frequency parasitic structures and other artifacts which can decrease the overall performance of the diffractive element.

Figure 13.35 shows a microscope photo of a Fourier type and a Fresnel type CGH fabricated through binary lithography with a single lithography/etching step in quartz.

Note that in the Fourier type CGH no fringe geometry can be seen, as expected and discussed in the previous section. On the other hand, the Fresnel type CGH shows fringe-like features, which can be modeled by the LGA method described in the modeling section.

DIFFRACTION EFFICIENCY CALCULATIONS FOR MULTI-LEVEL LITHOGRAPHIC FABRICATION Scalar theory of diffraction can be used effectively to predict the diffraction efficiency of binary or multilevel amplitude or phase elements fabricated by microlithography, in the realm of validity of scalar theory (see also Fig. 13.20).

Amplitude Gratings Efficiency Calculation Based on the Fraunhofer formulation of the diffracted field far away from the diffractive aperture (see modeling section^{3,9}), we will derive here the diffraction efficiency formulation for amplitude gratings (Fig. 13.36).

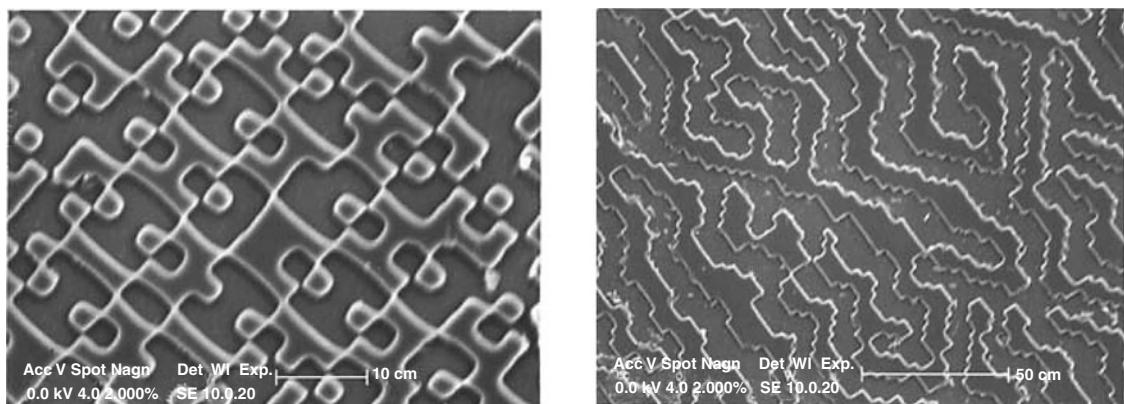
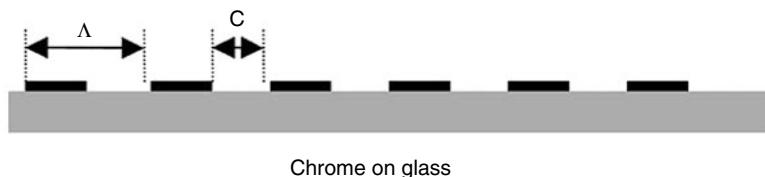


Figure 13.35

Fourier and Fresnel CGH Fabricated in Quartz Substrate as Binary Elements

Figure 13.36
Chrome on Glass
Binary Linear Ampli-
tude Grating



The amplitude of the grating function can be expressed as a Fourier expansion (Eqs. 13.16 and 13.17):

$$a(x,y) = \begin{cases} 1 & \text{if } m\Lambda \leq x \leq c + m\Lambda \\ 0 & \text{elsewhere} \end{cases} \quad (13.16)$$

$$a(x,y) = \frac{c}{\Lambda} \sum_{m=-\infty}^{+\infty} e^{-im(\frac{x}{\Lambda})} \frac{\sin(m\pi \frac{c}{\Lambda})}{m\pi \frac{c}{\Lambda}} e^{2im\pi \frac{x}{\Lambda}} \quad (13.17)$$

where each exponential term represents a plane wave, and m the index of the different diffraction orders present.

The magnitude of this m th diffraction order can be expressed as:

$$\eta_m = \left| \frac{\sin(m\pi \frac{c}{\Lambda})}{m\pi} \right|^2 \quad (13.18)$$

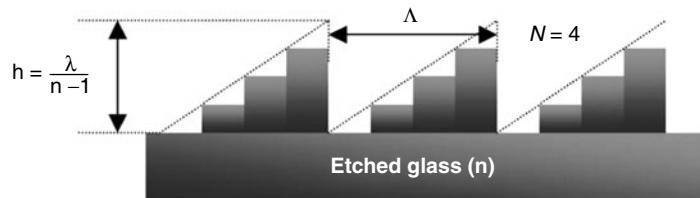
For regularly spaced opaque/transparent lines, (that is, a grating with a duty cycle of 50%), the diffraction efficiency formulation in the m th order becomes (Eq. 13.19):

$$\eta_m = \left| \frac{\sin(\frac{m\pi}{2})}{m\pi} \right|^2 \quad (13.19)$$

Multilevel Phase Gratings Efficiency Calculations Similarly to amplitude gratings, we can derive the diffraction efficiency for multilevel phase relief gratings (see Fig. 13.37).

If the grating is etched so that the maximum optical path difference yields 2π in transmission, which maximizes the efficiency (see introductory

Figure 13.37
Quadratic Phase Surface Relief Grating



section of this chapter), we get the following expression for the diffraction efficiency in the m th order for N phase levels (Eq. 13.20):

$$\eta_m^N = \left| \frac{\sin\left(\frac{m\pi}{N}\right)}{\frac{m\pi}{N}} \right|^2 \quad (13.20)$$

See also Table 13.4 in next section for the predicted values of diffraction efficiencies for multilevel phase relief diffractives.

We can easily extrapolate these results from infinitely long linear gratings to more complex structures limited in space, like Fresnel lenses and general CGHs (which can be approximated locally by such linear gratings).

When the number of phase level grows to 16 or more, the diffractive element can be easily considered as a quasi-analog surface relief element. The diffraction efficiency of an analog surface relief element (for $m = 1$ and $n \rightarrow \infty$) is shown in Fig. 13.38.

In fact, it does not make sense to fabricate a diffractive element with more than 16 levels by conventional multilevel masking techniques, since the successive systematic lateral misalignment errors and cascaded etching depth errors would reduce the diffraction efficiency dramatically. The diffraction efficiency increase when growing from 16 levels to more levels (32, 64, 128, or 256), even for perfect fabrication (which is an impossible task), is infinitesimal (a fraction of a percent), but the fabrication effort is enormous and does not make sense practically.

Figure 13.39 shows the diffraction efficiency of a blazed grating (or lens) for the diffraction orders +1 (fundamental), +2 and +3, as a function of the reconstruction wavelength.

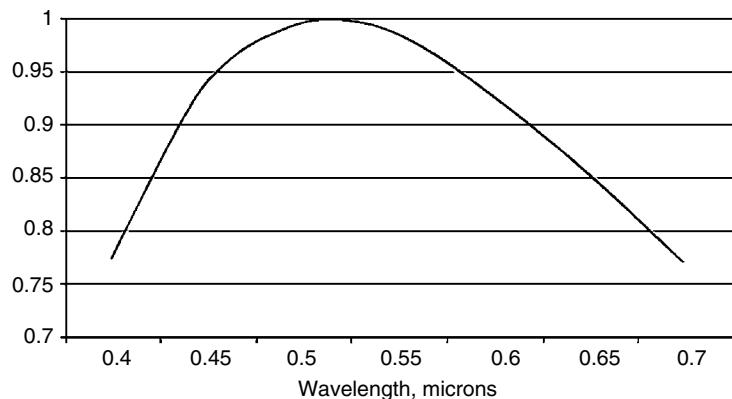
The blazed lens in Fig. 13.39 is designed and fabricated to yield maximum efficiency in the fundamental order in the green region of the spectrum (550 nm). As the efficiency in the fundamental drops when going from green to lower wavelengths, the efficiency in the higher orders increases.

TABLE 13.4 The Various Lithographic Fabrication Technologies and Their Characteristics

Microlithographic Technique	Nb of Surface Levels	η (%)	Nb of Real Mask	Litho/		Advantages	Limitations
				Etching Processes	Fabrication Costs		
Binary amplitude	Binary	<8%	<8%	1	0/0	Minimal (only mask)	Cost and resolution
Binary phase in resist	Binary	40.5%	\approx 30%	1	0/0	Minimal (only mask exposure)	Only one element, very low efficiency
Direct analog write (laser or e-beam)	Quasi analog	>80%	\approx 80%	1	0/0	Medium, requires dosage modulation	Only one element, profile in resist, difficult to etch
Binary phase in mask substrate	Binary	40.5%	\approx 35%	1	0/1	Small (only mask and RIE)	Cost and resolution, stable material
Multilevel (N masks for 2^N levels)	4	81%	\approx 75%	2	2/2	Medium	Efficiency Many elements
	8	95%	\approx 85%	3	3/3	High	High efficiency Many elements
	16	99%	\approx 90%	4	4/4	Very high	High efficiency Many elements
	>16	>99%	\approx 90%	>4	>4/4	Very high	High efficiency Many elements
Grey scale lithography	Analog	>90%	\approx 90%	1	1/1	Very high	High efficiency Many elements No misalignment Costs

Diffractive Optics

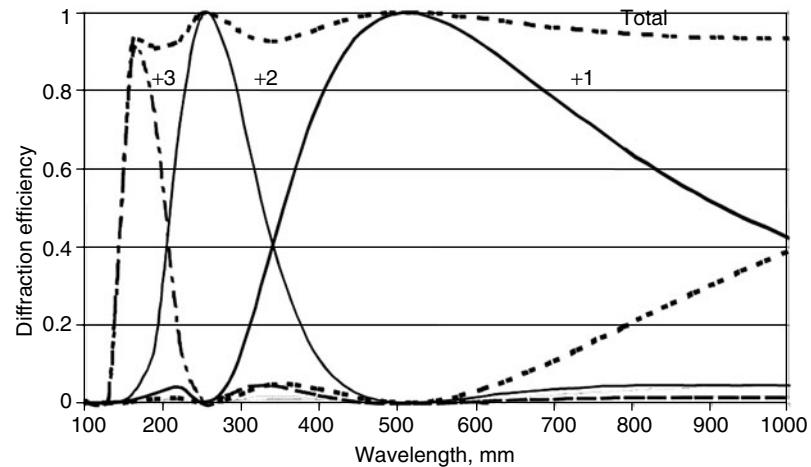
Figure 13.38
Diffraction Efficiency in the Fundamental Order for a Quasi Analog Surface Relief Profile (Lens or Grating). The Optimum Wavelength (Design Wavelength) Is 510 nm

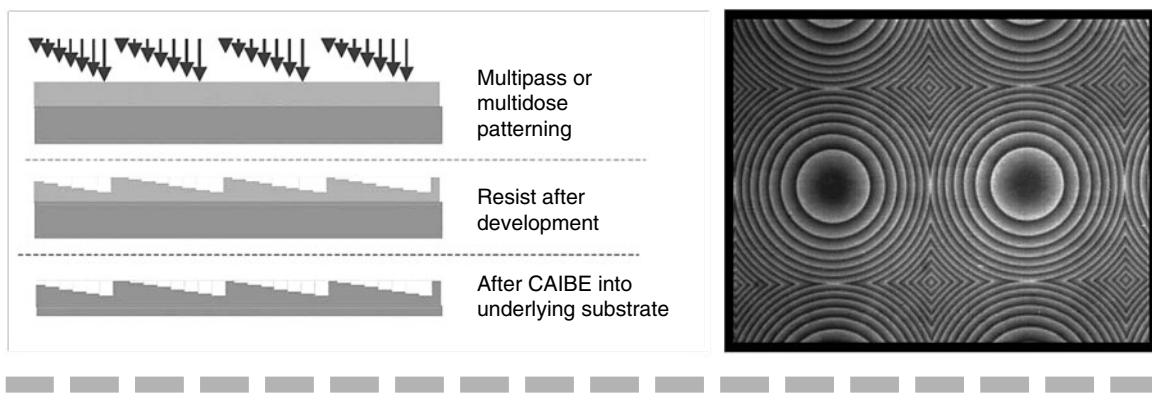


Direct Write Techniques In order to avoid the systematic lateral field misregistration errors occurring when using conventional multilevel lithography, direct write methods have been reported.²⁶ These methods write directly into a resist layer a quasi-continuous phase profile, by either with e-beam or laser beam dosage modulation or multipass patterning (see Fig. 13.40).

GRAY SCALE OPTICAL LITHOGRAPHY Several gray scale lithographic fabrication techniques have been reported since the mid 1990s (HEBS glass—high energy beam sensitive glass or inorganic resist with dry development). They have in common the fabrication of a gray scale mask or reticule used in projection lithography in order to expose

Figure 13.39
Diffraction Efficiency versus Wavelength for the Three Diffraction Orders +1, +2, and +3 for a Blazed Grating



**Figure 13.40**

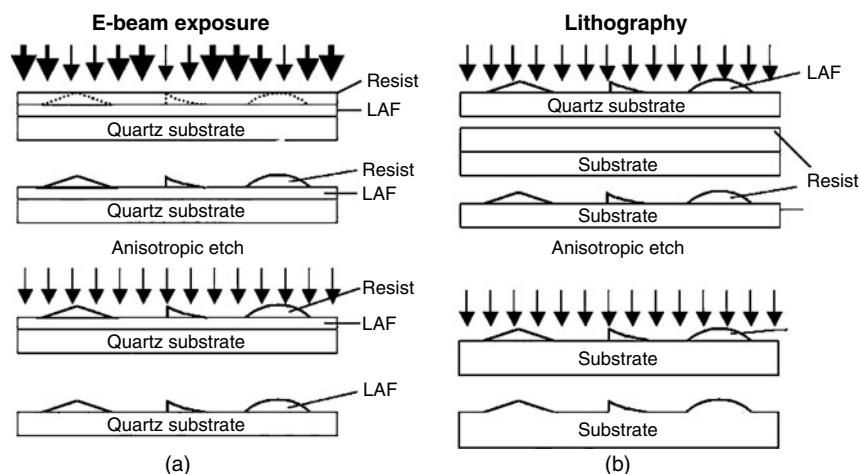
Direct Write Technique and Resulting 256 Levels Fresnel Lens Array Etched into Quartz

directly a photo resist in an analog way, and thus yield an analog surface relief modulation.

Figure 13.41 illustrates the schematic fabrication process of gray-scale technology utilizing a new carbon-based mask material and improved lithographic/etching processes.²⁷

The transmittance of the mask is related to the thickness d of their light attenuating film (LAF) according to $\exp[-\kappa d]$, where κ is the absorption coefficient of the LAF. When the gray scale mask is used in lithography and dry etching processes with special attentions paid to preserving the analog features in the mask, many gray scale device

Figure 13.41
 (a) Fabrication Process of a New Carbon-Based Gray Scale Mask with Conventional E-Beam Resist
 (b) Fabrication of Gray Scale Device Structures Using the Gray Scale Mask, an Analog Resists and an ICP-RIE System (Courtesy of Dr Sing Lee and Dr Jo Zhou, UCSD)



Diffractive Optics

structures can be fabricated without going through the cumbersome lithography and etching processes involving multiple masks. Figure 13.38 shows an AFM picture of an off-axis diffractive lens with a height of about 1 μm and periodicities ranging from 10 to 12 μm , fabricated in quartz by employing the new gray scale mask and 248-nm-stepper lithography.²⁸

Figure 13.42 shows a microscopic picture of a corner cube (with a depth of about 10 μm at its apex) array, fabricated in silicon by the use of the gray scale mask and an i-line aligner.

The gray scale structures have surface smoothness corresponding to 64 thickness or phase levels in Fig. 13.42 (left) and 256 levels in Fig. 13.42 (right).

There have also been pseudo gray scale lithography techniques reported in literature, which use simple binary masks to encode gray levels through either pulse width or pulse density modulation of opening in a chrome layer (see Fig. 13.43).

The gray scale intensity mapping is created when light is diffracted through these small openings on the photomask and then reimaged onto the wafer in the resist layer.

SUMMARY OF FABRICATION TECHNIQUES Table 13.4 shows the various lithographic fabrication technologies used in industry, the resulting typical diffraction efficiencies, as well as their respective advantages and limitations.

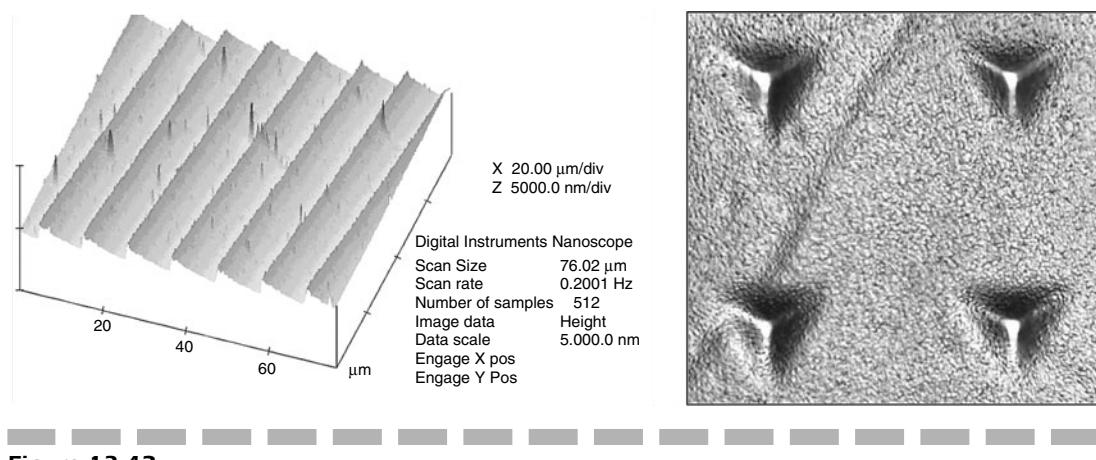
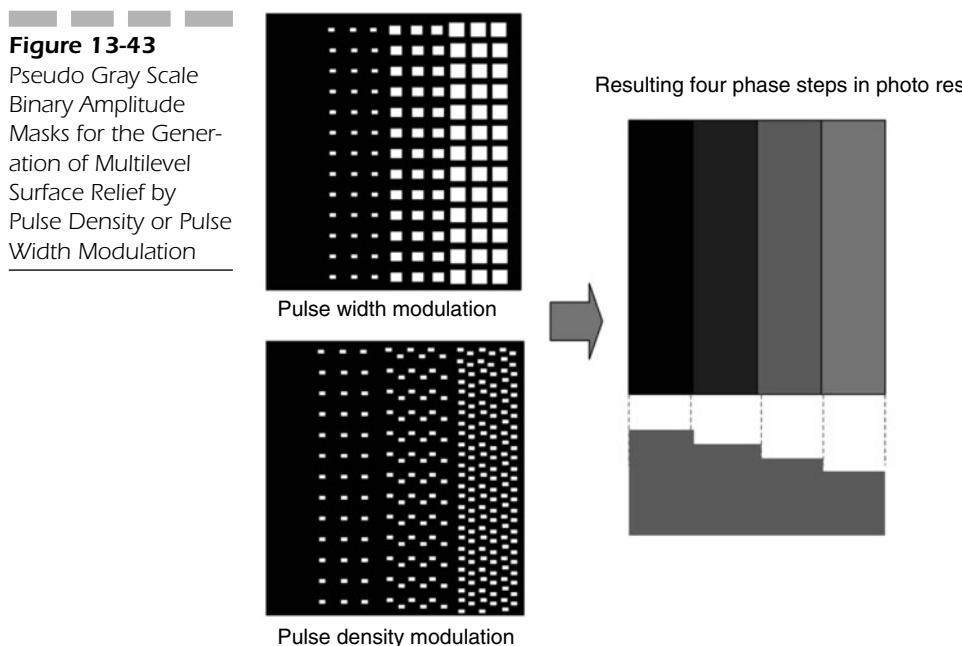


Figure 13.42

Gray Scale Lithography Process—Off-Axis Diffractive Lens on Quartz (Left) and Array of Microcorner Cubes (Right)

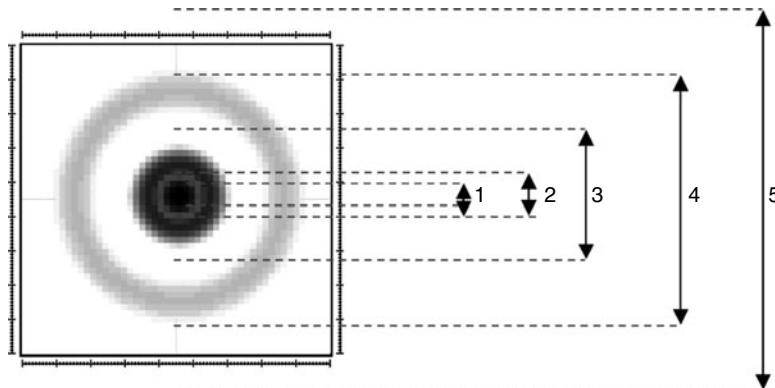


IMPORTANT REMARKS: NOTION OF DIFFRACTION EFFICIENCY FOR INDUSTRIAL APPLICATIONS It is important to remember that diffraction efficiency values are theoretical values for the efficiency in the m th diffraction order (for example, the fundamental positive order where $m = 1$). However, for a given application, the diffraction efficiency is actually very often defined as a fraction of this value or even as a combination of several diffraction orders. In order to illustrate this concept, we show below two examples where the efficiency linked to an application is respectively lower and higher than the theoretical efficiencies as calculated here before.

a. Analog surface relief Fresnel lens for data storage.

In this case, we are not interested in any imaging task from the lens, but rather in the amount of energy within the waist of the focal spot (see Fig. 13.44). The theoretical efficiency relates to the amount of energy in this spot (within the Airy disk), but also the energy within the successive side lobes. Therefore, the efficiency linked to this application is lower than the predicted efficiency of

Figure 13.44
Definition of
Application-Oriented
Diffraction Efficiency



the element. The figure shows the focal spot and five different definitions of efficiency in this case (1: FWHM, 2: waist, 3: Airy disk, 4: First side lobe, 5: etc.).

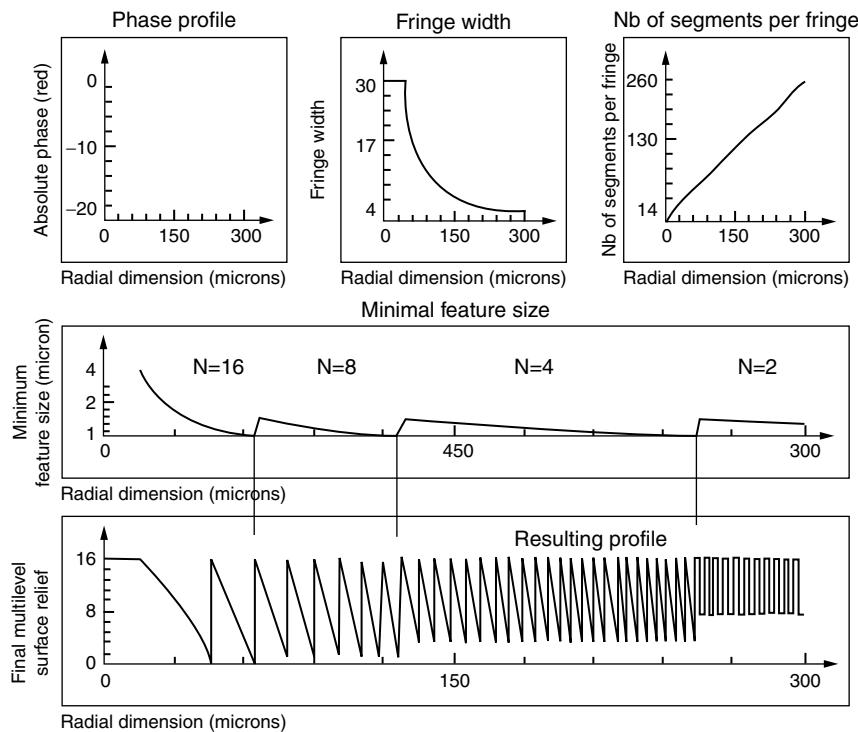
b. Binary Fourier CGH with on-axis symmetric reconstruction.

In this case, we are taking advantage of the fact that in the Fourier plane of such an element, two main reconstructions appear on-axis with the same efficiency (about 40%). These are the two conjugate fundamental orders (orders +1 and -1). If the reconstruction is symmetric in the Fourier plane, these two reconstructions can overlap exactly and double the theoretical efficiency from 40.5 to 81%.

This technique is often used for the binary (that is, relatively cheap) fabrication of high-efficiency on-axis spot array generators, logo generators, beam shapers, etc.

EFFICIENCY CALCULATIONS OF MULTILEVEL FRESNEL LENSES IN REAL WORLD In many cases, the microlithographic fabrication limitation is the lateral size of the smallest structure (note that a similar limitation also applies to the validity of the scalar theory of diffraction). As the fringe width (local period) of a Fresnel lens decreases radially, the maximum number of phase levels approximating the analog fringe relief surface also decrease. Therefore, a smart way to fabricate an efficient lens considering the limitation of the fabrication tool is to optimize the number of phase levels over each fringe. In such a lens, the efficiency is maximal in the center, and decreases radially according to the number of phase levels present (see Fig. 13.45).

Figure 13.45
Optimum Fabrication
of a Multilevel Fresnel
Lens for a Specific
Fabrication
Resolution



FINAL WAFER DICING Finally, AR coating and wafer dicing concludes usually the fabrication of diffractives through microlithography. Figure 13.46 shows a 4-in. quartz wafer AR coated and diced into individual rectangular dies (still held on sticky blue tape) and ready to be inserted in their respective products.

Mass Replication of Diffractives

The fabrication of diffractive optics makes sense in an industrial realm only if these elements can be cheaply replicated in mass. This is the *De Facto* condition for diffractive optics to emerge from the lab and address industrial needs in consumer electronics products and other application fields like automotive, factory automation, biomedical, and telecom, etc.

In the previous sections, we have emphasized the generation of master elements, either by holographic recording, diamond machining, or by microlithography.

Figure 13.46
Diced and AR Coated Quartz Wafer



There are several methods available to the optical engineer to replicate these master elements in volume. These include roll embossing, hot embossing, UV casting, injection molding, and Sol-Gel process (see Fig. 13.47).

Mass replication of a positive master element includes three successive processes:

- Generation of positive master through microlithography, diamond turning, etc.
- Nickel electroplating of negative master (negative structures on Nickel film)
- Generation of a final mold and replication (embossing or injection molding)

Figure 13.47
Mass Replication
Process by Emboss-
ing or Injection
Molding

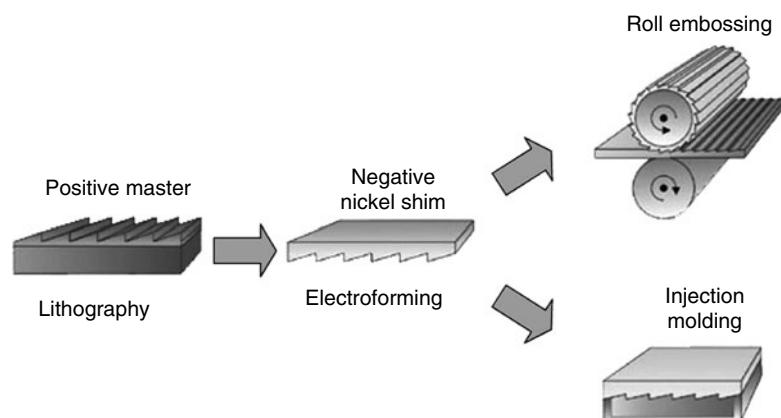


Figure 13.48 shows photos of diffractive optics replicated by CD-ROM injection molding, which is a very effective way to produce in mass diffractives on a polycarbonate substrate, the same thickness and size as a standard CD-ROM.

Table 13.5 summarizes the main diffractive optics replication technologies used in industry today.

Such replication technologies, which for most of them have been developed and optimized for other technologies like CD and DVD disk replication, provide an invaluable asset to diffractive optics to finally come out of the academic/research arena, and propose concrete industrial solutions to real market needs in many consumer electronics, factory automation and telecom applications, as we will see in the next section.

Where Are Diffractives Used?

Diffractive optics is used in numerous applications today, ranging from consumer electronics, sensors, factory automation, to telecom and biomedical applications.

Spectroscopic Applications

Historically, spectroscopy has been the first great application realm for diffractives, and more precisely for linear ruled gratings (see also Figure 13.7 in first section of this chapter). For many people, including many optical

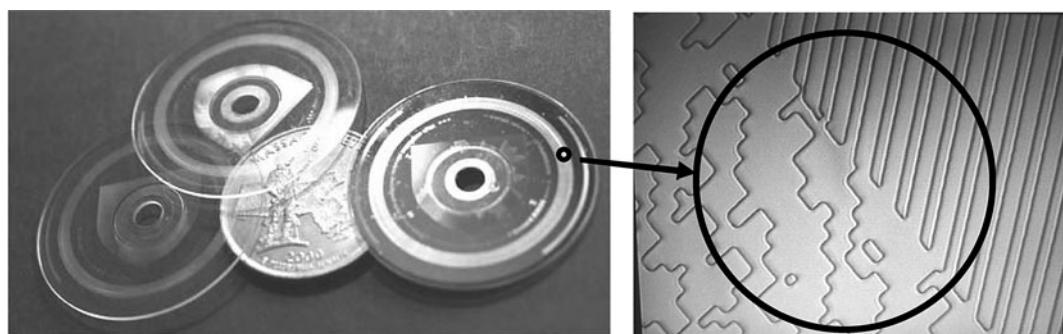


Figure 13.48

Replication of Diffractive Encoder Structures by CD Injection Molding

TABLE 13.5 Summary of the Major Diffractive Optics Replication Technologies Used in Industry Today

Replication Technology	Element Type	Typical lot Size	Price per Replica	Resulting Substrate	Advantages	Limitations
Conventional optical lithography	Binary/ multilevel surface relief	25 wafer cassette lot	High (\$0.1K–\$1K)	Any wafer (quartz, fused silica, glass, silicon, etc.)	Very high precision via optical lithography and dry etching	Pricing, lot size, dicing
Grey scale lithography	Analog surface relief	Individual wafers	Very costly (>\$1K)	Selected wafers	Very high efficiency	Costs and availability
CD/DVD injection molding	Any surface relief	>1000s to Millions	Cheap <\$1	Polycarbonate	Rigid substrate, proven technology	Need for specialized equipment and Nickel shim generation
Planar hot embossing	Any surface relief	1000s	Cheap \$1–\$10	Any plastic	Rigid substrate no need for specialized equipment	Nickel shim generation needed
Roll embossing	Any surface relief	Millions	Very cheap <\$0.1	Mylar, thin plastics,...	Cheap proven technology	Substrate not rigid, Ni shim generation
UV casting	Any surface relief	1000s	Cheap \$0.1 to \$10	Polymer	No need for special equipment	No rigid substrate, shrinkage, tear, wear
Sol gel	Any surface relief	100s	Average \$10–\$100	Glass	Resulting element is silica based (glass)	Shrinkage and geometrical control

engineers spectroscopic gratings remain the one and only application of diffractive optics today, a notion that has to be revised as we will see in the other sections.

Imaging Applications

Incorporating a diffractive lens in an imaging system is usually considered difficult when dealing with broadband illumination. However, as seen in Fig. 13.11 in this chapter, diffractives can help reduce chromatic and thermal aberrations and reduce the overall number of lenses needed. Canon Inc. of Japan (Fig. 13.49) has introduced this concept in several telefocus objective lenses for their line of digital single-lens reflex (SLR) cameras. Canon uses a set of sandwich Fresnel lenses with different refractive indexes in order to reduce the overall size and weight of the objective. Not only is the resulting objective shorter, it also incorporates lesser lenses.

In a general way, it is much simpler to design a hybrid achromat for single wavelength, for example, infrared cameras for automotive three-dimensional sensing in bumpers, camera lenses for laser-based factory automation sensors, etc.

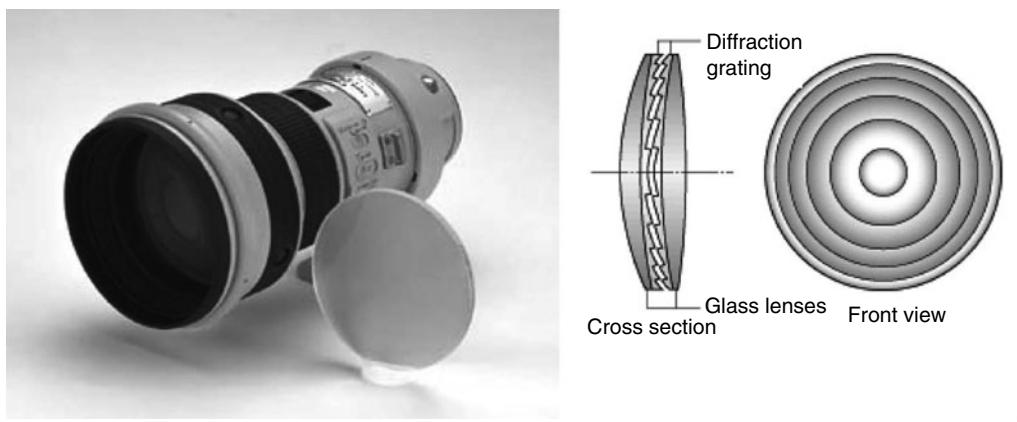


Figure 13.49

Canon Ltd. Diffractive Tele-Objective Lenses

Optical Computing and Optical Interconnections

Optical computing has become, through the 1990s a very hot topic of research and development, and has extensively used diffractive optics for both optical clock distribution²⁹ in multichip modules (MCM) and optical interconnections in massively parallel computing architectures³⁰. However, this scheme has never made it to a consumer electronic reality, and remains a development area mainly for high-demanding military computing applications.

Optical Data Storage

Optical data storage has become the first consumer electronics application realm for diffractive optics. Actually, many optical functionalities in a standard OPU unit for CD/DVD drives incorporate diffractives today [beam splitting for tracking purposes and hybrid refractive/diffractive lenses for dual spot lens for CD/DVD read-out (see section "How Are Diffractives Fabricated?" for more insight into this hybrid lens architecture and for more recent all diffractive data storage applications see www.inphase-technologies.com]. Figure 13.50 shows fast diffractive lenses used in Winchester heads for magnetic optical disk read-out (left), and dual focus lens for CD/DVD read-out.

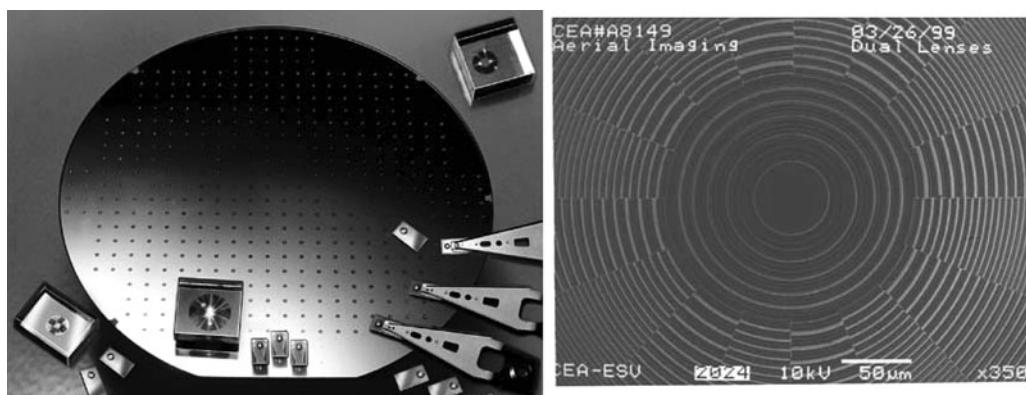


Figure 13.50

Fast Lenses for Magneto Optical Winchester Flying Heads and Dual Focus Lenses for CD/DVD Read-Out

Optical Telecom

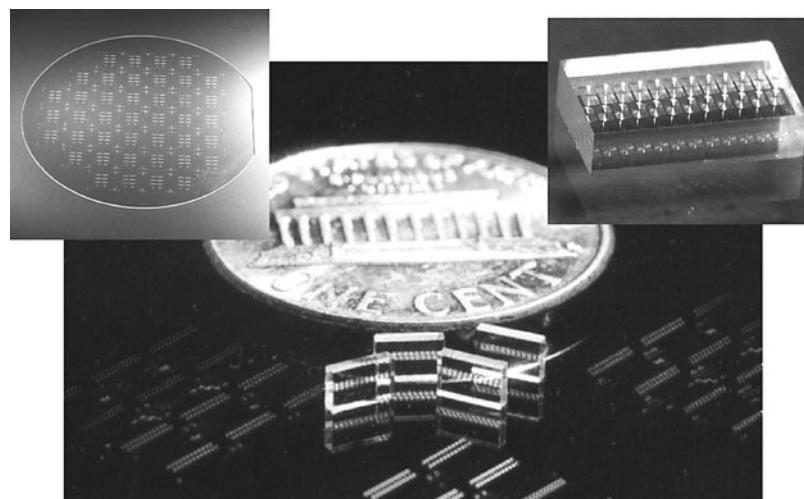
Optical telecom has seen massive surge of interest after the DWDM revolution at the end of the last decade, and more recently with the steady growth of 10-Gbs optical Ethernet lines. Diffractives have mainly been applied (as for spectroscopic gratings) to spectral Demux and Mux applications, as linear ruled gratings (www.jdsu.com) or phase gratings (www.bayspec.com). These Mux assemblies can then be applied to more complex functionalities like optical add-drop modules.

Another realm of application of diffractive in telecom is the introduction of Bragg gratings for very selective spectral filter implementation, both in Bragg reflection and Bragg coupling regimes.

More recently, diffractives have been applied to 10-Gbs optical Ethernet lines for fiber coupling and detector focusing, signal and other monitoring functionalities (www.doc.com). Figure 13.51 shows a 12 lines 10-Gbs optical Ethernet optical assembly block for 850-nm VCSEL laser to fiber coupling and fiber to detector using dual side substrate patterning.

In Fig. 13.51, the left part shows a 6-in. quartz wafer with many individual optical block assemblies, and the right part (as well as the centre parts) show a single diced out assembly ready to be integrated in a 12 array 10-Gbs Ethernet fibre bundle.

Figure 13.51
Multichannel 10Gbs
Ethernet Optical
Block Assembly for
850-nm VCSEL
Lasers/Detectors



Optical Anticounterfeiting

Holograms have been used since the 1970s to implement anticounterfeiting devices in products like credit cards, bank notes, passports, and other critical documents for authentication purposes. Recently, optical variable devices (OVDs), which are mainly spatially multiplexed gratings and lenses, provide additional security to such devices (www.ovd-kinegram.com). These OVDs and other synthetic holograms are for most of them not holographically generated, but rather generated by computer as series of DOEs or CGHs, mastered by conventional optical lithography and replicated by roll embossing. The current trend is to include in such visual OVDs additional machine readable (laser scanner readable) features which can include large amounts of digital data, much larger than the data that can be integrated in radiofrequency identification (RFID) devices. Of course, these new hybrid visual/machine readable diffractive security tags remain read-only, but are very cheap to replicate in mass by roll embossing.

Laser Material Processing (Cutting, Welding, Engraving)

Laser cutting, laser welding, and laser engraving are very desirable functionalities for the laser material processing industry, a market that is growing at a steady rate, since more than two decades. Laser beam shaping for high power CO₂ and YAG (yttrium aluminum garnet) lasers are now common elements, usually fabricated by diamond turning in ZnSe, ZnS, or Ge substrates.

More complex CGHs, usually fabricated as reflective elements etched in quartz with gold coatings, incorporating specific beam engineering functions (like complex logo engraving or accurate intensity redistribution for welding applications) show to be promising solutions for fast but accurate laser material processing tasks, without having any moving part in the optical path (no moving mirrors, shutters, etc.).

Industrial Optical Sensors Applications

Industrial optical sensors have been a large pool of application for diffractive optics. Gas sensors, position, displacement and motion sensors, strain, torque and force sensors, spectral sensors and index of refraction

sensors, Doppler sensors as well as diffractive microsensors in micro-electromechanical systems (MEMs) configuration have been reported in literature and industry (see for example www.microesys.com and www.appliedopto.com). Industrial optical sensors (especially motion encoders) are based on a large and steady market that is growing slowly and has not been affected by technological bubbles as the DWDM optical telecom, optical data storage, or biomedical markets.

Diffractive optics is being used increasingly in infrared image sensors for remote three-dimensional sensing in automotive and general factory automation. A CGH (type 3 element) is designed to project in the far field a set of grids through an IR laser diode. This scene is then acquired by a digital camera with an IR filter which analyzes the deformation of the IR lines and computes back the three-dimensional topology of the scene.

Biomedical Applications

More recently, biomedical applications have been showing great interest in diffractive optics, for specific applications in optical sensors, genomics, proteomics as well as individual cell processing.

Diffractive optics has been used to implement surface plasmon sensors and index sensors by using the Bragg coupling effect in integrated waveguides. Assay analysis in genomics is using CGHs as laser beam splitters in order to illuminate large numbers of individual arrays of samples and read out fluorescence measurements. Laser tweezers have used the specific advantages of dynamic diffractives in a microscope column in order to steer in real time individual cells in a two-dimensional pattern.

More generally, diffractives can be used in any biomedical apparatus that requires homogenization and/or shaping of the Gaussian laser beam into a specific intensity mapping (for example, fluorescence measurements in hematology by uniform laser illumination on blood sample).

Consumer Electronics Applications

When a technology makes it to standard consumer electronics devices, this technology enters the realm of mainstream technology. Although there has been several introductions of diffractive to consumer electronics as described in the previous optical data storage section, no other application has shown to the public the direct and real potential

of diffractives (and especially CGHs) more than the virtual keyboard application (www.canesta.com), the auto-focus pattern generator application (www.sony.co.jp), and of course the laser pointer pattern generator which has been sold in millions since its debuts in 1998 (see Fig. 13.52).

Projection Display Applications

The market expectations for LED and laser based projection displays, both in front (conventional projectors and pocket projectors) and in rear projection (RPTV) architectures, are very promising since current technologies like plasma screens, LCD and regular high pressure arc lamp projection engines are becoming obsolete, as both display screens get larger, projection engines get smaller and prices are falling.

Diffractive optics have many desirable facets which bring elegant solutions to many of the optical tasks involved in future projection display, especially as they will be laser based (vertical cavity surface emitting laser [VCSEL] et diodes). For example, CGHs can implement beam shaping, diffusing, homogenizing, steering and despeckeling, as well as direct far field image pattern generation functionalities. One of the first diffractive image pattern generators is presented in Fig. 13.53 (www.light-blueoptics.com, see also www.holoeye.com). Type 5 dynamic diffractive elements are used to implement such functionality (see also Fig. 13.6).

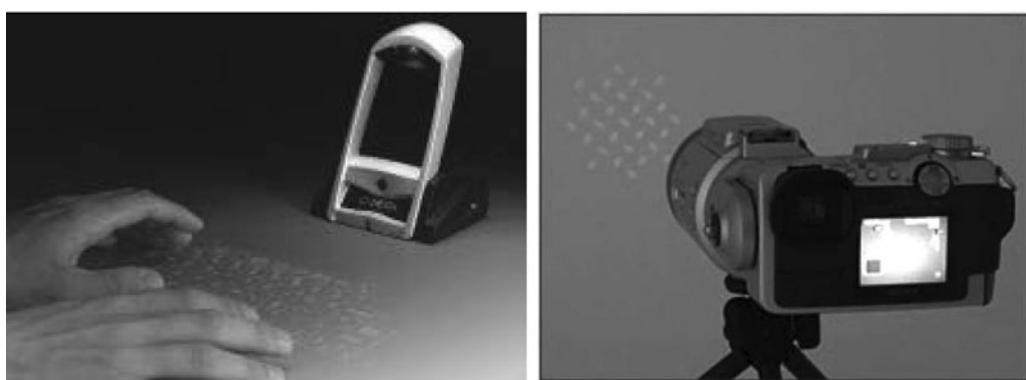


Figure 13.52

Diffractive Virtual Keyboard (Left) and Diffractive Focus Measurement Elements (Fourier Pattern Generators) for Digital Auto-Focus Camera (Right)

Figure 13.53

Diffractive Miniature Projector from Light Blue Optics



Since such a projector generates a diffracted image in the far field, there is no need for an objective lens and/or a zoom lens, and thus the size of the projector can remain very small. However, due to small diffraction angles generated by dynamical diffractives, an additional optical setup is usually necessary to increase these angles.

Niche Markets

Interests for diffractive optics for small niche markets have been demonstrated in the past years, and include complex illumination generation for deep UV stepper lithography for RET applications (reticle enhancement technology)—see www.doc.com.

Market Analysis and Future Applications

Technological waves and investment hypes have historically generated regained interests in diffractive optics technology during the last three decades. These successive technological waves can be described along seven main lines (see Fig. 13.54).

Diffractive Optics

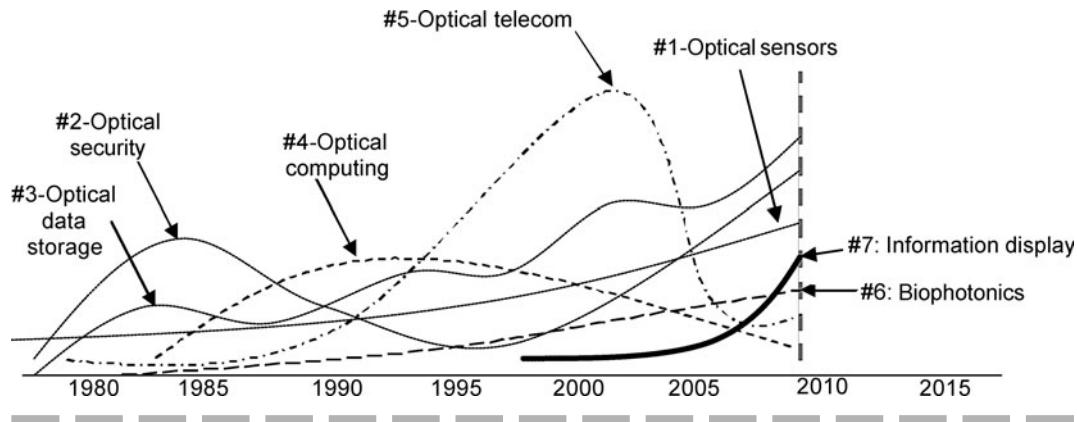


Figure 13.54

Optical Technological Market Waves that Have Fueled Interest in Diffractive Optics Technology over the Past 30 Years

- Optical security devices and optical data storage have been the first ones to generate interest in diffractives (other than spectroscopy applications). Optical data storage is actually generating successive interests in diffractives, from the 1980s for CD signal tracking, to the 1990s for DVD read-out heads, and to today's Blu Ray technology and holographic storage.
- Optical computing has generated a lot of interest in diffractives in the early 1990s, but has never been able to bring diffractives to mainstream market (see previous section on optical computing).
- The investment hype of the late 1990s in optical telecom and especially in DWDM Mux Demux devices has ill-fueled interest in diffractives, with the consequence of a disastrous bubble burst in 2003. Recent regain interest in 10Gbs optical Ethernet is correcting this market trajectory (see previous optical telecom section).
- Optical industrial sensors have shown the most steady growth during the past decades, and are still one of the most desirable and less risky markets for the implementation of diffractive optical technology.

- Biophotonics has shown much interest in diffractive very recently, and this interest is growing fast (see previous section on biomedical applications).
- The current interest in information display and especially in laser-based projection displays (rear and front projection engines), have the potential to build a large market and stable application pool for the use diffractive optics for the next decade.

References

1. M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon Press, Oxford, 1993.
2. A. W. Lohmann and D. Paris, Binary Fraunhofer Holograms generated by computer, *Appl. Opt.*, 6 (10), 1739–1748, 1967.
3. B. Kress and P. Meyrueis, *Digital Diffractive Optics*, 1st ed, John Wiley and Sons, Chichester, 1999.
4. G. J. Swanson, *Binary Optics Technology: Theoretical limits on the Diffraction Efficiency of Multilevel Diffractive Optical elements*, MIT Laboratory Report 854, MIT, Cambridge, MA, 1991.
5. D. Leseberg et al., Computer generated holograms: Cylindrical, conical and helical waves, *Appl. Opt.*, 26 (20), 4385–4390, 1987.
6. T. Stone et al., Hybrid diffractive/refractive lenses and achromats, *Appl. Opt.*, 27 (14), 2960–2971, 1988.
7. C. W. Londono et al., Hybrid diffractive/refractive lenses and achromats, *Appl. Opt.*, 27 (14), 2960–2971, 1988.
8. G. P. Berhmann et al., Color correction in athermalized hybrid lenses, *OSA Tech. Dig.*, 9, 67–70, 1993.
9. J. W. Goodman, *Introduction to Fourier Optics*, 2d ed. McGraw-Hill, San Francisco, 1996.
10. J. R. Fienup Iterative Method applied to image reconstruction and to computer generated holograms, *Opt. Eng.*, 19, 297–305, 1980.
11. K. H. Hoffmann and P. Salomon, The optimal simulated annealing schedule for a simple model, *J. Phys. A*, 23, 3511–3523, 1990.
12. U. Malhab et al., Iterative Optimization Algorithms for filter generation in optical correlators: a comparison, *Appl. Opt.*, 31 (8), 117–1125, 1992.

13. R. W. Gerchberg and W. O. Saxton, A practical algorithm for the determination of phase from image and diffraction plane pictures, *Optik*, 35, 237–246, 1972.
14. H. Farhosh et al., Comparison of binary encoding schemes for electron-beam fabrication of computer generated holograms, *Appl. Opt.*, 26, 4361–4372, 1987.
15. B. K. Jennison et al., Iterative approaches to computer generated holography, *Opt. Eng.*, 28 (6), 629–637, 1989.
16. R. Petit, Ed., Electromagnetic theory of gratings, in: *Optics in Current Physics*, Vol. 22, Springer Verlag, Heidelberg, 1980.
17. J. N.Mait and D. W. Prather, Eds., *Selected Papers on Subwavelength Diffractive Optics*, SPIE Milestone Series, Vol. M2 166. SPIE Press, Bellingham, WA, 2001.
18. W. C. Sweatt, Mathematical equivalence between a holographic optical element and an ultra high index lens, *J.Opt. Soc. Am. A*, 69, 486–487, 1979.
19. P. P. Clark et al., Ray tracing models for diffractive elements, *Opt. Soc. Am. Tech. Dig.*, 8, 2–3, 1993.
20. T. K. Gaylor and M. G. Moharam, Analysis and applications of optical diffraction by gratings, *Proc. IEEE*, 73, 894–937, 1985.
21. T. K. Gaylor and M. G. Moharam, Planar Dielectric grating diffraction theories, *Appl. Phys. B*, 28, 1–14, 1982.
22. W. B. Veldkamp, Binary Optics: The optics technology of the decade, in: *Proceedings of 37th International Symposium on Electron, Ion and Photon Beams*, San Diego, CA, 1993.
23. M. W. Farn, Binary Gratings with increase efficiency, *Appl. Opt.*, 31, 4453–4458, 1992.
24. Vasara et al., Binary surface relief gratings for array illumination in digital optics, *Appl. Opt.*, 31, 3320–3336, 1992.
25. M. W. Farn et al., Process error limitations on binary optics performance, in computer and optically generated holographic optics, *SPIE Proc.*, 1555, 80–88, 1991.
26. K. S. Urquhart and S. H. Lee, Computer Generated Holograms fabricated by direct write of positive electron beam resist, *Opt. Lett.*, 18, 308–310, 1993.
27. Zhou Zhou and Sing H. Lee, Fabrication of an Improved Gray-scale Mask for Refractive Micro- and Meso-Optics, *Opt. Lett.*, 29 (5), 457, 2004.

28. Zhou Zhou and Sing H. Lee, *New Gray-Scale Technology Part (1)-Two-beam-current method for the e-beam writing and its application to high-resolution optical structures*, to be submitted to *Applied Optics* in late March, 2007.
29. R. Khalil et al., Optical clock distribution for high speed computers, *SPIE Proc.*, 991, 32–41, 1988.
30. J. Goodman, Optical interconnections for VLSI systems, *IEEE Proc.*, 72, 850–866, 1984.

CHAPTER

14

Design of Illumination Systems

Introduction

Illumination optics is required in many varied system applications, including, for example, microscopes, projection systems, machine vision systems, industrial lighting. In optical systems where a light source is illuminating an object that has to be projected onto a screen as in a desktop projector, the design often requires a high brightness and uniformity across the image. High brightness implies a high collection efficiency of the light emitted from the source. Furthermore, these systems often require small packaging of the optical system.

Light sources have a wide range of types, shapes, and sizes, and the choice of the design of the illumination optics is very dependent on the source. Sources can be tungsten halogen lamps of different shapes, metal halide lamps, light-emitting diodes (LEDs), xenon lamps, frosted bulbs, different forms of arc lamps, or fusion (sulfur) lamps. Some sources have sufficient brightness and uniformity across their emitting area, and these can be imaged directly onto the object that has to be illuminated, but in most cases the sources need some kind of homogenization in the illumination optics to achieve the required brightness uniformity, while simultaneously minimizing throughput losses. The most common physical parameters that are used in photometry

321

to characterize the source and the illumination system are *flux*, *intensity*, *illuminance*, and *brightness*. *Photometry* deals with visual systems, and it is implicit that all relationships are weighted by the spectral sensitivity of the human eye. Clearly, if an object being viewed visually appears to be of a certain brightness, this must take into account the eye's spectral sensitivity. *Radiometry* is a closely related subject, which deals directly with power as it is emitted by a source, and ultimately irradiates a surface. The concepts and the basic theory are identical to those used in photometry; however, the units are totally different. We will define the basic photometric parameters and their corresponding units:

- *Flux* corresponds to power in radiometry, and it is the total power emitted by the source. The unit for flux is the lumen.
- *Intensity* is the flux per unit solid angle, or a solid angle flux density or the flux angular distribution, which assumes that the flux comes from a point source. The unit for intensity is the candela = lumen/steradian.
- *Illuminance* is the flux per unit area incident onto the surface being illuminated, or the area flux density. It does not relate to the angular flux distribution incident on the surface. The unit for illuminance is the foot-candela = lumen/ft², or lux = lumen/m².
- *Brightness (luminance)* is the flux per solid angle per unit area, or the area and solid angle flux density. The unit for brightness is the nit = candela/m².

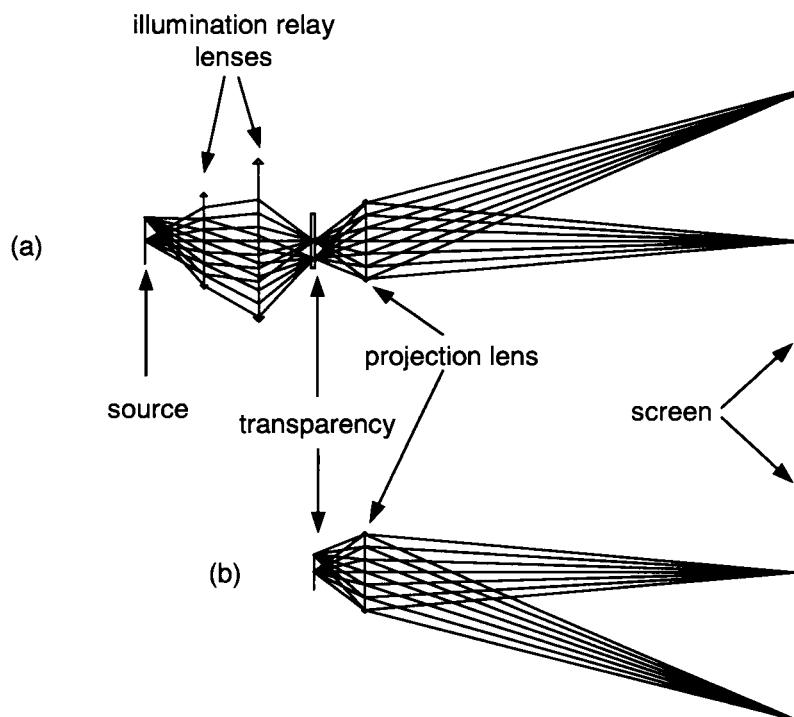
Köhler and Abbe Illumination

There are two classic approaches to the design of illumination systems. Most modern illumination system designs are modifications of one of these basic concepts.

Abbe illumination, which is sometimes called *critical illumination*, images the source directly onto the object to be illuminated, as shown in the paraxial model in Fig. 14.1. It is used in the cases when the source is sufficiently uniform for the system requirements. An illumination relay lens images the source onto the object. The object being illuminated can be film as in a movie projector, a 35-mm slide, or other similar transparency, or a transmissive or reflective LCD panel as in a desktop projector. Either

Design of Illumination Systems

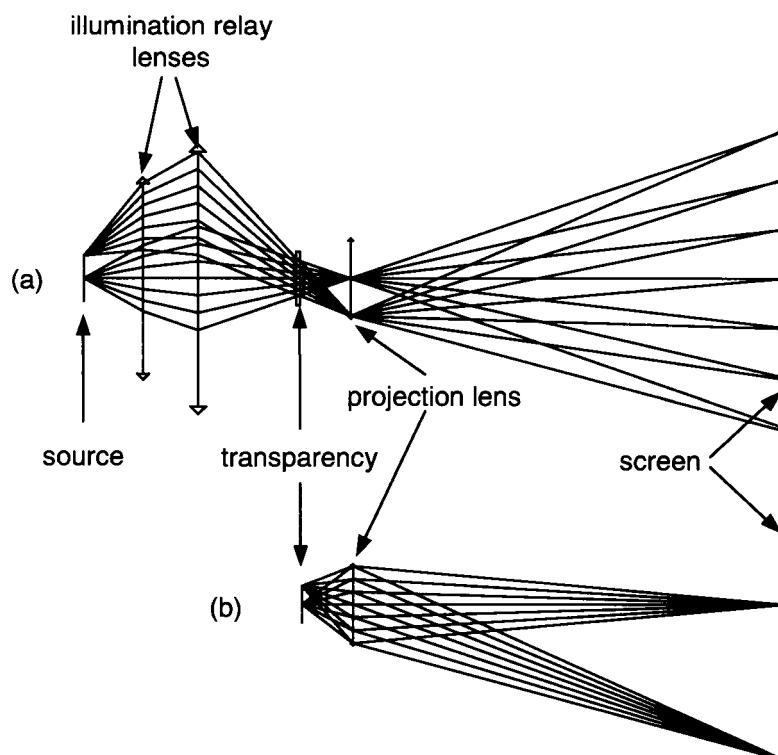
Figure 14.1
Abbe Illumination



the source must be inherently uniform as noted previously, or the condenser must have sufficient aberrations that blur the image of the source enough to eliminate the structure of the source at the transparency plane. The projection lens then images the object (transparency) onto the screen. At the same time, it images the image of the source onto the screen. This type of illumination works well with frosted bulbs or large sources, but it does not work with high brightness sources where high throughput is required and where there is significant structure in the source as with filament lamps and arc lamps. Generally, the resulting brightness and uniformity across the screen depends on the brightness and uniformity of the source, which means both spatial and the angular source uniformity.

In the case of a small bright, highly nonuniform source, such as an arc or a filament lamp, uniform brightness at the image is achieved with another form of illumination known as *Köhler illumination*, as shown in Fig. 14.2. Here, an illumination relay lens or condenser lens images the source into the *pupil* of the projection lens rather than onto the object being projected. This illumination relay system has its aperture stop at

Figure 14.2
Köhler Illumination



the location of the film, or the object, that is to be projected onto the screen, or into the eye. Since each point of the source illuminates the entire surface of the film, the film is by definition uniformly illuminated. Brightness nonuniformity with Köhler illumination can be caused by significant intense nonuniformity of the source.

Optical Invariant and Etendue

When a given light source emits a certain flux, not all of the emitted light reaches the screen or the detector. Some of that light is lost immediately after leaving the source. Since most sources emit into a large solid angle, often into a full sphere, we find that the first optical element, which may be a condenser system, has a difficult job of collecting and orienting or directing that light toward the aperture of the projection

Design of Illumination Systems

optical system. Some of the light is lost in the optical system as absorbed, scattered, diffracted, or vignetted light. There may also be some components, such as the beamsplitters, filters, or polarizers, whose purpose in the system is to transmit only a certain type of light and block the rest.

There are many different terms that are related to the coupling of the light from the source to the screen. These terms are the *optical invariant*, *etendue*, *light-gathering power*, *throughput*, *angle to area*, and *area-solid angle product*.

Let us first take an example of an imaging system such as a telescope. The flux that is transmitted through a telescope goes through the aperture stop of the telescope shown in Fig. 14.3. If our telescope has an entrance pupil diameter D_{in} and the field of view is $2\omega_{in}$, the exit pupil diameter is D_{out} and the exiting field of view is $2\omega_{out}$, then

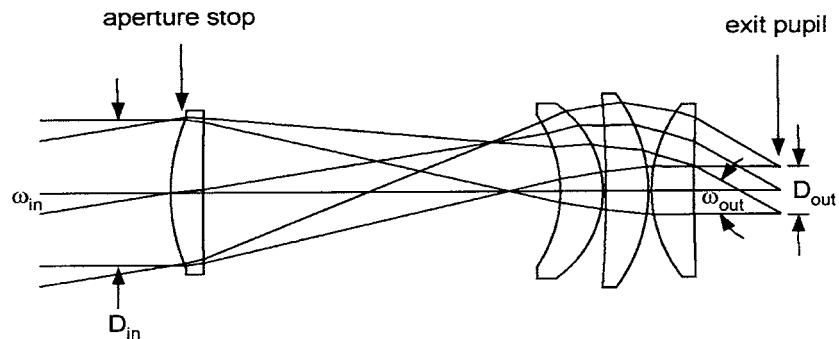
$$D_{in} \sin(\omega_{in}) = D_{out} \sin(\omega_{out})$$

In other words, optical invariant in the entrance pupil has the same value as in the exit pupil. This product maintains its value through the entire optical system. If we square the previous equation, we get

$$D_{in}^2 \sin^2(\omega_{in}) = D_{out}^2 \sin^2(\omega_{out})$$

The total flux that passes through the entrance pupil of the telescope will pass through the exit pupil if there are no losses within the system from absorption, vignetting, beamsplitting, filters, or scattering. This product $D_{in}^2 \sin^2(\omega_{in})$ is the measure of the throughput of the optical system and it is proportional to the area—solid angle product, or etendue. Etendue is, in effect, the conservation of flux within an optical system.

Figure 14.3
Flux at the Stop of a
Telescope and Tele-
scope Throughput



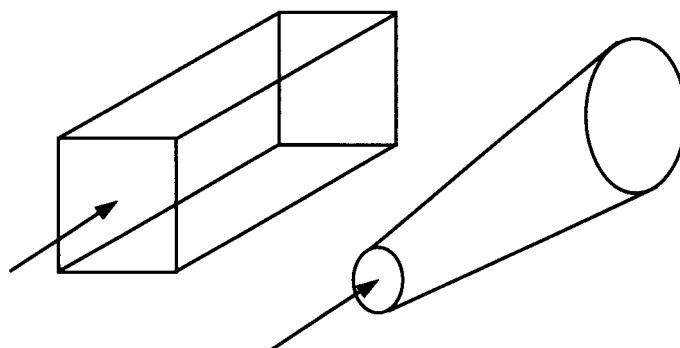
In a system with a light source, the goal is to use as much of the emitted light from the source as possible, and couple it into the optical system such as a projection lens for example. The etendue of the source (area of the source times the solid angle into which the source emits light) should be the same or slightly larger than the etendue of the projection lens. If the etendue of the source is significantly larger than the etendue of the projection lens, there are a lot of rays that are stopped by the apertures in the optical train, and these rays are lost and never reach the screen. On the other hand, if the projection lens is designed with the etendue larger than the etendue of the source, then the lens might be unnecessarily complex, with an underfilled aperture stop or underfilled field of view.

There are cases when the etendue is increased along the optical train. For example, if there is a need to polarize the light, then the unwanted polarization, instead of throwing it away, can be rotated at the expense of an increase in etendue. Another example is the presence of diffractive components in the system. If more than one diffractive order is used, the etendue in the system after the diffractive element is larger than before the diffractive element. Unfortunately, it is not possible to decrease the etendue without the light loss in an optical system.

The best correction of aberrations in most imaging optical systems is generally done at the center of the field of view. In illumination systems, on the contrary, the best aberration correction must be done at the edge of the field to get the sharp edge of the illuminated patch of light on the film, LCD or transparency.

Light pipes, as shown in Fig. 14.4, are commonly used as light homogenization components in illumination systems. Light pipes can be

Figure 14.4
Light Pipes



either hollow structures with reflective inside surfaces or they can be solid structures inside of which the light is totally internally reflected. Light pipes have two important roles in illumination systems. The first is that they are used as light homogenizers to change a spatial nonuniform distribution of the light at the input of the pipe into a uniform output. By definition, the etendue at the input surface of the light pipe is equal to the etendue at the output of the pipe.

The second very important function of light pipes is that tapered light pipes can change the angle of the input cone of light into a cone that can be accepted by the system into which the light from the source is injected. The optical invariant at the input of the pipe is equal to the invariant at the output

$$\frac{D_{\text{out}}}{D_{\text{in}}} = \frac{\sin(\omega)_{\text{in}}}{\sin(\omega)_{\text{out}}}$$

This formula gives the angle conversion in the case of a tapered light pipe and this is shown in Fig. 14.5. A tapered light pipe with multiple inside reflections is exactly like a kaleidoscope! Changing the angle between the rays and the optical axis after each internal reflection, it creates an array of virtual sources. This helps to spatially homogenize the brightness at the output of the pipe. Some shapes of the input (output) pipe surfaces spatially homogenize the light better than the others. Shapes such as a rectangle, a triangle, or a hexagon are the tileable shapes after unfolding due to internal reflections, and

Figure 14.5
Angle Conversion in
a Tapered Light Pipe

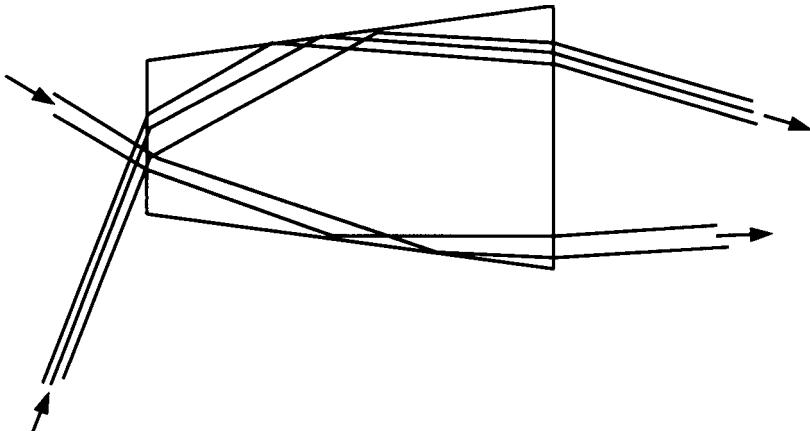
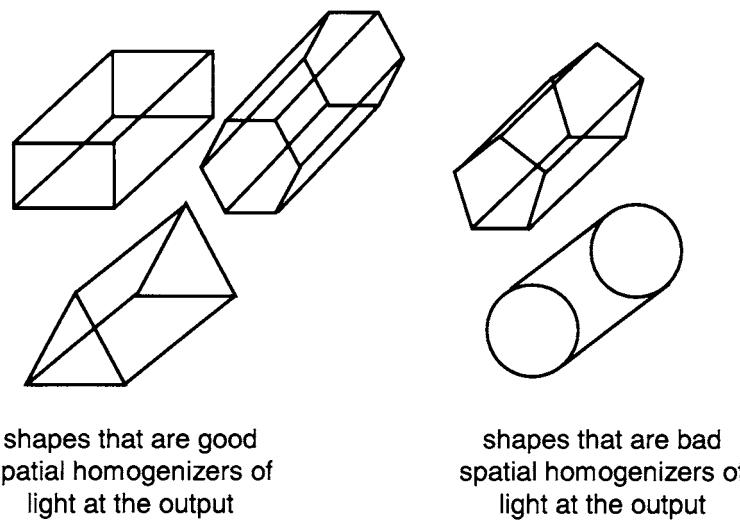


Figure 14.6
Different Shapes of Light Pipes

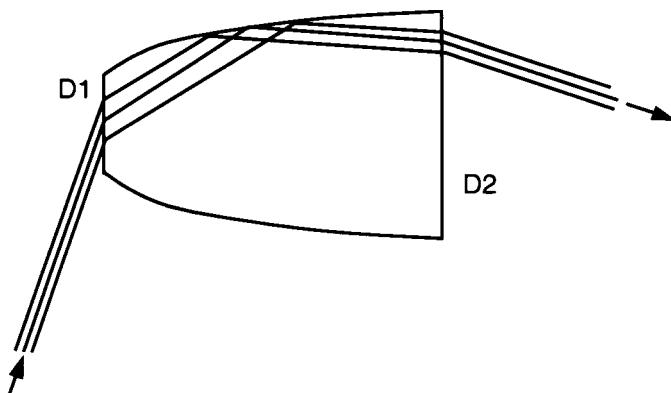


shapes that are good
spatial homogenizers of
light at the output

shapes that are bad
spatial homogenizers of
light at the output

they spatially homogenize the light very well. Different shapes are shown in Fig. 14.6. Straight walls of these pipes are good enough to give a uniform output. Other types that are not tileable, such as a circular input surface, do not homogenize the light as well, even when the pipe is very long. In the case of a circular input, improvement in uniformity is achieved with a pipe wall that is parabolic rather than a cone. This type of pipe is called a *compound parabolic concentrator* (CPC), as shown in Fig. 14.7.

Figure 14.7
Compound Parabolic Concentrator (CPC)



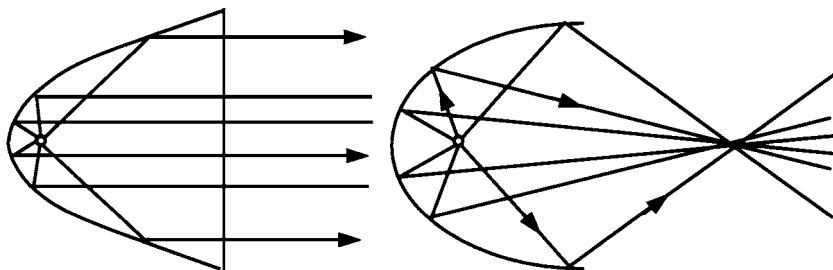
Other Types of Illumination Systems

The most efficient way to collect the light from a small source is to place the source at the focus of a parabolic or elliptical reflector. These reflectors collect the light emitted by the source into a large solid angle, and either collimate the light in the case of a parabolic reflector, or focus it to the second focal point of the ellipse shown in Fig. 14.8. There are different ways to reduce the nonuniformity of the image of the source and get a smooth illumination. One way of smoothing in collimating reflectors, such as flash lamps or street or automobile lighting, is the wedged reflector, where the reflector has a basic parabolic form, but it is divided into a lot of wedge-shaped segments. The other way is to combine the collimated light from the reflector with a smoothing plastic element located in front of the reflector, consisting of the extruded array of prisms or structures with combined prismatic and sinusoidal profiles.

In the case of a film projector, the goal is to send the light from the source into the rectangular area of the film, with the numerical aperture of the illumination matched to the numerical aperture of the projection lens. There are two common ways of getting a rectangular uniformly illuminated area in the film plane from a highly nonuniform, nonrectangular source.

The first method is to initially collect the light from the source with a reflector, then focus it down onto the input surface of the rectangular light pipe, as in Fig. 14.9. The magnification of the tapered light pipe is

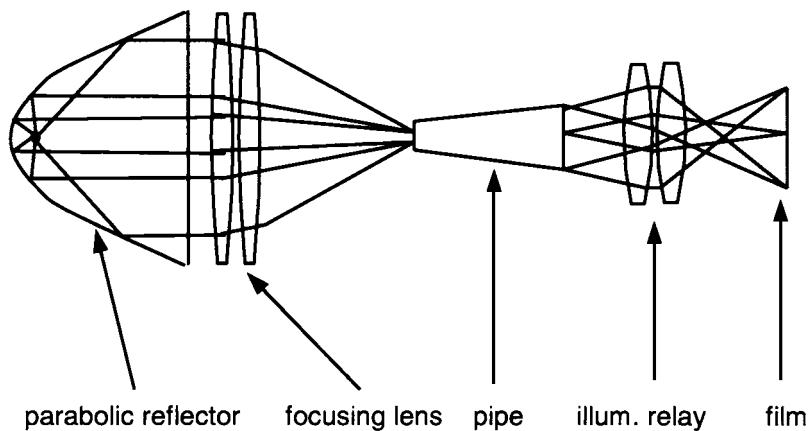
Figure 14.8
Parabolic and
Elliptical Reflectors



parabolic reflector

elliptical reflector

Figure 14.9
Illumination System
with a Tapered Light
Pipe



chosen such that the etendue at the output from the pipe is equal to the etendue of the optical system that relays the rectangular pipe output surface onto the plane of the film. The pipe has to be sufficiently long to get the spatially uniform output. A good rule of thumb is that the output will be uniform if there are at least three ray reflections along the pipe. The larger the magnification of the pipe, the smaller the minimum length of the pipe needed to achieve the uniform output.

The second method is often used in the illumination systems of front desktop projectors with transmissive image panels and with projection lenses of $f/3$ and above. It uses lenslet arrays along with a polarization recapture plate. This is shown in Fig. 14.10. The light from the source is collimated with the parabolic reflector. The expanded and collimated beam then goes through the lenslet array (lenslet array is shown in Fig. 14.11). Each lenslet focuses the beam inside the lenslet of the second lenslet array, producing an image of the source. The light that illuminates the image panels has to be linearly polarized. After the second array, there is an array of polarization beamsplitters, which splits the light into two orthogonal linearly polarized beams. One of the beams then undergoes the rotation of its plane of polarization, and two beams emerge in the same polarization state out of the polarization recovery array. The second lenslet array, together with the focusing lens, images the lenslets (which are rectangular) of the first array onto the image panel. The focusing lens superimposes images of the lenslets of the first array in the plane of the image panel. In summary, a circular distribution of flux from the parabola is sampled by rectangular elements of the lens arrays and is then superimposed at the

Design of Illumination Systems

Figure 14.10
Illumination System
with Lenslet Arrays
(Only the Rays Passing
through One
Lenslet Are Shown)

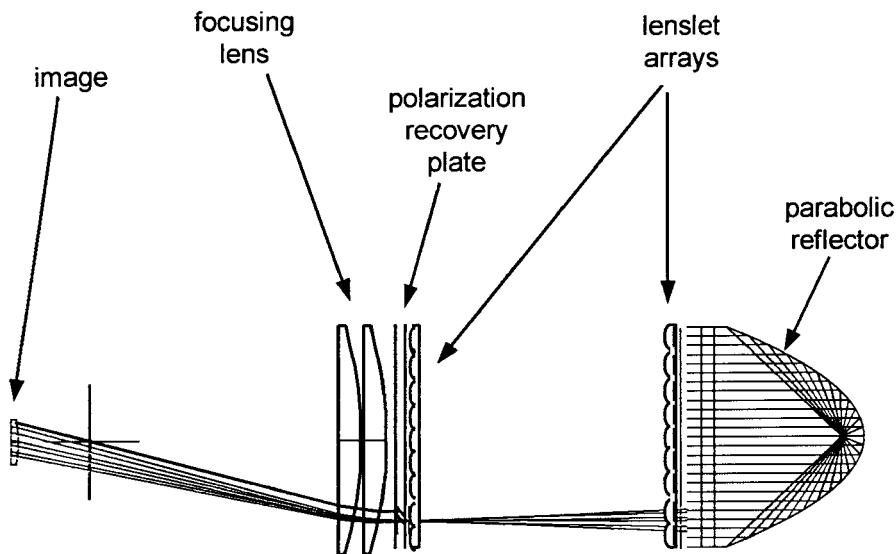
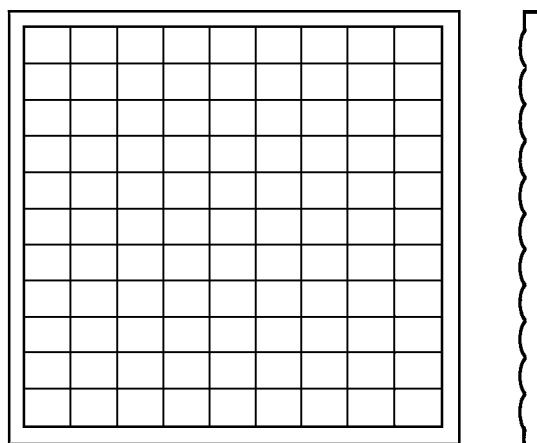


image panel, thus homogenizing the nonuniform parabolic output and transforming the geometry to match that of the image panel. An illumination system with a light pipe has a high throughput and small packaging. However, mounting of the light pipe is a problem, since all pipe surfaces are the optical surfaces, and any contact with the sides of the pipe frustrates the total internal reflection, resulting in a loss of light. Although the system with lenslet arrays requires more space, it accommodates the polarization recovery, and increases the throughput 30 to 40%.

Figure 14.11
Lenslet Arrays



This page intentionally left blank

CHAPTER

15

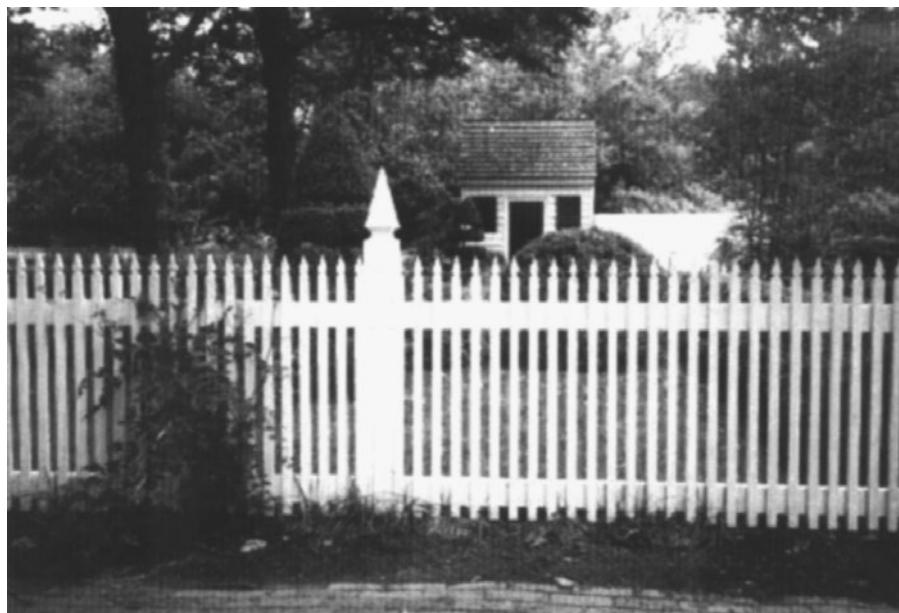
Performance Evaluation and Optical Testing

The quantitative characterization of optical performance, or image quality, is extremely important. Generally, the optical design engineer plays a key role in system testing, and for this reason we feel it is important to include the basics of optical testing in this book. Testing can range from the somewhat simplistic bar target to the more sophisticated means for characterizing the modulation transfer function (MTF).

Testing with the Standard 1951 U.S. Air Force Target

The simplest form of resolution target is perhaps the white picket fence shown in Fig. 15.1. The image of the fence consists of alternating bright and dark bars as formed by the white pickets and the dark background between the pickets. If we image this fence with a camera lens, the image will be demagnified by approximately the ratio of the camera

Figure 15.1
The White Picket
Fence Analogy to a
Bar Target for Optical
Testing

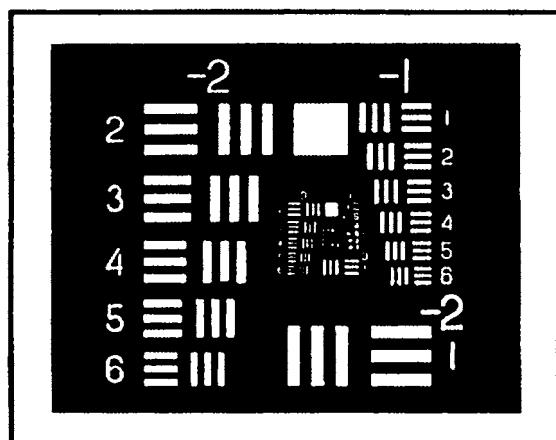


focal length to the distance to the fence. Let's assume the fence pickets are 75-mm wide and that we are imaging the fence with a 50-mm focal length lens from a distance of 20 m. The magnification is therefore $50/20,000 = 0.0025\times$. The fence is 150 mm/picket pair, or equivalently 150 mm/line pair. This equates to 0.006667 line pairs/mm. At the image formed by our lens, this becomes 0.375 mm/line pair, or 2.667 line pair/mm. Most camera lenses will resolve this spatial frequency just fine, as it is a rather low spatial frequency. Let's now move our lens to a distance of 200 m. Here the magnification is 0.00025 \times and the spatial frequency becomes 26.67 line pair/mm. A reasonably good camera lens will have a contrast of approximately 50% or higher at this spatial frequency. Needless to say, as the lens moves further from the fence the spatial frequency, in line pairs per millimeter increases and the contrast decreases as a result of aberrations, diffraction, assembly and alignment errors, and other factors.

If we had no other metric, a white picket fence would be a reasonable target to use for lens performance testing and characterization. In the laboratory, the most basic means for measuring image quality is through the use of the so-called 1951 U.S. Air Force Target. This form of target is readily available, low in cost, and easy to use. A typical Air Force target is

Performance Evaluation and Optical Testing

Figure 15.2
The Standard 1951
U.S. Air Force Target



shown in Fig. 15.2. The legend for the target is shown in Table 15.1 where it is evident that the target is divided into groups and elements so, for example, group 2, element 4 is a bar pattern of 5.66 line pair/mm.

How do we use an Air Force type target? Let us consider the example of measuring the performance of a 100-mm focal length 35-mm camera lens. Let us further assume that we must “resolve” 50 line pair/mm. As will be discussed in Chap. 22, this is a reasonable value for such a lens. We now construct a test setup shown in Fig. 15.3. We locate an Air Force target at the focus of a collimator lens. The collimator is used to simulate an infinite object distance. It is *critical* that the quality of the collimator *must* be independently validated and demonstrated to be better than the level of performance we are looking for. Generally, the focal length

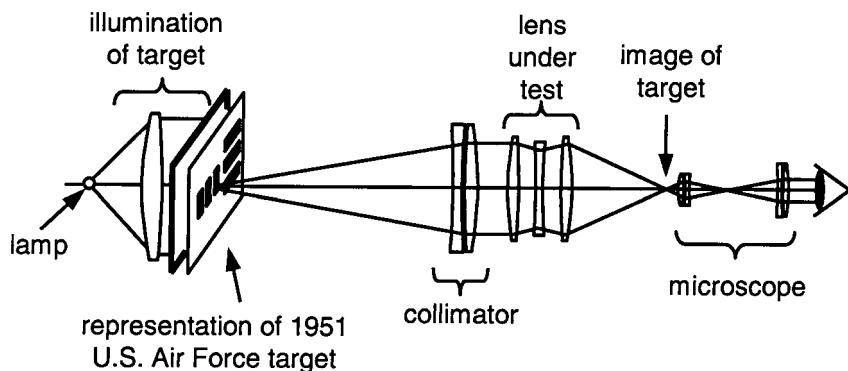
TABLE 15.1

Legend for 1951
Standard U.S. Air
Force Resolution
Target, lp/mm

Element Number	Group Number									
	-2	-1	0	1	2	3	4	5	6	7
1	0.25	0.5	1.0	2.0	4.0	8.0	16.0	32.0	64.0	128.0
2	0.28	0.561	1.12	2.24	4.49	8.98	17.95	36.0	71.8	144.0
3	0.315	0.63	1.26	2.52	5.04	10.1	20.16	40.3	80.6	161.0
4	0.353	0.707	1.41	2.83	5.66	11.3	22.62	45.3	90.5	181.0
5	0.397	0.793	1.59	3.17	6.35	12.7	25.39	50.8	102.0	203.0
6	0.445	0.891	1.78	3.56	7.13	14.3	28.51	57.0	114.0	228.0



Figure 15.3
Using the Air Force
Target to Test a
Camera Lens



of the collimator lens should be at least a few times longer than the focal length of the lens under test (a factor of 3 is the minimum). The collimated light now enters the lens under test and the image of the target comes to focus in the image plane as shown.

In order to compute which pattern corresponds to 50 line pair/mm, we can simply multiply 50 by the magnification from the target to the final image. This magnification is the focal length of the camera lens divided by the focal length of the collimator. Assume we have a collimator with a 1-m focal length. This gives us a magnification of $0.1\times$, which means that 5 line pair/mm is the spatial frequency on the target corresponding to 50 line pair/mm at the focus of our camera lens. Group 2, element 3 is 5.04 line pair/mm and is a sufficiently close match.

We now magnify this image with a microscope of suitable magnification so that the 50 line pair/mm is sufficiently well magnified so as not to be limited by the resolution of the eye. In order to compute the required magnification, first we need to find what angular subtense our 50 line pair/mm equates to: $50 \text{ line pair/mm} = 0.02 \text{ mm/line pair}$, which when viewed at a nominal distance of 254 mm corresponds to $0.02/254 = 0.00008 \text{ rad}$. The eye resolves approximately 1 min of arc, which is 0.0003 rad. Our target is thus about 4 times smaller than what the eye can resolve, and this is not acceptable. If we magnify the target by $40\times$ with the microscope, it will be seen at 10 times the resolution limit of the eye, which should be just fine. Ten arc minutes equates to 1 cm at a distance of 3 m. Thus, a $40\times$ microscope should be used. To determine the magnification of the microscope, simply multiply the magnification of the objective lens by the magnification of the eyepiece. Thus, a $4\times$ objective and a $10\times$ eyepiece will do the job for us. We may want to use a higher

magnification to be sure that the eye is in no way limiting and for more comfortable viewing of the image. The microscope objective has to have a numerical aperture that is larger than the numerical aperture of the lens under test.

The person performing the test now visually views the bar pattern through the microscope and judges whether he or she can distinguish three separate bars in the pattern in group 2, element 3 (for both horizontal and vertical orientations). If the answer is "yes," the lens passes, if the answer is "no," the lens fails. This test is indeed quite simple and easy to set up. The disadvantage of this form of test is that a real person can only judge whether the bar pattern can or cannot be resolved, not the level of contrast or sharpness of the image. Also, the test is somewhat subjective and sometimes different people will get different answers. A common rule of thumb is that the eye can resolve a modulation of approximately 5%, or 0.05. This is a useful metric, but what about the lens' ability to produce a good image with a higher contrast, say at 20 line pair/mm? The lens will likely resolve this spatial frequency fine; however, the user cannot judge the level of modulation or contrast. The only thing the user can judge is whether a specific spatial frequency is or is not resolvable or distinguishable into three bars.

The Modulation Transfer Function

The modulation transfer function provides a more quantitative measure of the image quality of an imaging lens system than a bar target. We discussed the basics of MTF in Chap. 10. From a laboratory standpoint, there are two principal ways of measuring MTF. First, we can use sinusoidal patterns of different spatial frequencies and image them with the lens under test onto a CCD sensor to characterize the resulting modulation as a function of spatial frequency. If you know the modulation of your target pattern, you can compute directly the modulation transfer function.

A more robust and easier way to measure MTF is by computing the Fourier transform of the line-spread function, which gives the MTF directly. The Fourier transform is a widely used mathematical transformation whereby we can compute the transformation from the spatial domain to the frequency domain, thereby giving the MTF directly. A lab setup such as in Fig. 15.4 is used, and is explained here. The test setup shown is for a finite conjugate lens, which is $f/20$ on its object side and $f/2.8$ on its image side:

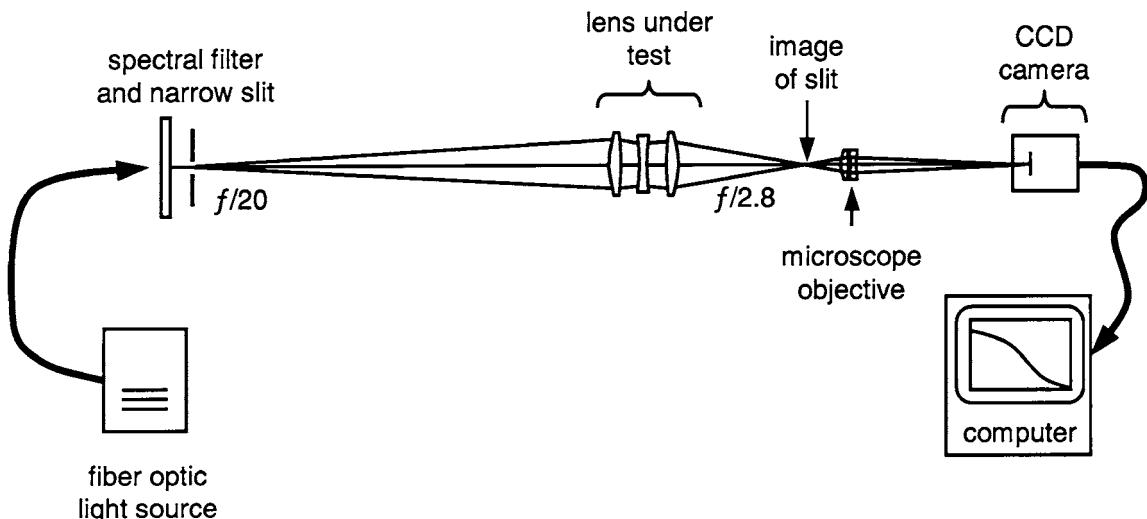


Figure 15.4
Example of Lab Setup to Measure MTF

- A light source with an appropriate spectral filter illuminates a narrow slit. The slit width should be narrow enough so that its finite width does not affect the results. As we know, the diameter of the Airy disk diffraction pattern is approximately equal to the f /number, in micrometers. A good rule of thumb is that if our slit width is 25% or less of this value, then its effect on the resulting Fourier transform will be negligible. For our $f/20$ object cone, the slit width required is approximately $5 \mu\text{m}$ or less since the Airy disk diameter would be about $20 \mu\text{m}$. If we cannot find a sufficiently narrow slit, the software can divide the resulting MTF by the Fourier transform of the rectangular slit, which is the MTF of the slit itself. This, in effect, divides out the effect of the finite slit width.
- The lens under test is now positioned on the optical bench.
- The image of the slit is now magnified with a high-quality microscope. The profile of the intensity distribution in the image of the slit, in the direction normal to the slit, is called the *line-spread function* (LSF). The magnified image of the slit is imaged onto a CCD array.

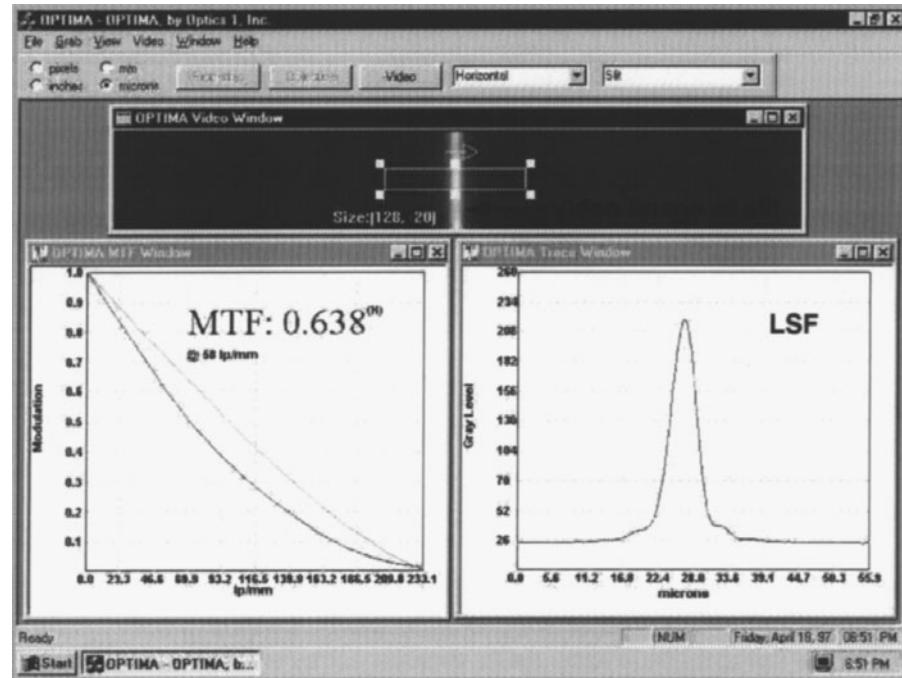
Performance Evaluation and Optical Testing

- The analog or digital output of the camera is input into a computer with a frame grabber, where the real-time video image of the slit is displayed in one of the windows, as shown in Fig. 15.5. A scan through the gray levels (line-spread function) is displayed in another window.
- The Fourier transform of the LSF is computed and is displayed in the remaining window. The perfect system MTF or diffraction-limited MTF is shown for reference and is computed based on user input of the central wavelength and the f /number.

The MTF is one of the most useful means for characterizing the optical performance of an imaging system. Most scenes consist of objects of many spatial frequencies, and the MTF data tell us how the modulation of the object is transferred from the object to the image as a function of the many varied spatial frequencies in the scene.

All lens design computer programs allow for the computation of the MTF for a given design. It is important to remember that these results

Figure 15.5
Example of MTF
Measurement
Software

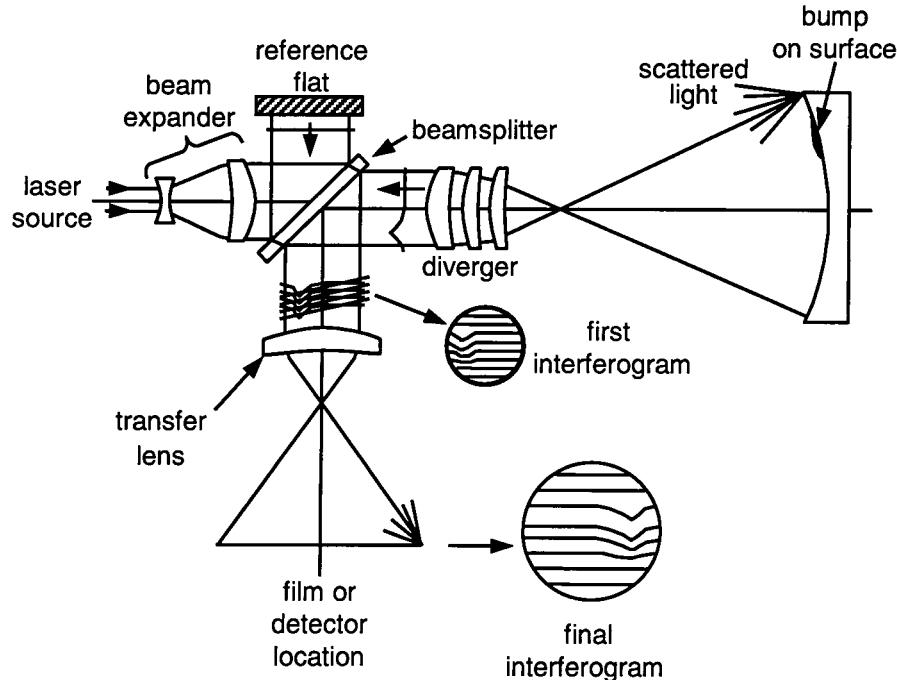


are for the design prescription, or paper design, and do not include the effects of fabrication, assembly, and alignment errors. This nominal design thus needs to have sufficiently higher MTF to account for the MTF degradations due to these manufacturing errors.

Interferometry

Interferometry allows us to measure quantitatively the optical path difference of a lens or, alternatively, the departure of a curved or flat surface from its nominal shape. While the geometry of many interferometers is different, their basic functionality is similar in that the wavefront from a perfect reference surface, such as an optical flat, is compared and interferes with the wavefront from the lens or surface under test. To illustrate just how an interferometer works, consider Fig. 15.6, where a Twyman-Green interferometer is shown:

Figure 15.6
Twyman-Green
Interferometer



- Light from a monochromatic and coherent light source, such as a HeNe laser, is input into a beam expander. The beam expander expands the approximately 0.8-mm-diameter beam to a diameter in the order of 25 to 50 mm. While the beam expander should be of high quality, it is not necessary that it be *perfect* as any small errors will cancel since the beam goes to both the reference arm as well as the test arm of the interferometer, and small wavefront errors will cancel each other.
- The expanded laser beam is incident onto a nominally 50–50 beamsplitter plate. Let us assume that the first (left-hand) surface has the beamsplitter coating and the opposite side is antireflection coated.
- Fifty percent of the light now travels upward to the reference flat. This mirror must be near perfect and is often in the order of $\lambda/40$ wave *P-V*. The light reflects downward from the reference flat and 50% of it passes through the beamsplitter, with the rest going back toward the laser. This is the so-called reference arm of the interferometer.
- The remaining 50% of the light, which passes through the beamsplitter, enters what is known as a *diverger* or *transmission sphere*. This is the test arm of the interferometer. We will assume that we are testing here a spherical mirror. The diverger is, in effect, a perfect lens that creates a perfect Airy disk at its focus and creates perfectly diverging wavefronts following the intermediate image. The purpose of the diverger is to create a wavefront that perfectly matches and nests into the nominal shape of the mirror under test, which means it appears that the wavefront emerges from the point coincident with the center of curvature of the mirror under test. If the mirror were a paraboloid or some other nonspherical shape, then a *null lens* is used instead of a diverger, with the same goal of creating a wavefront which nests into the nominal surface under test. We will show the use of a null lens in Chap. 25.
- We will assume that the mirror under test has a small bump on it, as shown, but is otherwise perfect. A portion of the wavefront will first hit the top of this bump and reverse its direction. The rest of the wavefront will now travel to the mirror surface and then turn around. Meanwhile the part of the wavefront that hits the top of the bump is already heading back toward the diverger. The bump

in the wavefront will thus be a factor of 2 times as large as the physical height of the bump itself.

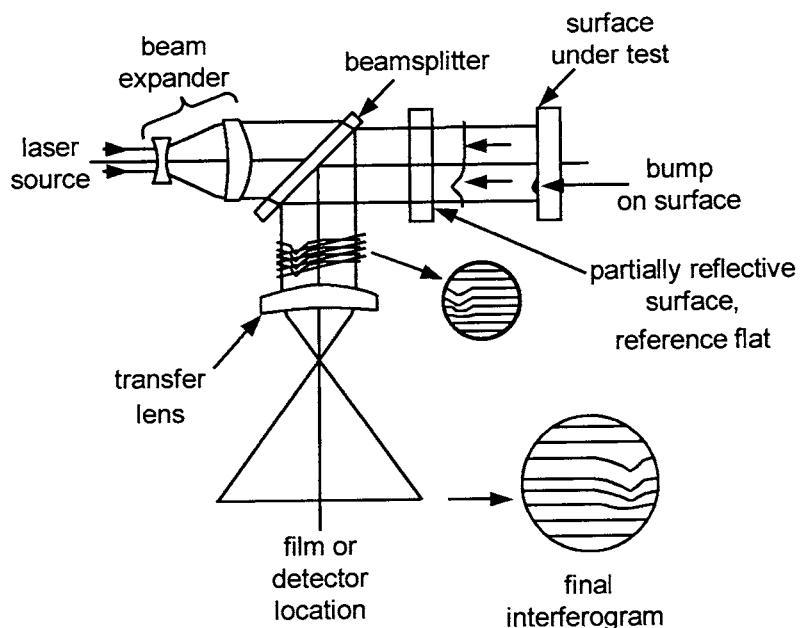
- The return wavefront passes through the diverger and is once again a series of plane parallel wavefronts, with the exception of the bump.
- Fifty percent of the light incident onto the beamsplitter will now reflect downward, with the rest going toward the laser. At this point, we have two wavefronts that are located between the beamsplitter and the transfer lens, one from the reference flat and the other from the mirror under test. Since the light is monochromatic and coherent, the light will interfere. What this means is that for regions where the two wavefronts are in phase we see a bright fringe and for regions where they are 180° out of phase we see a dark fringe. Since the wavefront from the reference arm of the interferometer is essentially perfect, any deviations are from the test arm, specifically from the mirror under test. If we were to place a white card in the beam here, we would see interference fringes indicative of the departure of the mirror under test from perfect sphericity. There is only one problem—due to scattering from the edge of the mirror under test we may see artificially curved fringes at the outer periphery of the interferogram. This effect could be interpreted as a turned up or down edge on the mirror, which is false.
- In order to eliminate the false turned up or down edge phenomena described, a *transfer lens* is used. This lens, or lens group, serves a dual role. It images the surface under test onto the sensor (CCD film, etc.), thus bringing the scattered light back to the edge of the interferogram where it belongs to eliminate false fringes. It also matches the size of the image to the sensor. If this is done properly, the final interferogram will show crisp edges with no artificial edge effects. Figure 15.6 shows how the scattered light at the edge of the surface under test is reimaged onto the interferogram plane by the transfer lens.

Another common form of interferometer is the Fizeau interferometer, as shown in Fig. 15.7. This system is similar to the Twyman-Green interferometer; only the reference beam is created from the partial reflection from a reference flat, which is also used in transmission.

It is important in any of the interferometer tests that the two beams (the reference and the test beams) are of approximately the same intensity in order to maximize the fringe contrast.

Performance Evaluation and Optical Testing

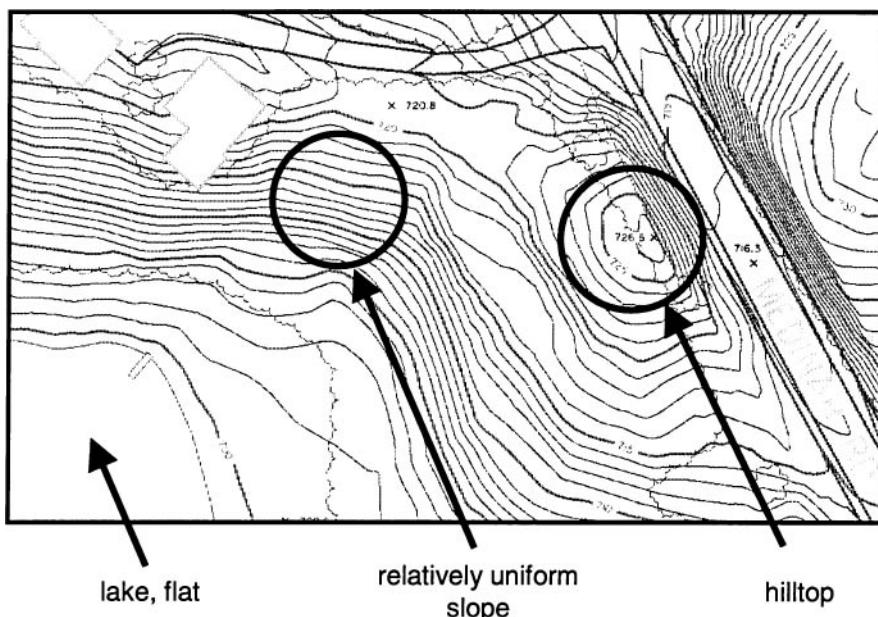
Figure 15.7
Fizeau Interferometer



The interpretation of an interferogram is very much like interpreting a topographical map with contours of equal altitude or elevation. In an interferogram, the fringes are lines or contours of equal height just as with a topographical map. Consider the topographical map shown in Fig. 15.8. The lake at the lower left is flat or level; hence there are no contour lines. The top center circle contains nearly straight contours, which indicates a flat region which is also tilted or sloped, and the circle at the right of the hilltop shows a somewhat dome-shaped hilltop which has a rapid falloff to the right. The interpretation of an interferogram is exactly the same as that of a topographical map with the only difference being the scale of the contours. In a topographical map the contour levels may be in units of meters or feet, and in an interferogram the units are in wavelengths of light.

Now consider Fig. 15.9 where we show four interferograms of a nominally flat mirror. Each fringe is due to one-half wave of surface departure from flatness, which results in one wave to the reflected wavefront, hence one fringe. The upper left interferogram shows a surface, which for the most part is tilted left to right. The upper right interferogram is quite flat, with a residual tilt from an 11 o'clock to a 5 o'clock direction.

Figure 15.8
Topographical Map
Analogy to Interferogram



And the lower left interferogram shows a saddle-shaped residual. It is important to note that interferograms contain no sign information whatsoever with respect to what is high and what is low on the surface. This information *must* be obtained when the part is being tested, as it is impossible to derive the sign information at a later time. The saddle could be down in the 10 o'clock and 4 o'clock directions and up in the 1 o'clock and 7 o'clock directions. Alternatively, the up and down directions could be reversed. In fact, it is physically possible (although unlikely) that the surface only departs up (or down) in the four regions.

Other Tests

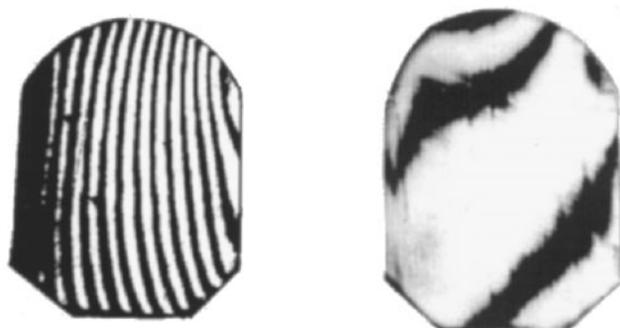
There are many other optical system testing methods, and the more important of these will be briefly outlined here:

- The *star test* is where you view the image of a point object (a pinhole) similar to a star through a microscope visually (or via a video camera). If a reasonably narrow-band filter is used, the image should be that

Performance Evaluation and Optical Testing

Figure 15.9

Typical Interferograms of Nominally Flat Mirrors



interferograms of thin aluminum mirror
(both interferograms are of same mirror)



interferogram of thin aluminum mirror with astigmatism or saddle

interferogram of thin beryllium mirror

of an Airy disk if the system is diffraction limited. This is an extremely sensitive test since very small asymmetries in the Airy disk are quite evident. The human eye is sensitive to many orders of magnitude in dynamic range, which makes the test quite robust. If the microscope is now focused inside and then outside of best focus, you will see a disk with concentric rings which becomes larger as you move further from best focus. This ringed disk should appear similar both inside and outside of best focus. If it is not, then you have a residual of spherical aberration. While not quantitative, the star test is an extremely sensitive test.

- The *Hartman test* is often used at observatories to test telescopes. A mask with a grid of small apertures is placed in front of the primary mirror, and the resulting imagery is recorded at several

through-focus locations. This is the analogy of creating a series of through-focus spot diagrams, and the residual aberrations can be derived using appropriate software.

- Other forms of optical system testing relate more to alignment of the system as a whole as well as its subcomponents or groups. Measurement of focal length, distortion, and other lens metrics is also important. We often use alignment telescopes, laser beams, and other methods for assuring that our optics is sufficiently well aligned.

CHAPTER

16

Tolerancing and Producibility

Introduction

Most of the material presented in the earlier chapters of this book is associated with achieving the optimum lens design. There has been a lot said over the years about lens design optimization theory and algorithms, global search algorithms, aberration theory, and other related topics. These are all directed toward achieving the optimum lens performance for what we sometimes call the “paper design.” The performance of the paper design is that of the design prescription, or the theoretical design, with the effects of absolutely no manufacturing errors of any kind included. While the performance of the paper design is indeed important, the effects of real-world hardware-related manufacturing errors and tolerances can and will, by definition, alter and degrade the level of performance from the theoretical paper design performance. We often find that the degradations due to manufacturing errors can many times even surpass the image degradations of the nominal design itself.

Tolerancing is the science (and art, to some extent) of distributing and error budgeting the manufacturing tolerances of all optical and optomechanical components and dimensions throughout the system to assure that your system will meet its required level of optical performance at a reasonable cost.

Unfortunately there is little, if any, direct correlation between the performance of the paper design and the robustness or insensitivity of the

design with respect to the level of manufacturing errors or tolerances. For example, we may have a design where steep bendings and high angles of incidence has allowed for the effective balancing of higher orders of aberration, thus yielding a high level of performance for the paper design. Unfortunately, this design may be extremely sensitive to tolerances due to the higher angles of incidence and the presence of higher-order aberrations. On the other hand, a different design configuration for the same lens requirements may have significantly reduced angles of incidence, reduced higher-order aberrations, and may result in a somewhat lower level of performance for the paper design. However, due to the reduced angles of incidence, this design may be less sensitive to manufacturing errors and tolerances. An example of this was shown in Chap. 9 where we discussed several designs submitted to the 1980 International Lens Design Conference. One of these designs was a very compact lens with small angles of incidence at the surfaces, while one of the other designs had extremely large angles of incidence on several surfaces. This latter design will, in all likelihood, have tighter manufacturing tolerances.

The key point is that manufacturing errors in the form of fabrication, assembly, and alignment errors can be extremely important and are often a major contributor to the overall level of performance of an optical system, even if the paper design is excellent.

In tolerancing an optical system, we need to assign tolerances to all optical and mechanical components within the system. This includes all lenses and/or mirrors as well as the mechanical components, which directly or indirectly support the optics. The overall goal for the system is that the optical performance is met (MTF and/or other image quality criteria), optical component costs are minimized, assembly and alignment costs are minimized, and yields are maximized. Tolerancing is necessary whether you are producing a lens in a high-production environment or a one-of-a-kind lens.

Ultimately, the goal of the tolerancing effort is to aid in establishing a performance error budget whereby you can, with confidence, predict the expected level of optical performance.

What Are Testplates and Why Are They Important?

Prior to embarking on an extensive tolerance analysis, your design needs to be completed and finalized. One of the very last steps in this process is

Tolerancing and Producibility

that of matching radii to existing testplates or tooling. Virtually all optical shops have in their inventory hundreds of so-called testplates. These testplates (sometimes called *test glasses*) are a convex and concave mating pair of tooling with radii ranging from very short to very long. They are often made of low-expansion Pyrex glass, they have very low surface irregularity, and their radii have been measured to a high level of precision.

Once a design is finalized and the shop to manufacture the optics has been selected, the designer should proceed to match as many radii as possible to existing testplates. Let's take an example: Assume that we have a six-element double gauss lens with 10 different radii. Further, assume that after we compared the radii in the final design to the testplate list of the selected lens vendor, we found that the closest radius was 25.21 mm, and our vendor has a testplate of radius 25.235 mm, a difference of only 2.5 μm . Five sections of the testplate list from OPTIMAX is shown here to give you an indication of the number of plates available and the density, or closeness, of radii in the different radii regions.

Plate ID	Radius	Diameter	CC	CX
1	1.5000	2.5000	X	X
2	2.0000	3.2000	X	X
3	2.0000	3.9000	X	X
4	2.0470	3.5000	X	X
5	2.0470	4.0000	X	X
6	2.5150	3.2000	X	X
7	2.5150	5.0000	X	X
168	10.0150	14.2000	X	X
169	10.0150	14.9000	X	X
170	10.0600	13.2000	X	X
171	10.0600	14.9000	X	X
172	10.1800	13.9000	X	X
173	10.1800	14.7000	X	X
174	10.2240	13.5000	X	X
175	10.2240	16.6000	X	X
176	10.3100	17.0000	X	X

Plate ID	Radius	Diameter	CC	CX
177	10.3100	15.0000	X	X
178	10.3900	13.6000	X	X
179	10.3900	14.9000	X	X
180	10.5000	13.4000	X	X
181	10.5000	13.7000	X	X
483	25.0550	40.0000	X	X
484	25.0550	41.1000	X	X
485	25.1000	17.8000	X	X
486	25.1000	16.8000	X	X
487	25.2350	38.3000	X	X
488	25.2350	44.3000	X	X
489	25.2600	33.5000	X	X
490	25.2600	44.8000	X	X
491	25.3800	38.7000	X	X
492	25.3800	41.7000	X	X
493	25.4050	37.7000	X	X
494	25.4050	41.8000	X	X
495	25.4460	33.1000	X	X
496	25.4460	35.0000	X	X
948	75.3200	57.6000	X	X
949	75.3200	61.2000	X	X
950	75.5000	56.2000	X	X
951	75.5000	44.7000	X	X
952	75.8260	70.0000	X	X
953	75.8260	71.5000	X	X
954	76.2950	63.0000	X	X
955	76.2950	63.0000	X	X
956	76.4350	62.0000	X	X

Plate ID	Radius	Diameter	CC	CX
957	76.4350	63.7000	X	X
958	76.6150	49.9000	X	X
959	76.6150	52.0000	X	X
960	77.0930	69.0000	X	X
1491	500.2100	85.0000	X	X
1492	500.2100	87.0000	X	X
1493	501.5200	78.3000	X	X
1494	501.5200	56.6000	X	X
1495	508.1400	87.3000	X	X
1496	511.8840	88.0000	X	X
1497	511.8840	88.0000	X	X
1498	514.5900	69.3000	X	X
1499	514.5900	69.8000	X	X
1500	520.7050	68.3000	X	X
1501	520.7050	68.9000	X	X
1502	528.5250	102.4000	X	X
1503	528.5250	103.5000	X	X
1504	537.4110	73.6000	X	X
1505	537.4110	75.4000	X	X

We then proceed to change the radius in the design to the testplate value of 25.235 mm and then freeze it from any further changes. In effect, we will reoptimize the lens while constraining this radius to exactly match the testplate. All other variables and constraints of the design are the same as for our final optimization cycles, and our error function remains the same. After the first radius has been matched, we once again search for the closest radius of those remaining to any on the testplate list and match it to this testplate radius. Note that during the reoptimization process following the first testplate insertion, all of the remaining radii will change or “shuffle” a small amount, giving a whole new scenario with respect to which surface is now closest to an existing testplate. This process continues until we have matched all of the radii to existing testplate radii.

It is our experience that in most cases 100% of all radii should be able to be matched to existing testplates. If for some reason you have problems with the last one, two, or three radii, you can release several radii that you have already matched and then match the problem radius or radii, with the intent of matching 100% of all radii to existing testplates. A very important point to keep in mind is that when we are matching our last one or two testplates, there are barely enough variables to correct the aberrations, much less any constraints such as focal length. Thus, for the last few testplates we highly recommend either to allow some or all of the element thicknesses to vary (in addition to the already varying air spaces), or alternatively we may need to release or relax one or more of the system constraints such as focal length. Any further changes in the design will likely cause a negligible change in focal length. We had a system some years ago where we inadvertently left the constraint on focal length during the match of the final testplate. In order to meet this constraint, the design took on a highly degraded level of optical performance, which would have caused it to fail its performance specification. During the testplate fit, you may elect to match the closest or the farthest radius first; this is your choice. Our experience shows the closest algorithm to work just fine.

Most software programs have an automated testplate fitting routine, and these are quite robust and work well. But use them with care, as they are quite automated and you could run into one of the problems cited previously if you are not careful (such as the match of the last one or two testplates, producing poor performance due to overconstraining a first-order parameter such as focal length). If you do have a particular problem with matching a specific radius, you might consider running your optimization and varying this surface radius along with a constraint to match it to the testplate. By working the match into the optimization, your chances are best.

Why do we emphasize so strongly matching all radii to existing vendors, testplates, and how does this relate to tolerancing? The reasons are several:

- Testplates cost several hundred dollars or more each, and they take time to manufacture, perhaps up to several weeks.
- Testplate radii can be measured more accurately than they can be manufactured. For example, a radius of 25 mm can be *manufactured* to within about ± 0.025 mm (precision level for manufacturing); however, it can be *measured* to within about ± 0.00625 mm or less. We all realize that, given enough time and money, we could of course manufacture the radius to within the ± 0.00625 -mm tolerance or even better; however, this is generally not

economically feasible. Thus, matching testplates is like an insurance policy whereby your level of confidence in your lens working as predicted is enhanced after matching the radii to existing testplates.

When you perform your final tolerance analysis, the way you can best model the real-world situation with respect to the surface radii is to first assign a power fit to the testplate in units of fringes. You will then also assign a radius tolerance, which in this case means the *accuracy* to which the testplate radius is known to have been measured. This is a measurement of accuracy or capability tolerance, and it should be discussed with your lens manufacturer. And if you really want to assure that your system works well, have the shop remeasure the testplates you have matched and incorporate these newly measured results into your design. After all, equipment and techniques constantly improve, so why depend on a testplate radius measurement from 20 years ago?

On a different, yet related matter, for extremely high-precision systems the designer often incorporates the measured refractive indices of the glasses into the design. This is called a *melt design*, and is yet a further assurance that the design will perform as predicted. As with testplates, glass refractive indices can be measured more accurately than they can be manufactured.

How to Tolerance an Optical System

The basic procedure for tolerancing an optical system is shown in outline form in Fig. 16.1 and is described as follows:

1. We first assign viable tolerances to all toleranceable parameters within our system. This includes all optical as well as mechanical components. This first candidate set of tolerances should be reasonably achievable at a rational cost. If we know we have a very sensitive system and/or we require a very high level of performance, a somewhat tighter set of tolerances should be used; conversely if we have an insensitive system or a poorer level of performance, then looser tolerances can be used.

For an optical system with a reasonable level of performance, a candidate tolerance set may be derived from Table 16.1.

Note in Table 16.1 that the radius tolerance is listed as testplate measurement accuracy. This is the accuracy to which the testplates have

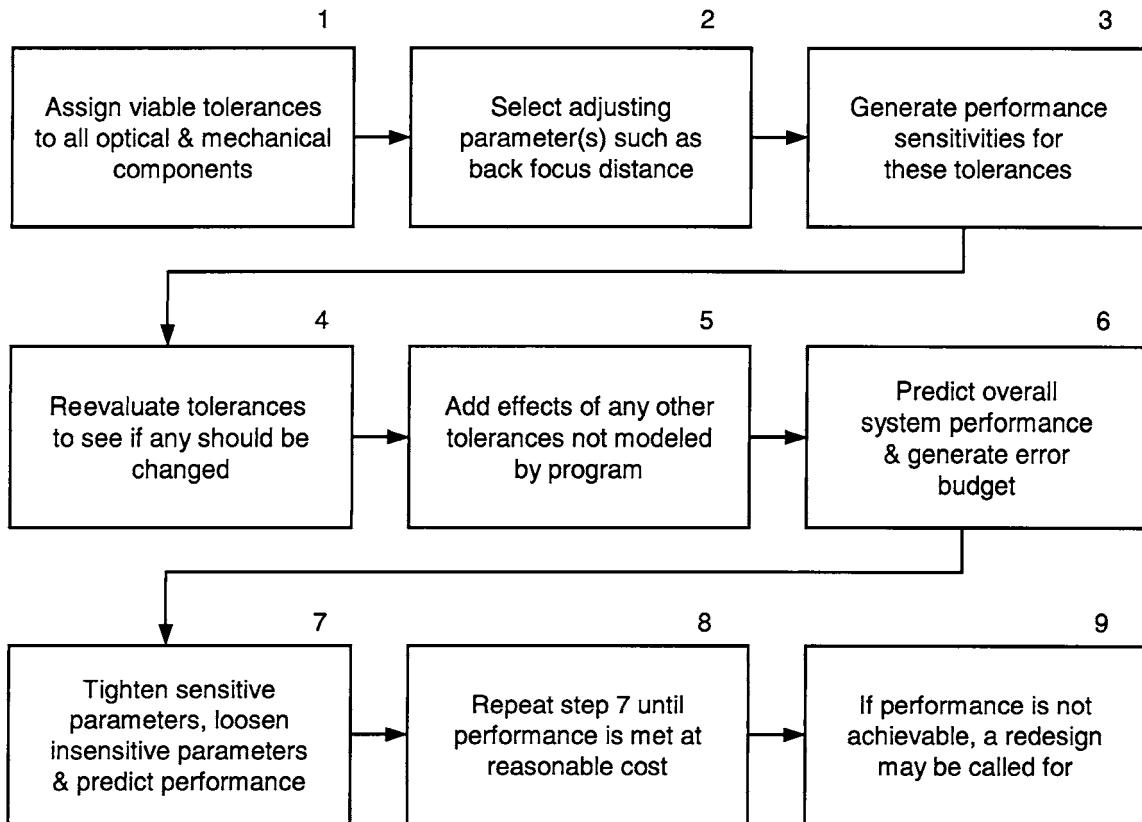


Figure 16.1
Tolerancing Procedure

TABLE 16.1

Candidate Tolerances for a Reasonable Performance Lens

Parameter	Tolerance	Parameter	Tolerance
Radius	Testplate measurement accuracy	Tilt	0.05-mm TIR
Power fit to testplate	Three fringes	Decenter	±0.05 mm
Surface irregularity	One fringe	Refractive index	±0.001
Thickness	±0.05 mm	Abbe number	±0.8%
Air space	±0.05 mm	Glass inhomogeneity	±0.0001
Wedge/centration	0.025-mm TIR		

been measured at the shop. Using a value of ± 0.01 mm is probably a reasonable assumption, but check with your shop to be sure. For long radii, a better assumption might be the radius change corresponding to ± 0.25 wave of sag at the outer periphery of the element clear aperture, or the testplate, whichever is smaller. This is because for long radii testplates the effective *f*/number of the light cone from the center of curvature of the surface is high, thus yielding a sizable depth of focus, which ultimately impacts the ability of determining the radius.

2. We then generate performance degradation sensitivities for all toleranceable parameters within the system using the tolerance routine in our lens design computer program. In other words, each and every fabrication, assembly, and/or alignment-related tolerance on all components is evaluated and sensitivities are determined. The tolerance forms include radius, power fit to testplate, surface irregularity, element thickness, airspace, element wedge, element tilt, element decentration, refractive index, Abbe number, and glass inhomogeneity. Other specific factors, such as the effects to performance due to the thermal environment, may also need to be included here.
3. As part of the preceding tolerance sensitivity analysis, we need to select the appropriate *adjusting parameter* or *parameters*. An adjusting parameter is an adjustment which you plan to allow for during the final lens assembly and testing. Back focus is the most common adjusting parameter and will almost always be used as a final adjustment during final system assembly. There can, in principle, be other adjusting parameters ranging from airspaces to tilts and decentrations. More on adjusting parameters later.
4. We now generate the performance sensitivities for this initial set of candidate tolerances. Thus, if our performance criteria were the MTF at 30 line pair/mm, we would determine the drop in MTF at this spatial frequency for each toleranceable parameter within the system.
5. We now go back and look carefully at each tolerance to determine if it should be changed from a manufacturing or assembly point of view. For example, larger airspaces may require somewhat looser tolerances. This is the all-important time to talk with your optical shop and your mechanical designer as well as your machine shop to reach a mutual understanding of the optical components and the mechanical design with respect to the anticipated levels of tolerances, which will affect directly the lens elements.

6. We now add the performance degradations from any other effects, which may not have been covered by our computer model. These may include atmospheric turbulence, surface irregularities that may not have been modeled in the computer analysis, and/or other effects. The net result is to predict the expected level of optical performance based on the assumed tolerance set.
7. We then predict the overall system performance and generate a performance error budget.
8. We now proceed to tighten any sensitive parameters and loosen insensitive ones and, again, predict performance. Here you need to look carefully at the effect on your performance metric such as MTF for each toleranced parameter to determine its effect. You will sometimes find that only a small subset of the overall tolerance set is sensitive, and the majority of the tolerances are insensitive. In these situations, tightening only a small number of tolerances can have a big payoff in improved performance. It is important to review the tolerances you intend to tighten with your optical shop and/or machine shop.
9. Step 8 is repeated as necessary until we meet the performance goal at a reasonable level of cost. The results of step 8 comprise the final set of manufacturing and assembly tolerances which, if applied and adhered to, will result in a system which meets your performance goals and objectives at a reasonable cost.
10. If the required performance is not achievable at a reasonable level of cost, you may need to return to your initial specifications to see which ones you may be able to relax. Also, a redesign may be called for with the specific goal of loosening the manufacturing tolerances.

How Image Degradations from Different Tolerances Are Summed

When you compute tolerance sensitivities as outlined earlier, you will ultimately need to predict the net system performance. Ideally, this should represent the predicted level of performance for some reasonably high percentage of manufactured systems such as a 95% cumulative probability or confidence level. But how do we add performance degradations in order to accurately predict performance? In other words, if we have, for example, 10 tolerances and each one degrades the performance

by a different amount, how do we use these data to predict the net result of assembling a large number of systems?

One common method of adding degradations is known as *RSS addition*, which stands for root sum square. In this method, we take the *square root of the sum of the squares* of each of the individual degradations, and if the degradations are of the same form, this should lead to a 95% confidence level. The degradations could be a peak-to-valley optical path difference (OPD), RMS OPD drop in MTF or some other criteria. While this can all be done, the criteria are not all mathematically correct, as will be discussed later.

To illustrate RSS addition and how powerful the method is, let us assume that we have a stack of pennies, and each penny has thickness $T \pm 0.1$ mm. If we have 100 pennies, the predicted thickness of the stack of 100 pennies is $100 \cdot T \pm 10$ mm worst case or $\sqrt{100} T \pm 1$ mm for a 95% confidence based on an RSS addition, since $\sqrt{0.1^2 \times 100} = 1$. The tolerance on the thickness of a penny is of a simple form in that the penny is either thicker or thinner than nominal; in other words, the degradations are all of the same type—they are all thickness errors.

In lenses, the degradations to performance are of different forms and are caused by different types of parameter errors. For example, performance degradation introduced by center thickness and airspace errors will likely introduce defocus plus spherical aberration on axis. On the other hand, element wedges, tilts, and decentrations will likely introduce coma and/or astigmatism on axis and across the field of view. The question now becomes: can we use an RSS addition to predict the net expected performance due to axial spacing errors as well as asymmetrical errors such as decentrations? The answer is a qualified “yes,” we can indeed RSS performance degradations caused by unlike errors (such as thickness errors and decentrations), but the more important question is whether the results are accurate and meaningful. The bottom line is that RSS addition simply does not handle properly the mix of aberrations that we often encounter in a lens system. For certain cases, RSS addition is quite appropriate and valid, and one such case is in computing the expected refocusing required following assembly. If we take the RSS of the refocusing required for each and every individual tolerance, the result should be valid since all of the refocusing is of the same form.

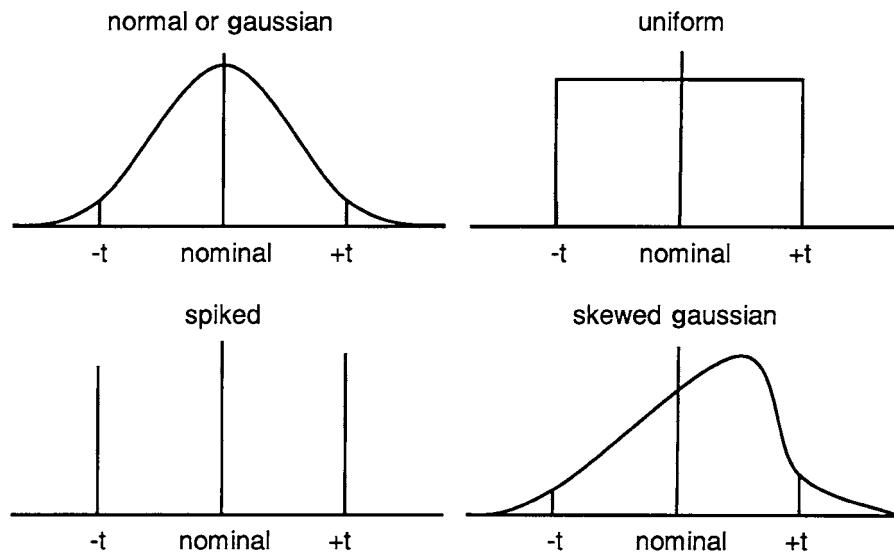
Fortunately, there is another approach known as *Monte Carlo tolerancing*, which is a method of simulating the performance statistics of a lens system in a high level of production. In a Monte Carlo simulation, tolerances are assigned to all toleranceable parameters along with a likely

probability function such as normal, uniform, spiked, or a skewed gaussian, as shown in Fig. 16.2. While a normal distribution seems to be the best model, we often find that tolerances sometimes tend to end up toward one end of the allowable range. Element center thickness is a good example, as the optician often leaves extra thickness just in case a scratch occurs and the element needs to be reground. Thus, the probability distribution for CTs is a skewed gaussian distribution with the highest probability on the thick side of nominal as shown in Fig. 16.2.

The following is the procedure used in computing a Monte Carlo tolerance analysis:

- Every parameter is independently and randomly perturbed according to its assigned tolerance, based on a likely probability distribution. At this point, we have created a random lens assembly on the computer, based on our current set of tolerances.
- The performance is now computed after applying any compensators such as refocusing. The system model at this point is a simulation of a single manufactured system with its tolerances randomly distributed as discussed previously.
- The preceding process is repeated 25 or more times, and for each Monte Carlo sample, we have, in effect, a simulated system that has been manufactured. We can use the resulting output to compute

Figure 16.2
Tolerance Distribution Models



the level of performance versus the cumulative probability. With these results, we can easily determine the level of performance for, say, 90 or 95% cumulative probability of occurrence.

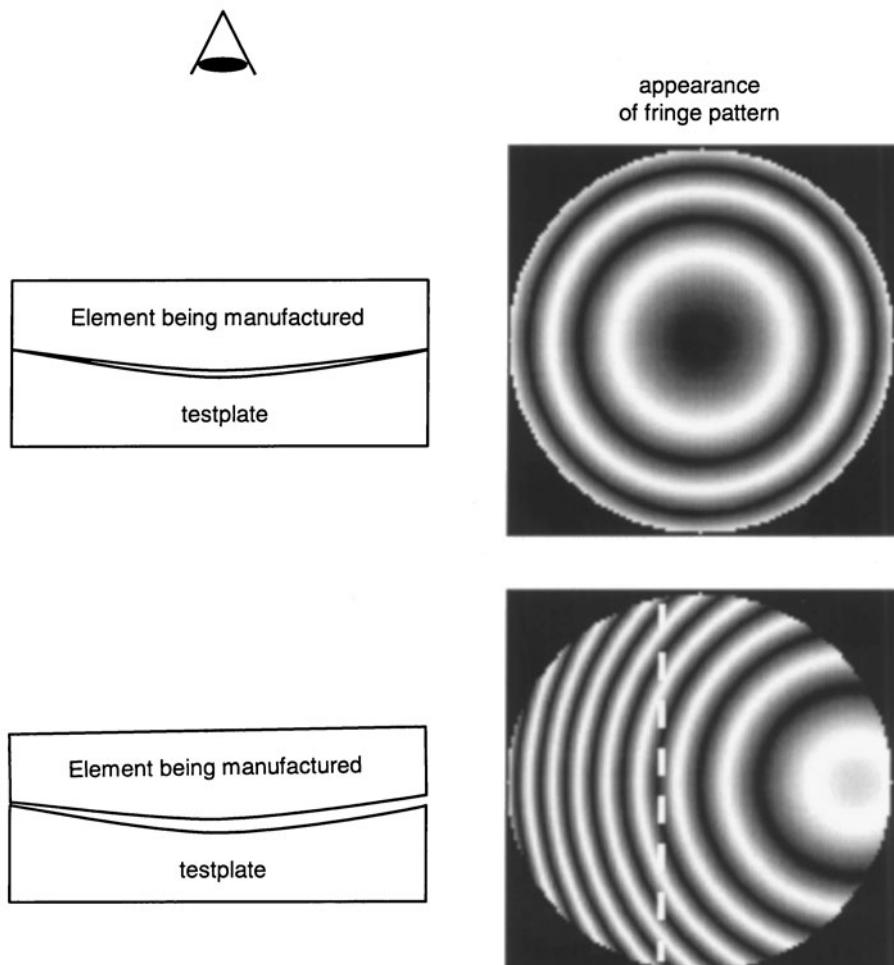
The beauty of the Monte Carlo approach is that it is fully valid regardless of the relative nature, mix, or form of the aberrations because we are simulating a manufacturing environment. Each and every Monte Carlo sample is, in effect, a new and different manufactured system.

Forms of Tolerances

We have discussed the various tolerances, along with ways and means for adding the performance degradations. Next we will show the various forms of the tolerance parameters:

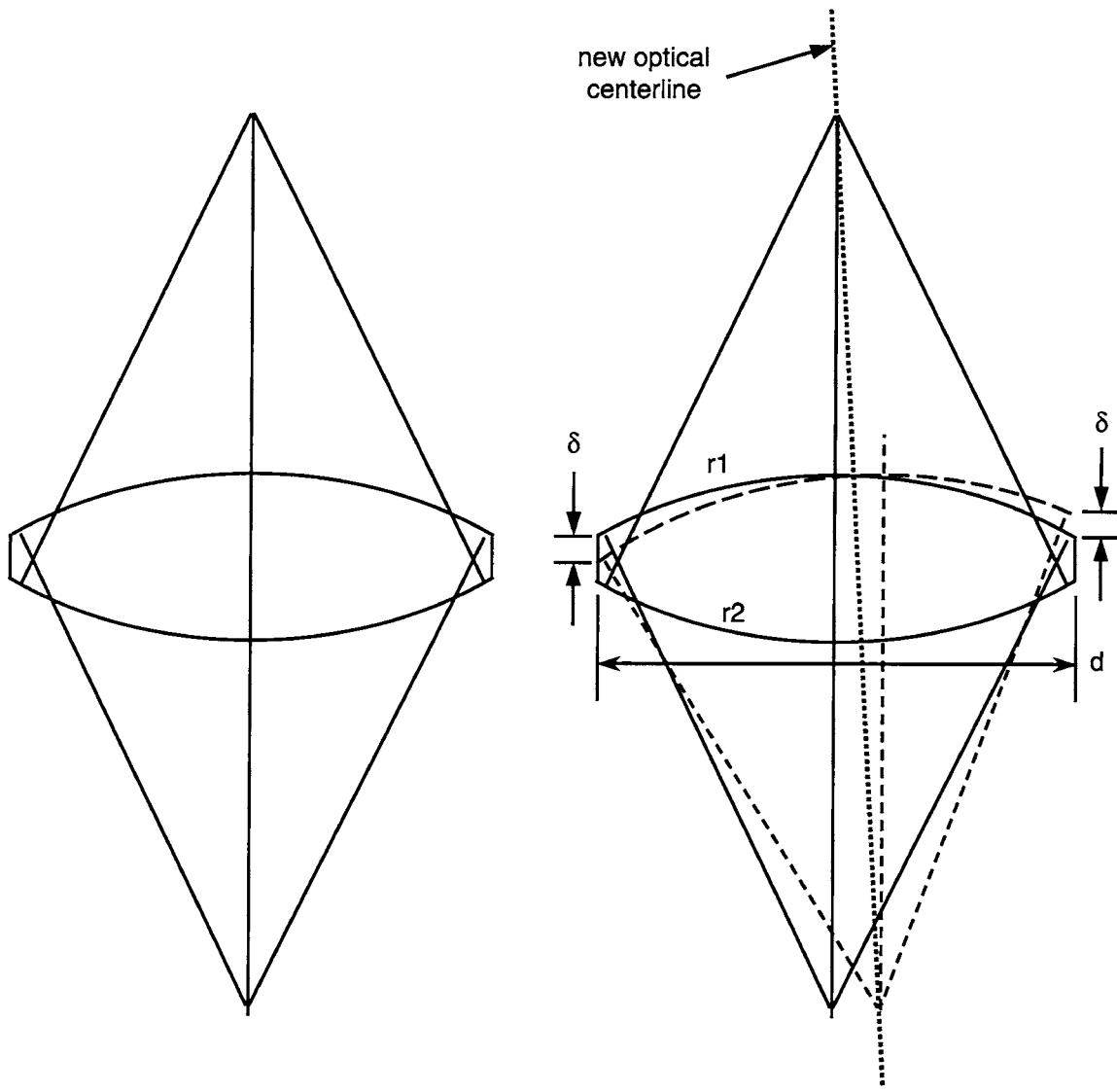
- *Symmetrical errors* relating to fabrication and assembly include radius, power fit to testplate, thickness of elements, airspaces, refractive index, and Abbe number. Figure 16.3 shows how a testplate is used to characterize a lens element in manufacturing. The testplate is used as a standard of a known radius to high precision. The surface under test is placed, or “nested,” into the testplate as shown, and every time the airspace or gap between it and the surface under test changes by $\lambda/2$ we see one full fringe. In the example shown, we see approximately two rings, or fringes, which means that approximately 1.0 wave of optical power or mismatch exists between the testplate and the surface under test. If the gap between the testplate and the surface under test is rotationally symmetrical, the interference pattern is indeed round, with circularly symmetric rings, as shown in the upper part of Fig. 16.3. If, on the other hand, we tilt the surface under test (this is the same as moving the ring center way off the part as shown), we see curved fringes as in the lower part of Fig. 16.3. The dashed vertical line is a reference to tell us how many fringes of power we have, and, as you can see, we have about two fringes of power.
- *Asymmetrical errors* in assembly and alignment include element wedge, element decentration, element tilt, surface irregularity, and inhomogeneity of refractive index. In effect, an element with a wedge is really the same as an element with its optical centerline tilted with respect to its mechanical centerline. Further, the element

Figure 16.3
Power Fit to Testplate



has an *edge thickness difference* as we rotate the element around.

Consider Fig. 16.4 where on the left we show a nominal perfect lens element. The two centers of curvature, when connected, represent a line straight through the mechanical centerline of the element. This line is, by definition, the optical axis of the element. On the right is an element with a severe tilt on its top surface. Note that this surface tilt or element wedge results in an edge thickness difference from left to right (the edge is thinner on the left, thicker on the right). The wedge, in radians, is the edge thickness difference, 2δ , divided by the element diameter, d . Thus, the wedge, in radians, equals $2\delta/d$.

Tolerancing and Producibility**Figure 16.4**

Left: Nominal Element; Right: Wedged Element

As we have shown, a lens element with spherical surfaces will, by definition, have a single optical axis, and it is the process of lens centering which will bring the mechanical edge of the element to be concentric with the optical axis of the lens. Figure 16.5 shows two mechanical methods. In Fig. 16.5a the lens is located on a precision spindle so that its lower surface runs true. Spinning the element will result in a large total indicator runout (TIR) reading on the dial indicator. Moving the element to the left in its shown rotational position will enable the top surface to run true, and the element is then edged using a diamond wheel. Figure 16.5b shows a method of centering called "cupping," in which the upper and lower rings (similar to the edge on a cup) will only fully contact the lens surfaces when the two surfaces are on a common and vertical axis. When this alignment condition is achieved, the element is edged to the proper diameter.

In many of today's optical shops techniques using HeNe lasers are often used to aid in centering lens elements. With one surface running

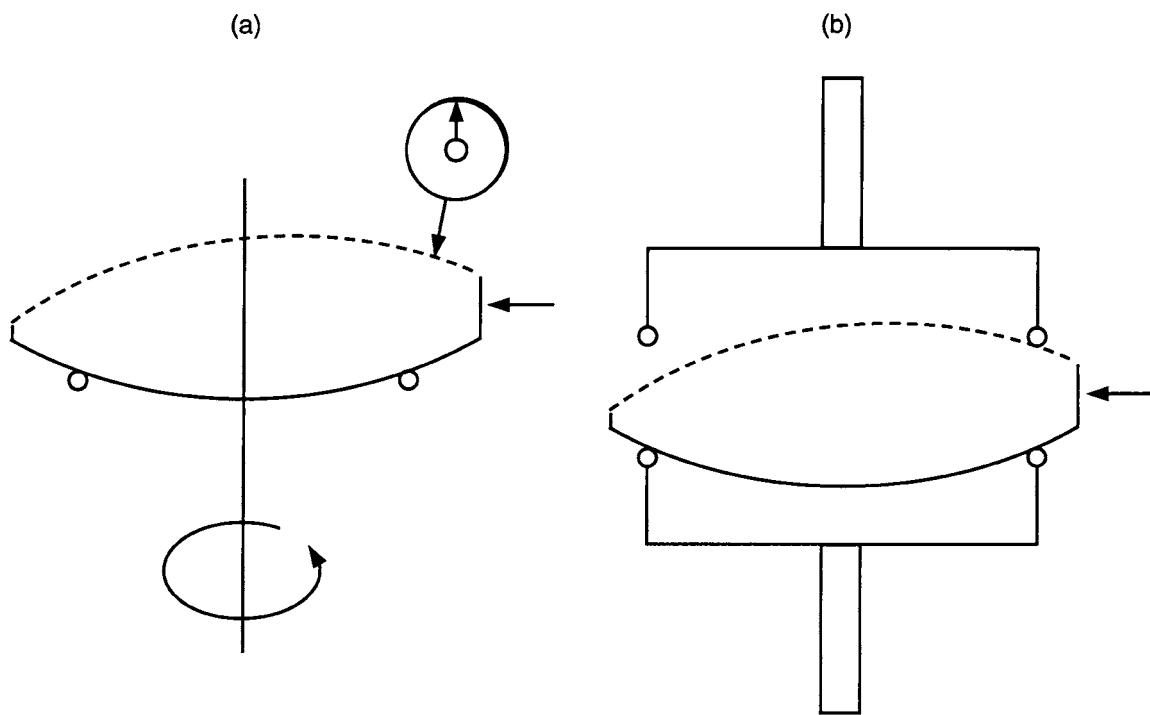


Figure 16.5

Left: Centering on a Spindle; Right: by Cupping

Tolerancing and Producibility

true, the laser beam is either reflected off of the other surface or transmitted through the element, while the element is being rotated. The laser beam will nutate, or "wobble," until the element is properly centered. By using sensors such as CCDs or quad cells, extremely accurate indications of the element centration can be made.

Element decentration can be either a simple lateral decenter (up and down) or it can be a "roll" whereby the element maintains contact with a housing seat, as shown in Fig. 16.6. Note that while the net effect is quite similar, the two decentration models are actually quite different. In the roll situation the left-hand radius, which is in perfect contact with the housing seat, ends up perfectly aligned, with the surface tilt occurring only on the right-hand surface. Many of the computer software packages can model both pure decentrations as well as roll.

Finally, element tilt is self-explanatory, and it is expressed either as a total indicator runout (TIR) or sometimes as the tilt, in minutes of arc.

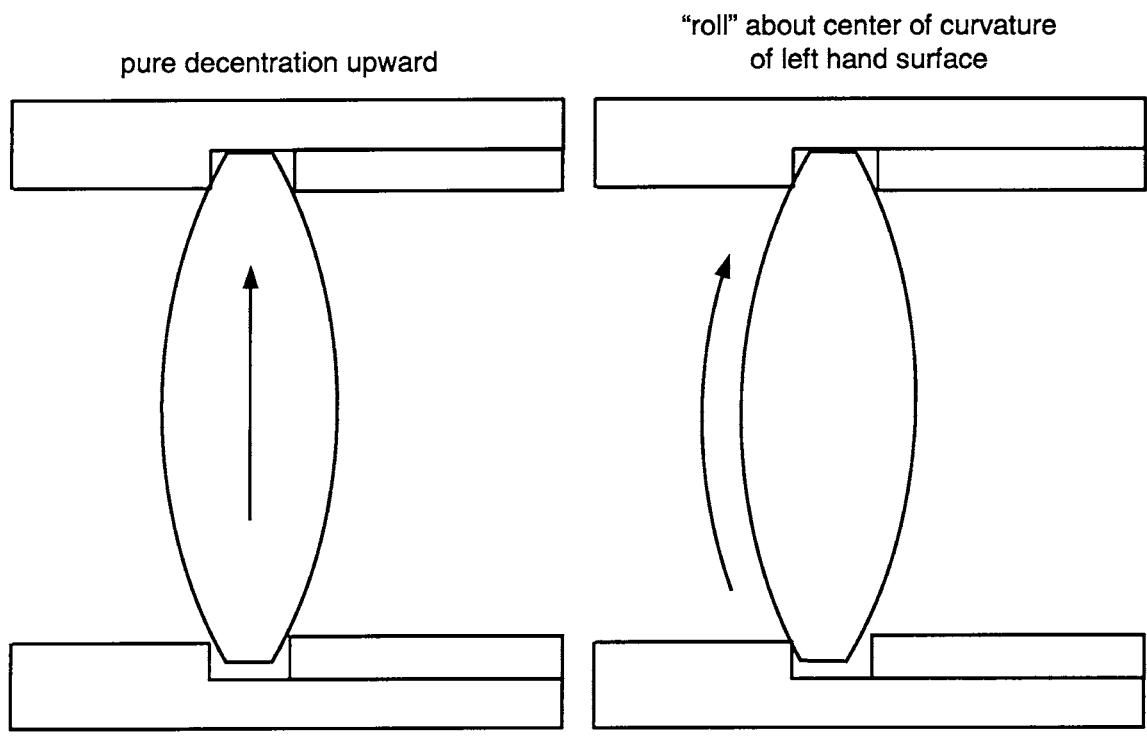


Figure 16.6
Element Decentration Models

Since 1 min of arc is 0.0003 rad, we simply multiply the tilt, in radians, by the diameter to derive the TIR.

Adjusting Parameters

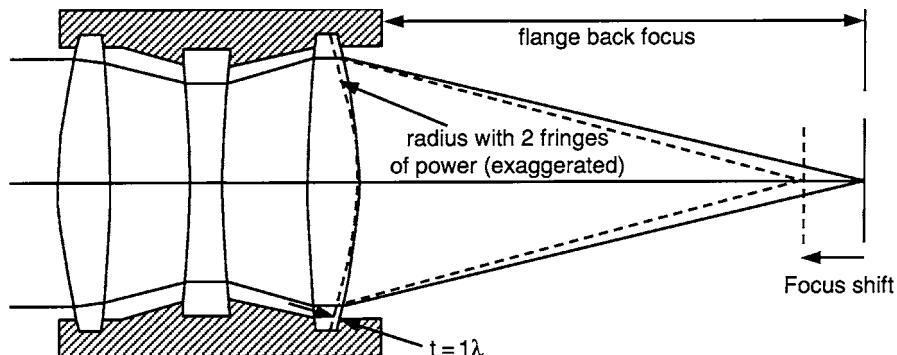
Adjusting parameters, sometimes called *compensators*, are those parameters used to optimize the optical performance in the laboratory or on the assembly line at some time during the final lens assembly and testing.

The most common adjusting parameter is refocusing during the final assembly and testing procedure. Let's consider the manufacturing of a 35-mm camera lens or a similar high-quality lens for a CCD camera. The hypothetical example shown in Fig. 16.7 illustrates the situation. The lens will be mounted to the camera using a bayonet or similar accurate mounting methodology, and the distance from the rear flange to the image plane (the film or the CCD chip) is a tightly held mechanical dimension. Needless to say it is imperative that when the completed lens is fastened to the flange on the camera body, the image of an infinite object is in perfect focus on the film or CCD when the lens is focused at infinity.

Referring to Fig. 16.7, we show a nominal lens imaging onto a sensor at the nominal flange back-focal distance. We also show in dashed lines the last radius of the lens with a steeper or more powerful radius than nominal by two fringes of power. The effect of this is to move the image inward, and the image will be out of focus on the sensor.

The optical path difference (OPD) = $(n - 1)t$, where n = the refractive index and t = the separation between the nominal surface and the manufactured surface at the edge of the aperture as shown. If we have two

Figure 16.7
Refocusing As an
Adjusting Parameter



fringes of power, this is due to one wave of surface error, or sag, at the edge of the element. And since $OPD = (n - 1)t$, and $t = 1.0\lambda$, the $OPD = 0.5\lambda$ peak to valley. If the flange back-focus distance is absolutely correct for the nominal camera and for the nominal lens assembly, the image will be out of focus by 0.5λ peak to valley, which is a factor of 2 from the Rayleigh criteria, due *only* to the last radius being two fringes of power from nominal!

The entire lens in this simplified example of course has six radii as well as the two airspaces, element thicknesses, and other tolerances, all of which will further contribute to defocus errors. If we RSS the effect of the six radii being two fringes of power each, we find that the final image could be out of focus by 1.22 waves of defocus, which is of major significance. What this means is that if we consider only the power fit to testplate, then with a tolerance of two fringes per surface, we predict a focus error of 1.22 waves of defocus with a 95% confidence level. This is about 5 times the Rayleigh criteria, and the imagery will be poor relative to the diffraction limit. We *must* refocus the lens, and this is accomplished by using an adjusting parameter, in this case the back-focus distance, during final assembly.

It is not uncommon to sometimes use an element decentration as an adjusting parameter. This is done when the system is quite sensitive to what is sometimes called *axial coma*, or coma which occurs on axis (and carries somewhat uniformly across the field of view) due to asymmetrical tolerances such as element wedge, tilt, and decentration. By allowing a strategically located element within the system to be adjusted in *X* and *Y* during the final assembly and testing, you can often cancel the coma introduced by all of the other tolerances within the system. Not only will this permit you to produce a system which otherwise may not work properly, but you may be able to relax some of your other tolerances from their otherwise tight levels, thus lowering the cost and enhancing the producibility. The trade-off often results in producing a system with extremely tight tolerances versus looser tolerances and one adjustment made during final assembly and testing.

In large field-of-view projection optics, it is sometimes imperative to have a lens or a group of lenses that are adjusted in centration during final assembly of the system in order to relax the tolerances on the other components, and have a reasonable cost to the system.

In microscope objective manufacturing a common test is the "star test" discussed in Chap. 15. In some high-precision shops, if the objective shows axial coma, it is sometimes tapped or "banged" lightly on a table and retested one or more times until the performance is met. While somewhat of a "brute force" method of compensator usage, it does work.

Typical Tolerances for Various Cost Models

There have been several papers presented over the years showing the effect on cost associated with various levels of tolerances. One of these, presented by John Plummer, is shown in Fig. 16.8. Note that there is a similar updated table in Chap. 17. You may find it interesting to compare

diameter tolerance in millimeters	± 0.1 100	± 0.05 100	± 0.025 103	± 0.0125 115	± 0.0075 150		
center thickness tolerance in mm	± 0.2 100	± 0.1 105	± 0.05 115	± 0.025 150	± 0.0125 300		
stain charac. of the glass	0 100	1 100	2 103	3 110	4 150	5 250	5+ 500
# of lenses per block 100	25 105	18 115	11 130	6 175	3 300	1	
eccen. Toler. in light dev.	6 min. 100	3 min. 103	2 min. 108	1 min. 115	30 sec. 140	15 sec. 200	
Figure tolerance in λ (pow/irreg.)	10-5 100	5-2 105	3-1 120	2 1/2 140	2 1/4 175	1 1/8 300	
Dia. to thick. ratio (fig. 3.1)	9-1 100	15-1 120	20-1 150	30-1 200	40-1 300	50-1 500	
Beauty defects (MIL-C-13830A)	80-50 100	60-4 110	40-30 125	20-10 175	10-5 350		
Raw glass cost in 1000 lb lots	\$3.00 100	\$5.00 108	\$8.00 115	\$15.00 125	\$25.00 135	\$50.00 200	\$100.00 350
Coating speci- fications	unctd. 100	Mg. Fl. 115	3-4 layer 150	>4 layers 200-500			

Figure 16.8

Relative Cost of Manufacturing As a Function of the Level of Tolerance

the two sets of data. Plummer's company at the time, was primarily involved in high production of reasonable quality optics which we might call "riflescope quality" in other words not ultratight tolerances and not "loosey goosey" either. For 10 manufacturing parameters, the level of tolerance is shown on the first line and the relative cost on the second line for various levels of tolerances. Let's look at several examples:

- The standard diameter tolerance is ± 0.05 mm. If we require ± 0.025 , the cost increase is only about 3%; however, if we need ± 0.01 , then the cost increase is about 25 to 30%.
- Meeting tight tolerances for element thicknesses is difficult and costly. Achieving a tolerance of ± 0.05 costs 15% above the standard level of ± 0.2 , and achieving ± 0.025 will cost an additional 50%.
- Even stain characteristics on glass will affect cost. The reason for this is that a glass type with a 5 stain will acquire a stain within minutes, and for this reason the elements must be sent almost immediately into the vacuum chamber for coating after they have been polished. This, of course, affects the workflow in the shop, and for this reason it becomes costly.

You should take the time to read through the table in Fig. 16.8 and become familiar with the various tolerances and their relative cost versus quality trade-offs. More in-depth material relating to optical manufacturing and tolerances is presented in Chap. 18.

Example of Tolerance Analysis

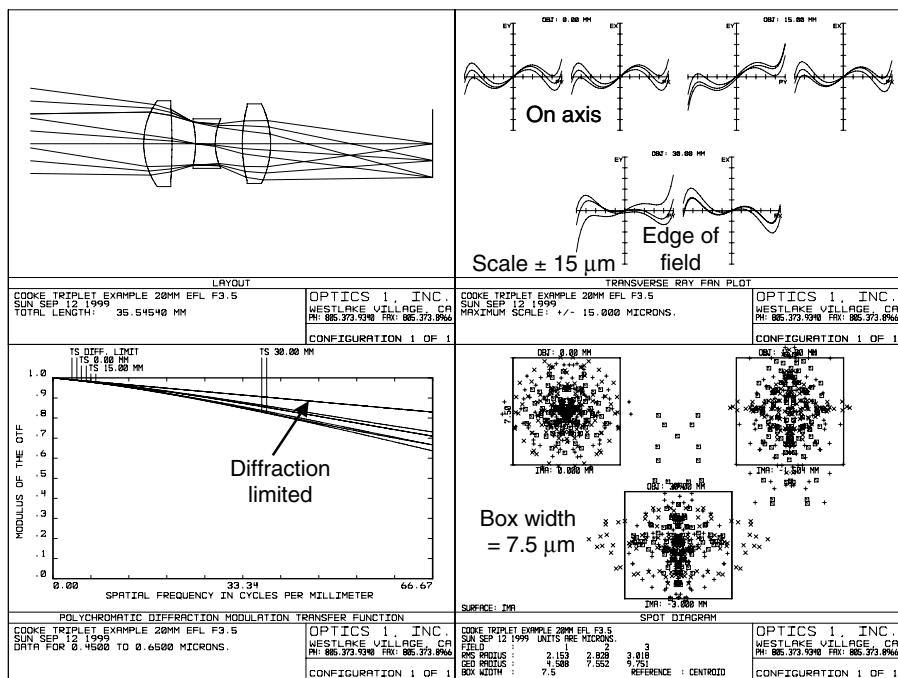
In order to best show how to tolerance a lens system, we will go through the tolerances for a $10\times$ reduction lens Cooke triplet for a machine vision application. Our basic specifications are as follows:

Parameter	Specification
Object distance	200 mm
Object full diagonal	60 mm
Magnification	$0.1\times$
Image full diagonal	6 mm
f/number	f/3.5 at used conjugate

Parameter	Specification
Focal length	≈ 20 mm
Full field of view	$\approx 17.06^\circ$
Spectral band	550 to 650 nm, uniform weights
Sensor	1/3-in CCD
Number of pixels	640×480
Pixel pitch at image	$7.5 \times 7.5 \mu\text{m}$
Pixel pitch at object	$75 \mu\text{m}$
Nyquist frequency at image	66.67 line pair/mm
Nyquist frequency at object	6.667 line pair/mm
MTF spec at image	>0.3 at Nyquist

Figure 16.9 shows the layout and performance of the final design for this lens. Note that the design is quite close to diffraction limited as evidenced from the MTF. The geometrically based spot diagrams

Figure 16.9
Cooke Triplet for
Machine Vision
Application



Tolerancing and Producibility

show square boxes measuring $7.5 \times 7.5 \mu\text{m}$, which is one pixel at the CCD sensor. Figure 16.10 is a listing of the specifications and prescription data for the lens in case you want to set it up and work with it.

We will now apply a standard set of manufacturing tolerances to the lens. We show in the following the resulting tolerance sensitivities. In order to conserve space, data will be shown for the central element only (this is for most of the tolerances, the most sensitive). In the data, we see in the “Field” column “All,” which is the average of all of the fields of view. Below that are each of the individual fields (field 1 is on axis, 2 and 3 are $\pm 70\%$ of the full diagonal, and 4 and 5 are \pm full field). The “MTF” columns are the MTF after applying the tolerance and refocusing for compensation (by the “Change in Focus” value), and the “Change” columns are the change in MTF

Figure 16.10
Prescription of Cooke
Triplet for Machine
Vision Application

System/Prescription Data					
GENERAL LENS DATA:					
Surfaces	:	8			
Stop	:	4			
System Aperture	:	Image Space F/# = 3.5			
Eff. Focal Len.	:	19.00051 (in image space)			
Back Focal Len.	:	12.4271			
Total Track	:	35.48153			
Image Space F/#	:	3.5			
Stop Radius	:	1.777314			
Parax. Ima. Hgt.	:	3.018941			
Parax. Mag.	:	-0.1006314			
Entr. Pup. Dia.	:	5.428717			
Entr. Pup. Pos.	:	15.4601			
Exit Pupil Dia.	:	6.196183			
Exit Pupil Pos.	:	-23.54149			
Field Type	:	Object height in Millimeters			
Maximum Field	:	30			
Primary Wave	:	0.55			
Lens Units	:	Millimeters			
SURFACE DATA SUMMARY:					
Surf	Type	Radius	Thickness	Glass	Diameter
OBJ	STANDARD	Infinity	190		60
1	STANDARD	Infinity	10		0
2	STANDARD	6.572511	2.5	SK5	7.303856
3	STANDARD	-51.05087	2.041158		6.332172
STO	STANDARD	-8.89512	1.75	SF2	3.696617
5	STANDARD	5.048865	2.408424		4.172604
6	STANDARD	12.99962	2.5	SK5	6.31239
7	STANDARD	-8.044117	14.28195		6.824242
IMA	STANDARD	Infinity			6.020948

In the following data, the tolerance designations are:

- TFRN is the number of fringes of power fit to the testplate.
- TTHI is the airspace or element thickness, in millimeters.
- TEDY is the decentration of an element, in millimeters.
- TETY is the tolerance of element tilt, in degrees.
- TIRY is the tolerance on the total indicator runout or surface tilt for wedge or element centration.
- TIRR is the tolerance on surface irregularity in fringes.

Type	Sf1	Sf2	Field	Minimum			Maximum			
				Value	MTF	Change	Value	MTF	Change	
TFRN	3	All	-4.00	0.713	-0.027	4.00	0.710	-0.031	4 fringes power	
	1		0.739	-0.023			0.724	-0.038	to testplate	
	2		0.659	-0.060			0.726	0.006		
	3		0.659	-0.060			0.726	0.006		
	4		0.751	0.008			0.681	-0.061		
	5		0.751	0.008			0.681	-0.061		
Change in Focus				:	0.126				-0.124	
TFRN	4	All	-4.00	0.732	-0.008	4.00	0.727	-0.013	4 fringes power	
	1		0.754	-0.008			0.759	-0.003	to testplate	
	2		0.731	0.011			0.688	-0.031		
	3		0.731	0.011			0.688	-0.031		
	4		0.714	-0.028			0.740	-0.002		
	5		0.714	-0.028			0.740	-0.002		
Change in Focus				:	-0.096				0.097	
TTHI	3	4	All	-0.050	0.720	-0.020	0.05	0.720	-0.021	thickness
	1		0.764	0.001			0.761	-0.001		
	2		0.746	0.026			0.679	-0.040		
	3		0.746	0.026			0.679	-0.040		
	4		0.662	-0.080			0.726	-0.016		
	5		0.662	-0.080			0.726	-0.016		
Change in Focus				:	0.045				-0.045	
TTHI	4	6	All	-0.050	0.729	-0.011	0.050	0.732	-0.008	airspace
	1		0.763	0.000			0.762	-0.000		
	2		0.694	-0.025			0.736	0.015		
	3		0.694	-0.025			0.736	0.015		
	4		0.736	-0.006			0.702	-0.040		
	5		0.736	-0.006			0.702	-0.040		
Change in Focus				:	-0.042				0.042	
TEDX	3	4	All	-0.025	0.607	-0.134	0.025	0.607	-0.134	element
	1		0.633	-0.129			0.633	-0.129	decenter x	
	2		0.587	-0.132			0.587	-0.132		
	3		0.587	-0.132			0.587	-0.132		
	4		0.602	-0.140			0.602	-0.140		
	5		0.602	-0.140			0.602	-0.140		
Change in Focus				:	0.000				0.000	
TEDY	3	4	All	-0.025	0.589	-0.151	0.025	0.589	-0.151	element
	1		0.633	-0.129			0.633	-0.129	decenter y	
	2		0.550	-0.169			0.607	-0.112		
	3		0.607	-0.112			0.550	-0.169		

			4	0.563	-0.179	0.559	-0.183
			5	0.559	-0.183	0.563	-0.179
				:	0.000		0.000
			Change in Focus				
TETX	3	4	All	-0.150	0.584	-0.157	0.150
			1	0.757	-0.005	0.757	-0.005
			2	0.703	-0.016	0.477	-0.242
			3	0.477	-0.242	0.703	-0.016
			4	0.417	-0.325	0.532	-0.210
			5	0.532	-0.210	0.417	-0.325
				:	0.000		0.000
TETY	3	4	All	-0.15	0.661	-0.080	0.150
			1	0.757	-0.005	0.757	-0.005
			2	0.648	-0.071	0.648	-0.071
			3	0.648	-0.071	0.648	-0.071
			4	0.598	-0.144	0.598	-0.144
			5	0.598	-0.144	0.598	-0.144
				:	0.000		0.000
TIRX	3	All	All	-0.015	0.566	-0.174	0.015
			1	0.638	-0.124	0.638	-0.124
			2	0.552	-0.167	0.552	-0.167
			3	0.552	-0.167	0.552	-0.167
			4	0.518	-0.224	0.518	-0.224
			5	0.518	-0.224	0.518	-0.224
				:	0.000		0.000
TIRY	3	All	All	-0.015	0.570	-0.170	0.015
			1	0.638	-0.124	0.638	-0.124
			2	0.629	-0.090	0.508	-0.211
			3	0.508	-0.211	0.629	-0.090
			4	0.482	-0.260	0.555	-0.186
			5	0.555	-0.186	0.482	-0.260
				:	0.000		0.000
TIRX	4	All	All	-0.015	0.683	-0.058	0.015
			1	0.736	-0.026	0.736	-0.026
			2	0.664	-0.055	0.664	-0.055
			3	0.664	-0.055	0.664	-0.055
			4	0.654	-0.088	0.654	-0.088
			5	0.654	-0.088	0.654	-0.088
				:	0.000		0.000
TIRY	4	All	All	-0.015	0.627	-0.114	0.015
			1	0.736	-0.02	0.736	-0.026
			2	0.708	-0.011	0.539	-0.180
			3	0.539	-0.180	0.708	-0.011
			4	0.507	-0.235	0.604	-0.138
			5	0.604	-0.138	0.507	-0.235
				:	0.000		0.000
TIRR	3	All	All	-1.000	0.636	-0.104	1.000
			1	0.669	-0.093	0.651	-0.111
			2	0.554	-0.165	0.721	0.001
			3	0.554	-0.165	0.721	0.001
			4	0.704	-0.038	0.595	-0.147
			5	0.704	-0.038	0.595	-0.147
				:	0.000		0.000
TIRR	4	All	All	-1.00	0.709	-0.032	1.000
			1	0.719	-0.043	0.730	-0.032
			2	0.748	0.028	0.624	-0.095
			3	0.748	0.028	0.624	-0.095
			4	0.665	-0.077	0.730	-0.012
			5	0.665	-0.077	0.730	-0.012
				:	0.000		0.000

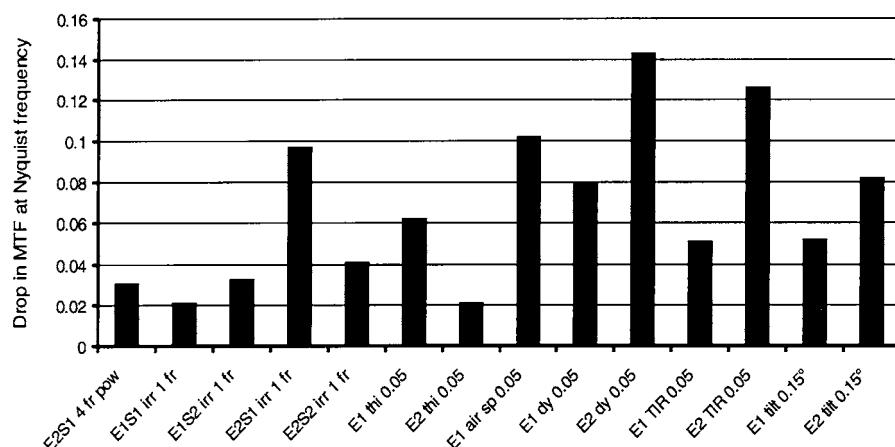
The preceding are the tolerance sensitivities for the central element. There are many numbers... what do they all mean?

At this point, you should look over all of the sensitivities to see if any are especially sensitive. For example, decentration of the central element by 0.025 mm drops the average MTF about 0.134 for an *x* decentration (in and out of the figure) and 0.151 for a *y* decentration. The effect over the field is reasonably uniform (this is not always the case). A less sensitive tolerance is the four fringes of power on surface 4 (the rear of the second element), where a maximum MTF drop of only 0.013 average (0.031 maximum drop) has resulted. We should be able to loosen this tolerance if our shop feels it is worthwhile.

In Fig. 16.11 we show in a graphical form those parameters which drop the average MTF over the field of view by 0.02 or more for our lens. There are many tolerances not even represented in the data. Note that there are only about five to six tolerances that are most sensitive, and these are primarily tilts, decentrations, and wedges. The most sensitive tolerances are for element 2, as discussed earlier.

The best measure of the overall lens performance and the most reliable means for predicting performance is via the Monte Carlo analysis. Here we computed 20 Monte Carlo samples and the resulting statistics are shown in Table 16.2. Recall that each Monte Carlo sample is, in effect, a simulated fabricated system. Each individual parameter is changed according to a normal probability distribution between its minimum and maximum values.

Figure 16.11
All Cooke Triplet Tolerances which Drop the MTF by >0.02



Tolerancing and Producibility

TABLE 16.2

Monte Carlo Results	Average	Field				
		1	2	3	4	5
Nominal	0.741	0.762	0.720	0.720	0.742	0.742
Best	0.602	0.760	0.664	0.649	0.616	0.641
Worst	0.326	0.360	0.297	0.327	0.196	0.165
Mean	0.491	0.585	0.484	0.498	0.444	0.410
Standard deviation	0.068	0.108	0.101	0.086	0.105	0.107
Compensator Statistics						
Change in back focus:						
Minimum		−0.232282				
Maximum		0.222737				
Mean		0.031034				
Standard deviation		0.138210				

Ninety percent of Monte Carlo lenses have an MTF above 0.306.

Fifty percent of Monte Carlo lenses have an MTF above 0.393.

Ten percent of Monte Carlo lenses have an MTF above 0.556.

The Monte Carlo results show that 90% of our lenses should have an MTF at 66.7 line pairs/mm of 0.306 or better. This just meets our MTF goal of 0.3 or better! It still may be beneficial to review each tolerance to see if it can be loosened, and the most sensitive tolerances should be tightened if possible. Once these changes have been made, another Monte Carlo analysis is in order as this is our best way of modeling the predicted lens performance.

It appears that 20 Monte Carlo samples is a small number to represent accurately the statistics of assembling lenses in production. So, we ran 500 Monte Carlo samples with the same tolerances. Here 90% of the lenses are predicted to have an MTF at Nyquist greater than or equal to 0.243. While this result is, indeed, slightly less than 0.306 for 20 samples, it does tell us that 20 samples give a reasonably accurate answer. Indeed, we will need to tighten some of the sensitive tolerances after all.

The compensator is the back focus, and a total range of ± 0.23 mm was encountered with a standard deviation of 0.14 mm.

Surface Irregularities

In Chap. 4 we discussed the concept of optical path difference and its influence on image quality. One of the most important and influential rules of thumb is the Rayleigh criteria which tells us that if the peak-to-valley optical path difference is less than or equal to one-quarter wave, then the image quality will be nearly indistinguishable from perfect diffraction-limited performance.

Optical path difference (OPD) can be introduced by the following factors:

- The fundamental *aberrations present in the basic lens design* such as spherical aberration, coma, astigmatism, defocus, and other image-degrading aberrations.
- *Assembly and alignment errors*. Included in this category are the various tolerance forms such as element thicknesses, airspaces, wedges, tilts, decentrations, and other tolerance types.
- *Environmental effects* such as thermal soaks and/or gradients.
- *Effects external to the system* such as atmospheric turbulence.
- *Surface irregularities and other wavefront errors not included in the previous errors*. Included here are the residual manufacturing errors, which cause a surface to deviate or depart from its nominal shape. Most often, this is a deviation from sphericity or flatness, but it can, of course, also be a deviation from a prescribed aspheric profile if the nominal surface is aspheric.

The basic lens design performance residual, which is due to the various forms of aberrations, will consist of the orders of aberration present in the design such as defocus, spherical aberration, coma, astigmatism, field curvature, and the chromatic aberrations of axial and lateral color which are changes in the basic aberrations with wavelength. Thus, the design of a double gauss camera lens may have a mix of third, fifth, and higher orders of aberration, both on axis with spherical aberration, as well as off axis with other aberrations.

Recall that the OPD polynomial is one exponent higher than the transverse ray aberrations. In other words, third-order spherical aberration affects the wavefront proportional to the fourth power of the aperture, third-order astigmatism is linear with aperture, and the effect to the wavefront is quadratic with aperture, and so on. The net

result is that the basic lens design will have a mix of all of the residual aberrations present in the prescription. Fortunately, lens design software programs are quite robust and can model these aberrations and accurately predict the MTF and other measures of performance. The optical path difference introduced by our design prescription is accurate and quantifiable.

The optical path difference introduced by assembly and alignment errors include symmetrical errors such as power fit to testplates, element thicknesses, and airspaces, as well as asymmetrical errors such as element wedge, tilt, and decentration. In most cases, the effects of these tolerances will be the introduction of primarily low-order aberration and, to a reasonable extent, this will be true throughout the system. For example, the resulting effect of the designated number of fringes of the power fit to testplate will, to a very large degree, contribute to the on-axis image defocus and, to a lesser extent, third-order spherical aberration. The axial image will also show the effects of the asymmetrical tolerances, including wedge, tilt, and decentration. These perturbations will contribute primarily low-order coma and astigmatism to the axial image. Similarly, off axis we will see the introduction of mostly lower-order aberrations due to the manufacturing and alignment errors.

The results of environmental effects, such as thermal soaks and gradients, will also be primarily low order.

This brings us to *surface irregularities*, which we will divide into several categories:

- *Conventional optical manufacturing* Here we are talking about spherically surfaced lenses in approximately the 10- to 30-mm-diameter range manufactured in a production environment using conventional machinery. For longer radii, elements are blocked for cost economies. The residual surface irregularities for these elements are primarily cylindrical in form. Thus, the surface is literally a toroid which is best thought of as a cylindrical departure from sphericity. Clearly, to the axial imagery, this will introduce astigmatism, and the aberration is of low order.
- *Larger elements*. As lens elements become larger, the surface irregularities tend to depart some from the classical cylindrical shape. We can have asymmetries, which can in many cases become more highly asymmetric and nondescript and less well correlated.

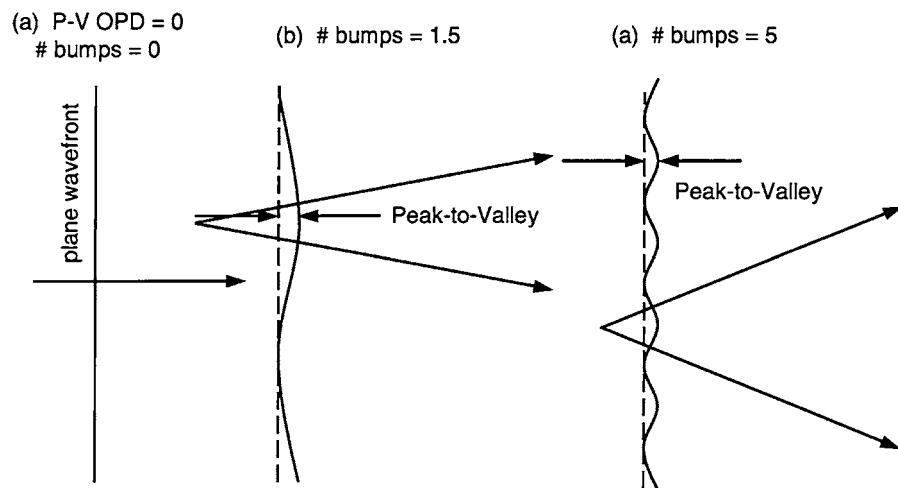
- *Larger surfaces such as telescope mirrors.* Whenever larger surfaces are involved, especially if there are aspheric surfaces, we often encounter much higher-order surface irregularities. These effects can, for example, be due to the manufacturing process where subdiameter polishing tools are often used. In the process of reaching the desired surface profile (often a paraboloid or similar conic section, sometimes with an intentional higher-order residual), high-frequency irregularities are often left in the surface.
- *Thin lenses, windows, mirrors, and plastic optics.* Very often, we require very thin lenses, windows, and mirrors, and also in this category we have injection-molded plastic lenses (as well as compression-molded glass lenses). We may, for example, use flat glass manufactured by a process called *float glass*. In many of these cases the residual surface irregularity is less straightforward to predict and can often have nonrotational symmetrical residuals from the nominal surface shape.

The primary difference between small, high-production lenses, where a cylindrical departure from sphericity results, and some of the latter examples, such as injection-molded plastic lenses and thin-float glass mirrors or beamsplitters, is that in these latter scenarios we may have both larger departures from the ideal surface profile as well as a lower “correlation” surface due to more “bumpiness” on the surfaces and therefore to the wavefront.

How Does Correlation Relate to Performance?

Consider Fig. 16.12 where we show three sinusoidal wavefront profile models: Fig. 16.12a is plane or flat, Fig. 16.12b has a given peak-to-valley surface irregularity with 1.5 bumps across the surface, and Fig. 16.12c has the same P-V irregularity with five bumps across the surface. The term “correlation,” in very simple terms, is the inverse number of bumps across the surface. Thus, the surface in Fig. 16.12b has a correlation of 0.666 and the surface in Fig. 16.12c is less correlated with a correlation of 0.2.

Figure 16.12
Sinusoidal Wavefront Shapes



If the preceding represents the deviation of the wavefront at the exit pupil from its ideal spherical shape, then multiplying the total angular ray deviation by the focal length of our optical system will give the maximum image blur diameter or extent at the image. The larger the slope errors, the larger the image blur. Unfortunately, these effects are often random and change from part to part or system to system, so some form of modeling and approximation is in order.

Effect to Spot Diameter

Assume that we have a wavefront that departs from perfect flatness or sphericity sinusoidally as shown in Fig. 16.12. Assume the following:

A = the peak-to-valley variation of the wavefront

n = number of bumps over the wavefront extent

D = the total extent of the wavefront (exit pupil diameter)

The maximum slope of the wavefront can be shown to be

$$\text{Maximum slope} = \frac{\pi A \lambda n}{D}$$

Thus, the total angular spread of the rays proceeding to the image is twice the preceding result. If we assume that the rms wavefront error is one-fifth of the peak-to-valley (a reasonable assumption), the angular spot diameter containing 100% of the light is

$$100\% \text{ spot diameter (rad)} = \frac{5\pi\sigma\lambda n}{D}$$

where σ is the rms wavefront error. If we now assume that the energy is uniformly distributed in the image, we can rewrite the previous relationship in its final and most useful form:

$$\text{Spot diameter} = \sqrt{p} \left(\frac{5\pi\sigma\lambda n}{D} \right)$$

where p is the fraction of the total energy. This gives us the diameter of the spot inside of which there is a given percent energy.

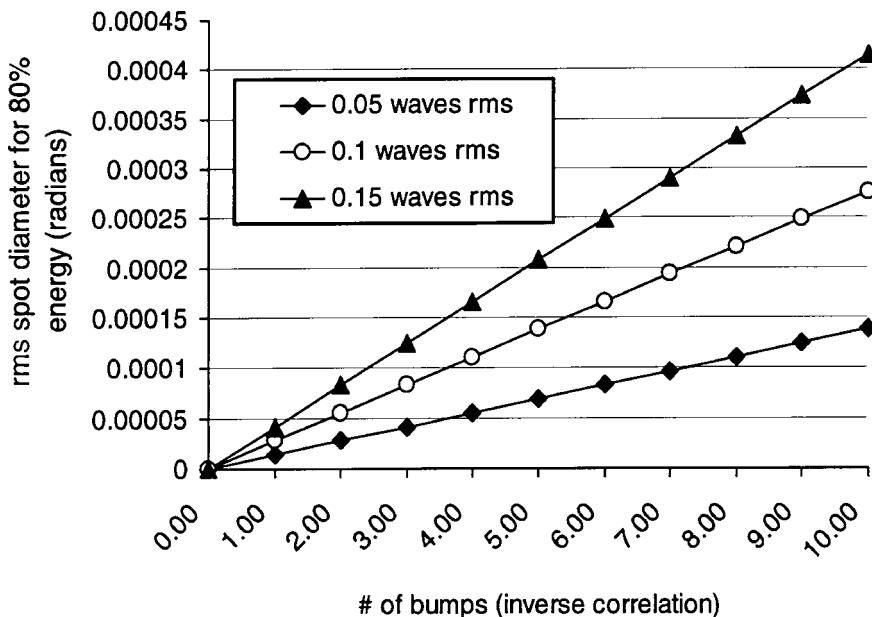
Let's do a quick sanity check. Assume we have a wavefront with three bumps, which just meets the Rayleigh criteria exiting an $f/10$ lens with a 100-mm exit pupil diameter. The rms wavefront error is thus 0.05 wave, and the wavelength is 0.5 μm . Our focal length is 1000 mm. Thus, the diameter of the spot containing 80% of the energy is

$$\begin{aligned} 80\% \text{ spot diameter} &= \sqrt{0.8} \left\{ \frac{5 \times 3.141 \times 0.05 \times 0.5 \times 3}{100,000} \right\} \times 1000 \\ &= 0.0105 \text{ mm} \end{aligned}$$

The diffraction-limited Airy disk diameter for the previous $f/10$ lens is approximately 0.012 mm, which is very close to the derived value of 0.0105 for a situation which should be essentially diffraction limited. As the number of bumps increases or the other parameters change, we can compute the approximate predicted blur diameter. While this derivation is not rigorous, and it is based solely on geometry, it is extremely useful when you really do not know the exact form of the wavefront error but you do have an idea of the correlation and the wavefront error.

Figure 16.13 shows graphically the 80% energy blur diameter, in radians, as a function of rms wavefront error and the number of bumps on the wavefront. These data are based on the derivation shown earlier.

Figure 16.13
Predicted Blur
Diameter As a
Function of rms
Wavefront Error and
Number of Bumps



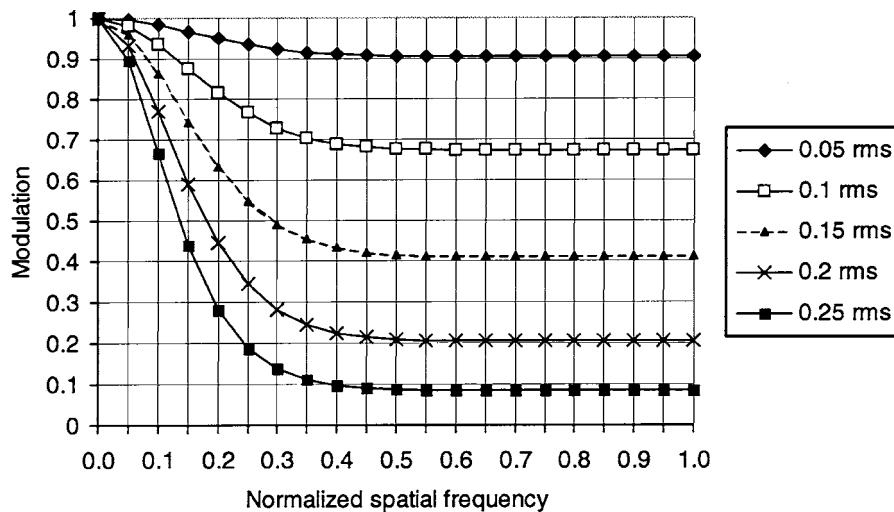
Results smaller than the Airy disk diffraction diameter are fictitious and in these situations the prediction should revert back to the Airy disk diameter.

Effect to MTF: The Optical Quality Factor

A number of years ago, Hufnagle of Perkin-Elmer developed an empirical relationship whereby the MTF degradation can be derived as a function of the rms wavefront error and the correlation of the wavefront. This relationship yields what has become known as the optical quality factor (OQF), and is given by

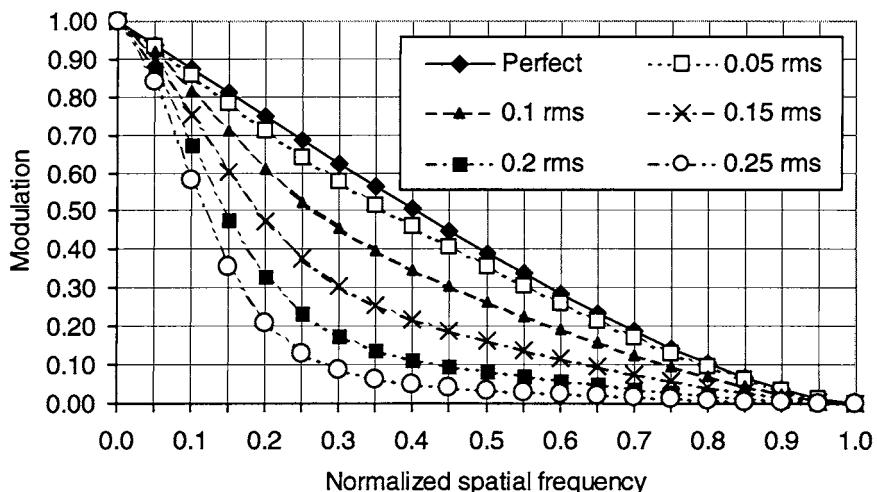
$$\text{OQF} = e^{-(2\pi\text{rms})^2(1 - e^{-2n^2 s^2})}$$

Figure 16.14
OQF As a Function of
rms Wavefront Error
and Number of
Bumps



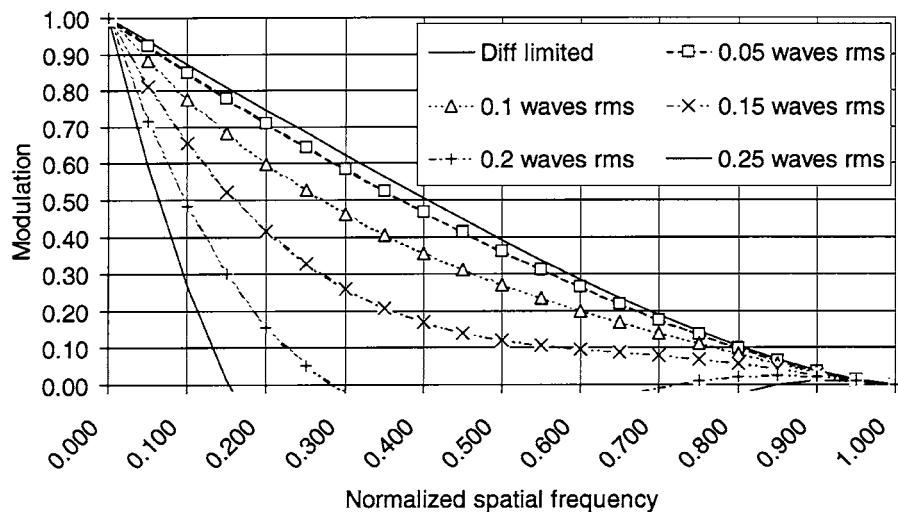
where n is the number of bumps over the exit pupil (the inverse correlation) and s is the normalized spatial frequency relative to cutoff. Figure 16.14 shows the OQF as a function of normalized spatial frequency and rms wavefront error for a correlation of 0.333, or three bumps across the pupil. If we multiply these data by the perfect system MTF we get the results in Fig. 16.15.

Figure 16.15
Predicted MTF As
Function of rms
Wavefront Error and
Number of Bumps



Tolerancing and Producibility

Figure 16.16
MTF Versus rms
Wavefront Error



There is an empirically derived approximation by Shannon which can also be used, and this is shown in Fig. 16.16. The equation for this data is

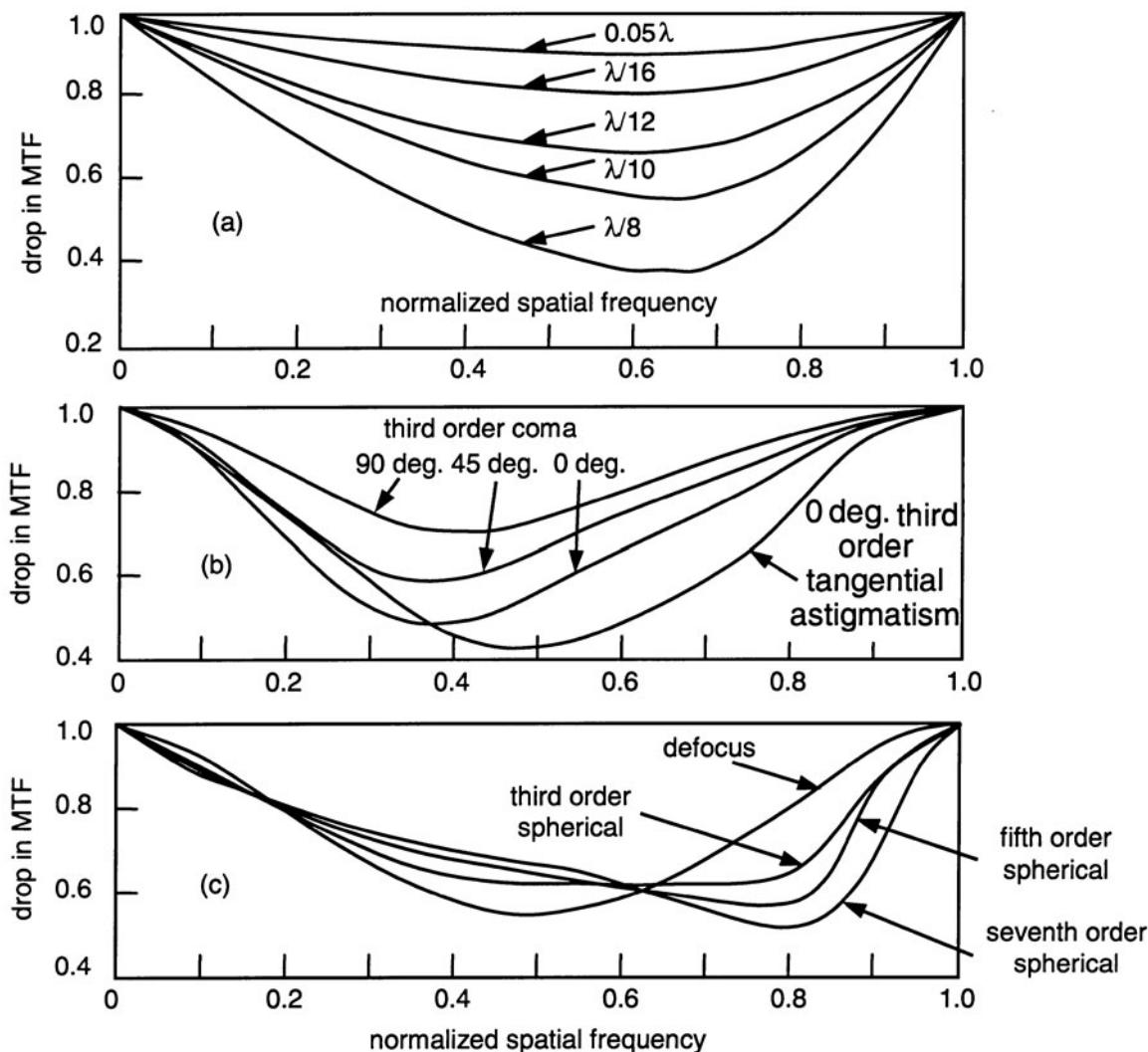
$$\text{MTF}(\nu) = \frac{2}{\pi} \left[\arccos(\nu) - \nu \sqrt{1 - \nu^2} \right]$$

$$\text{ATF}(\nu) = \left\{ 1 - \left[\left(\frac{W_{\text{rms}}}{0.18} \right)^2 \right] [1 - 4(\nu - 0.5)^2] \right\}$$

$$\nu = \frac{\text{spatial frequency}}{\text{cutoff frequency}} = \frac{N}{[1/(\lambda) f/\#]}$$

In this relationship, the diffraction-limited MTF is given by the first equation as a function of the normalized spatial frequency, ν . The equivalent to the OQF described earlier is given by the second equation as a function of the normalized spatial frequency ν and the rms wavefront error.

Another form of MTF degradation, or OQE, is shown in Fig. 16.17a, where we show the effect of a mix of third-, fifth-, and seventh-order aberrations. The wavefront degradations are in the form of rms wavefront error. In Fig. 16.17b and c we show the OQE for 0.1 wave rms of nonrotationally symmetric aberrations and rotationally symmetric

**Figure 16.17**

MTF Drop As a Function of Wavefront Error

aberrations, respectively. It is interesting to note that for 0.1 wave rms the MTF drops to approximately 0.6 of its nominal value, regardless of the form of aberration.

The data in this section can be very useful for predicting the MTF drop due to wavefront errors which you may have derived from a

performance error budget, but you may not know the specific form of the error so you cannot model it directly on your computer program. For example, if you were to have approximately one-half wave P-V of random irregularity due to a mirror in your system, this equates to approximately 0.1 wave rms, and the OQF at midfrequency would be in the order of 0.6. These data allow us to quickly and easily assess and predict the effect of wavefront errors on the optical system.

Beam Diameter and Surface Irregularity

There is one additional important point, which needs to be discussed—how the beam diameter relates to the surface and wavefront irregularity. The critical factor is how much wavefront error is introduced to the wavefront as it proceeds through our system, surface by surface. A parameter that becomes extremely important is the beam diameter or the “footprint” of light at each surface. Consider the lens in Fig. 16.18. Note how almost the entire aperture of the cemented doublet is used regardless of the field position. The element closest to the image is nearly the same diameter as the doublet; however, the beam diameter going to any given image point uses only about 20% of its diameter.

The impact of this is clear from Fig. 16.19a, where we show a simulated interferogram with five waves of astigmatism. Let us assume that an equivalent wavefront error is introduced by the final lens element over its full diameter. Since the beam diameter reaching any point in the field of view is only 20% of the lens diameter, what we have is the

Figure 16.18
Beam Diameter

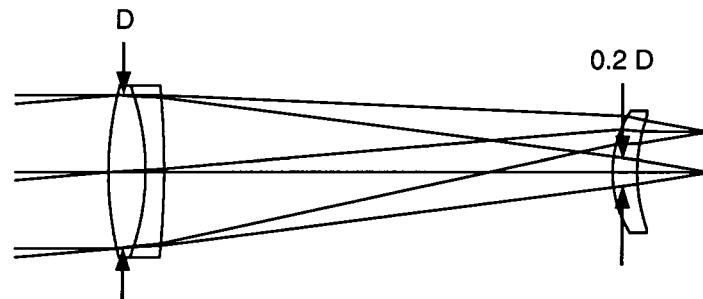
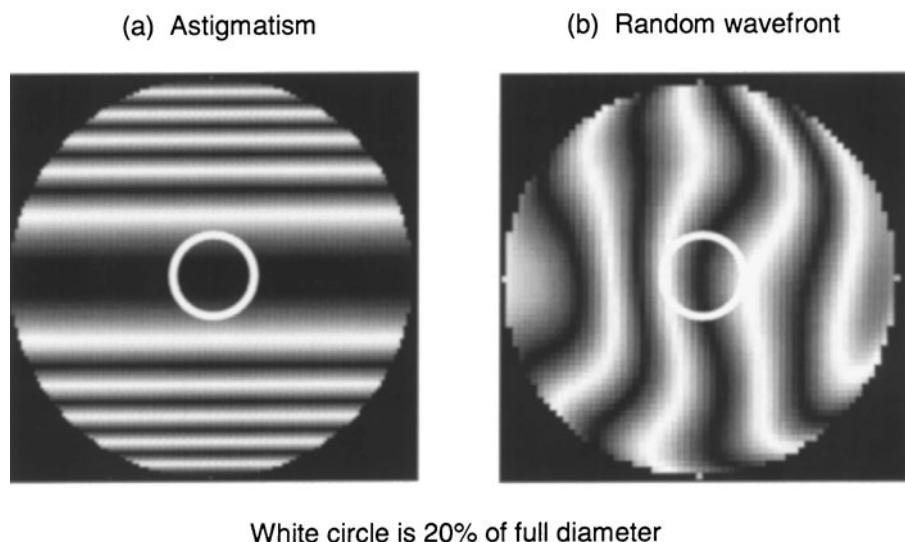


Figure 16.19
Reduction of OPD
for Smaller Beam
Diameters



White circle is 20% of full diameter

effect of the same interferogram but only within the white circle. If we have astigmatism, as shown, the net wavefront error to the system will be reduced from 5 waves to 0.2 wave since the wavefront polynomial for astigmatism is quadratic with aperture, and $0.2^2 = 0.04$, which means that only one-twenty-fifth of the astigmatism is introduced to the wavefront.

Figure 16.19b shows a random wavefront where we can also see that, over a reduced beam footprint, the net wavefront error can be significantly reduced. In this case, the full diameter has a residual of approximately 0.8λ peak-to-valley, and over 20% of its aperture (the white circle) we find approximately 0.02 wave, a reduction of 40 times.

The Final Results

Once we predict the net system degradation, the performance can be shown in various formats. In order to show how these results all compare, consider Figs. 16.20 and 16.21, where we show the performance for a perfect $f/5$ system with no diffraction or aberrations (a), as well as a diffraction-limited system (b), 0.25λ P-V (c), 0.5λ P-V (d), and 0.75λ (e).

Tolerancing and Producibility

Figure 16.20

Image Quality for Different Amounts of Peak-to-Valley Third-Order Spherical Aberration

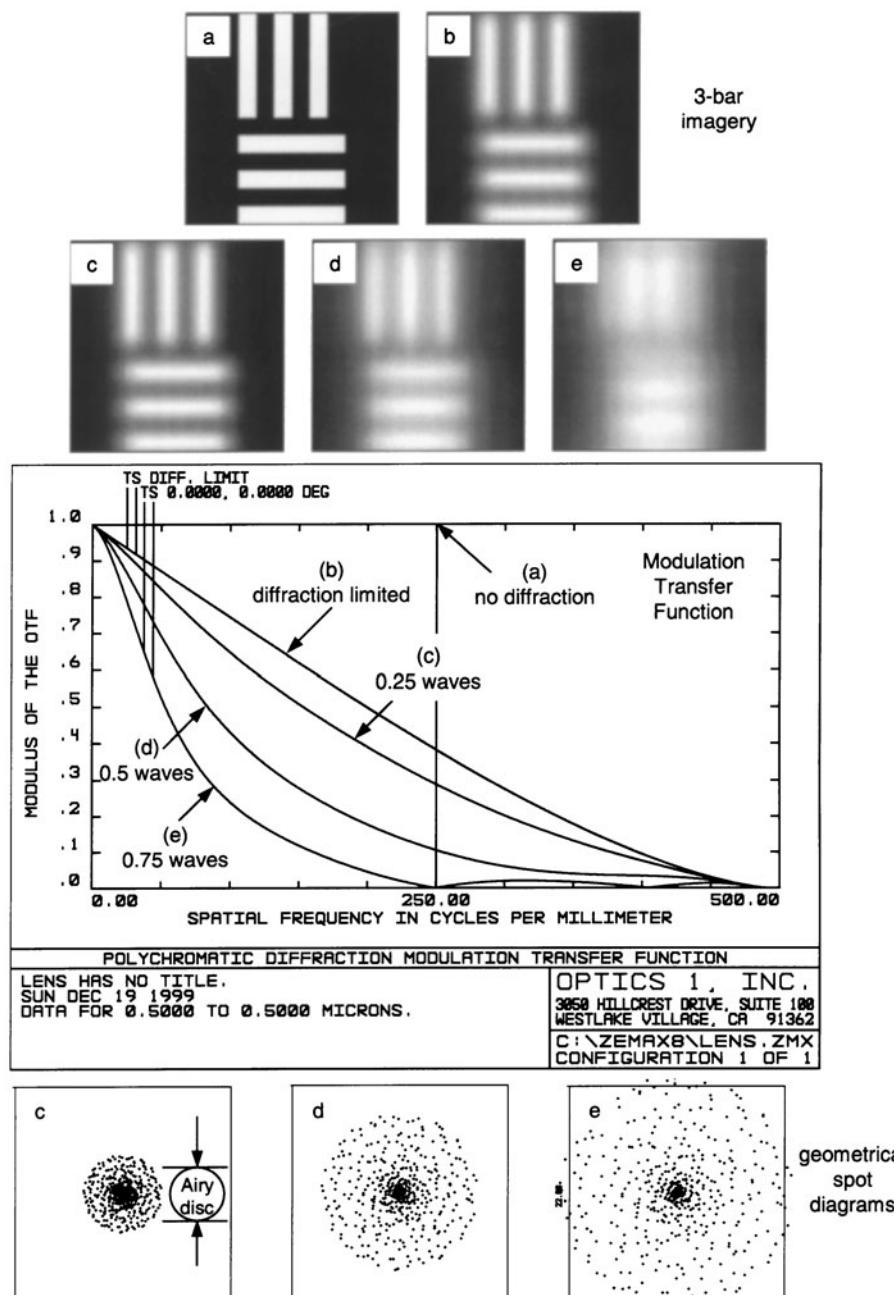
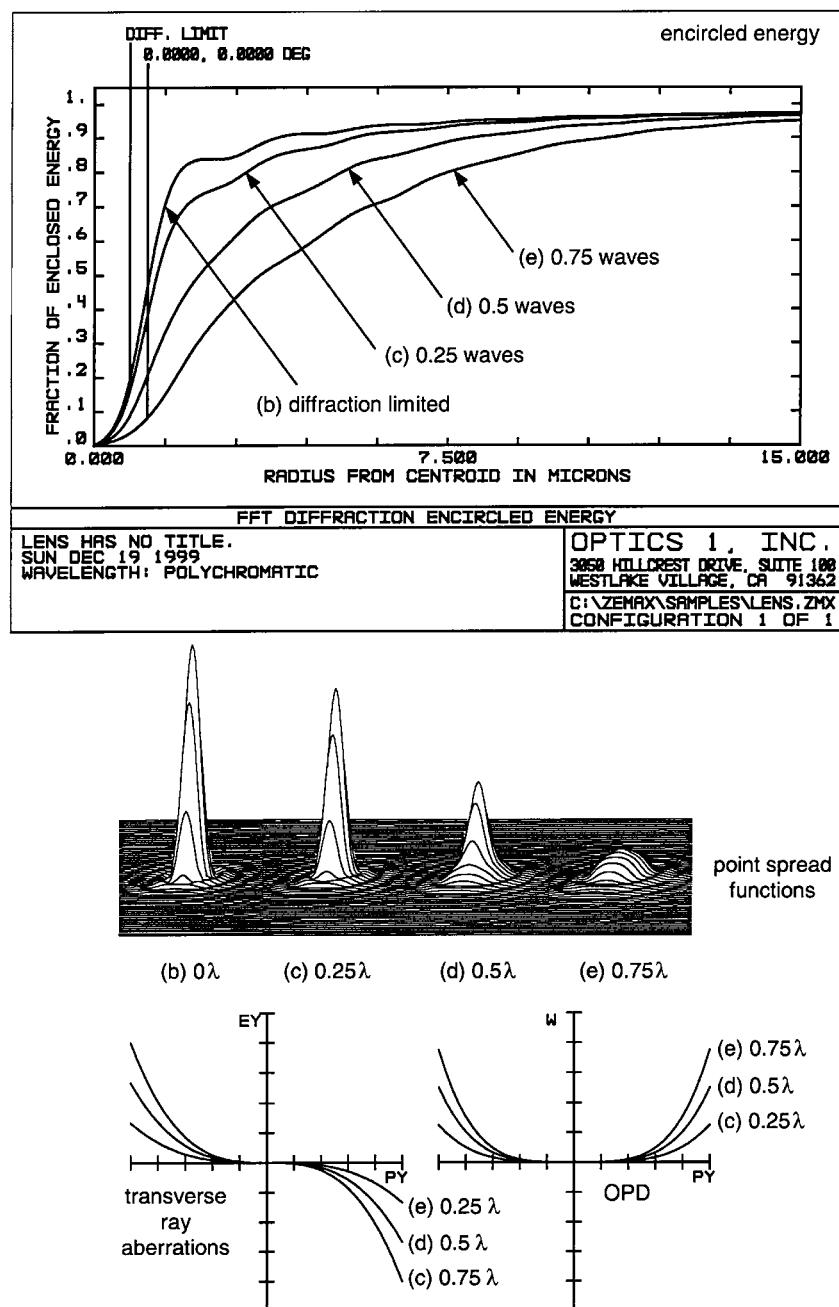


Figure 16.21
Image Quality for
Different Amounts of
Peak-to-Valley Third-
Order Spherical
Aberration



The aberrations are all third-order spherical elements. The following performance metrics are shown in Figs. 16.20 abd 16.21:

- *A graphical representation of a three-bar Air Force-type target whose bar pattern is selected to be at 50% of the cutoff spatial frequency at 250 line pairs/mm.* Note that for these data, as well as the other data presented here, the 0.25-wave situation is only slightly degraded from the diffraction-limited metric, as predicted by the Rayleigh criteria. It is interesting to note that if there were no diffraction effects, and we had an aberration-free system, then the three-bar pattern would image as shown in Fig. 16.20a. This corresponds to a modulation of unity and, for all practical purposes, represents the target for the data in Figs. 16.20 b through d.
- *A plot of the MTF.* The reduction in contrast is quite evident, reaching zero for the 0.75-wave case. Note how the bar pattern imagery correlates with the MTF data.
- *Geometrical spot diagrams, with the Airy disk size shown for reference.* Note in the case of the spherical aberration that we have here somewhat of an intense center to the pattern rather than a uniform intensity pattern which would be evident from pure defocus. As the pattern grows from the Airy disk diffraction pattern diameter, the MTF drops and the contrast degrades, as already shown.
- *A plot of encircled energy.* Here, too, we can see the increasing blur diameters as the spherical aberration increases. As before, the 0.25-wave case is close to diffraction limited, as predicted by Rayleigh.
- *Plots of the point spread function (PSF).* This is one of the more instructive examples of the Rayleigh criteria. The PSF for the perfect system and the 0.25-wave system are very nearly identical in overall appearance. While we do see a drop of intensity at the center of the pattern, the overall appearance of the PSF is nearly the same as for the perfect system. The first ring is still quite evident. However, as soon as we reach 0.5 wave and above, the whole character and appearance of the pattern degrades, and more energy is thrown out from the central maximum of the pattern.

- *Plots of the transverse ray aberrations.* Here we show the transverse ray aberrations for the different amounts of wavefront error. These data are cubic with aperture.
- *Plots of the optical path difference.* Finally we show the optical path difference for the different amounts of wavefront error. These data are proportional to the fourth power of the aperture.

All of the data here are for *precisely* the same system with the same third-order spherical aberration. These data, along with the OQF and other relationships discussed earlier in this chapter, can be used to help predict the performance of imaging optical systems for various amounts of wavefront error.