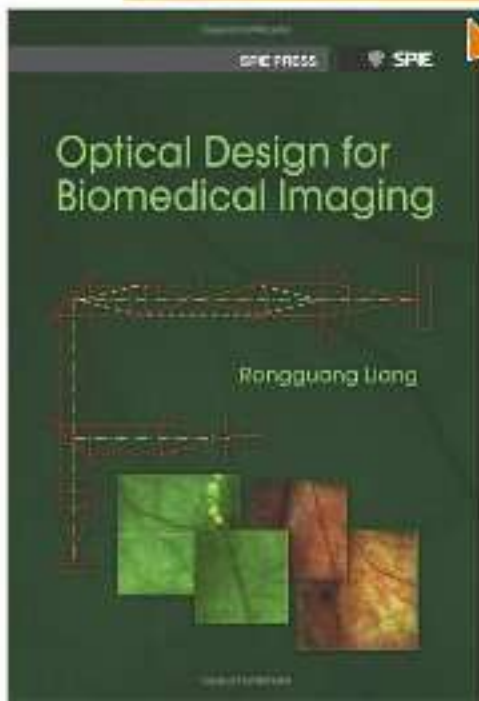


Click to **LOOK INSIDE!**



Optical Design for Biomedical Imaging

Optical Design for Biomedical Imaging

Rongguang Liang

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Liang, Rongguang.

Optical design for biomedical imaging / Rongguang Liang.

p. cm. -- (Press monograph ; 203)

Includes bibliographical references and index.

ISBN 978-0-8194-8369-0

1. Imaging systems in medicine. 2. Optical instruments--Design and construction. I.

Title.

R857.O6L527 2010

616.07'54--dc22

2010039811

Published by

SPIE

P.O. Box 10

Bellingham, Washington 98227-0010 USA

Phone: +1 360.676.3290

Fax: +1 360.647.1445

Email: Books@spie.org

Web: <http://spie.org>

Copyright © 2010 Society of Photo-Optical Instrumentation Engineers

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

The content of this book reflects the work and thought of the author(s). Every effort has been made to publish reliable and accurate information herein, but the publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Printed in the United States of America.



Contents

<i>Preface</i>	xiii
<i>Acknowledgments</i>	xiv
<i>List of Acronyms and Abbreviations</i>	xv

Chapter 1 Introduction to Biomedical Optical Imaging 1

1.1	Introduction	1
1.2	Optical Signals in Biomedical Optical Imaging	2
1.2.1	Optical properties of tissues	2
1.2.1.1	Refractive index.....	2
1.2.1.2	Absorption and absorption coefficients.....	3
1.2.1.3	Scattering and scattering coefficients	3
1.2.1.4	Anisotropy	4
1.2.1.5	Reduced scattering coefficient and penetration depth.....	5
1.2.2	Optical signals	5
1.2.2.1	Reflection, absorption, and scattering	5
1.2.2.2	Polarization effects	7
1.2.2.3	Fluorescence	7
1.2.2.4	Diseases and the optical properties of tissues.....	8
1.2.3	Tissues in optical systems	8
1.3	Biomedical Optical Imaging Techniques	9
References	12

Chapter 2 Introduction to Optical Design..... 13

2.1	Introduction to Optical Design	13
2.1.1	Paraxial optics	13
2.1.2	Newtonian and Gaussian equations.....	15
2.1.3	Stop and pupil.....	16
2.2	Optical Properties	18
2.2.1	Numerical aperture	18
2.2.2	Depth of focus/field.....	19
2.2.3	Lagrange invariant.....	20

2.3	Radiometry of Optical Systems.....	21
2.3.1	Basic concepts.....	21
2.3.2	Radiometric approximations.....	23
2.3.3	Radiometry of optical systems.....	24
2.4	Introduction to Aberrations.....	25
2.4.1	Spherical aberration $W_{040}\rho^4$	29
2.4.2	Coma $W_{131}H\rho^3\cos\theta$	31
2.4.3	Astigmatism $W_{222}H^2\rho^2\cos^2\theta$	34
2.4.4	Field curvature $W_{220}H^2\rho^2$	36
2.4.5	Distortion $W_{311}H^3\rho\cos\theta$	37
2.4.6	Chromatic aberrations.....	39
2.5	Seidel Aberrations.....	40
2.6	Performance Evaluations.....	43
2.6.1	Transverse ray aberration.....	43
2.6.2	Spot diagram.....	44
2.6.3	Encircled/ensquared energy.....	47
2.6.4	Modulation transfer function.....	48
2.6.5	Root-mean-square wavefront error/spot radius.....	48
2.6.6	Strehl ratio.....	49
2.7	Correction of Aberrations.....	51
2.7.1	Lens bending.....	51
2.7.2	Lens splitting.....	53
2.7.3	Lens compounding.....	55
2.7.4	Aplanatic surface.....	56
2.7.5	Aspherical surface.....	57
2.7.6	Optical materials.....	58
2.7.7	Stop shift.....	58
2.7.8	Symmetry principle.....	59
2.7.9	Field flatteners.....	61
2.8	Optical Design.....	61
2.8.1	Determining the system specifications.....	61
2.8.2	Selecting a starting design.....	61
2.8.3	Optimization.....	62
2.8.3.1	Establishing variables, constraints, and merit functions.....	62
2.8.3.2	Optimization and evaluation.....	64
	References.....	65
Chapter 3 Optical Fibers in Biomedical Imaging.....		67
3.1	Introduction.....	67
3.2	Fundamentals of Optical Fibers.....	68
3.3	Light Propagation in the Fiber.....	72

3.3.1	Single-mode fibers	72
3.3.2	Multimode fiber.....	72
3.3.3	Fiber tips.....	77
3.3.4	Fiber bundles	79
3.3.5	Fiber transmission	82
3.3.6	Autofluorescence of fiber bundles.....	84
3.4	Illumination Using Fiber Bundles	84
3.5	Fiber-Optic Imaging Systems.....	92
3.5.1	Introduction	92
3.5.2	Optical properties of imaging fiber bundles	92
3.5.3	Depth of focus/field.....	95
3.5.4	Optical properties of imaging systems using fiber bundles	98
	References	100

Chapter 4 Microscope Optics **103**

4.1	Introduction	103
4.1.1	Key specifications	105
4.1.2	Key parameters.....	108
4.1.2.1	Numerical aperture	108
4.1.2.2	Magnification	109
4.1.2.3	Resolution.....	110
4.1.2.4	Depth of field.....	111
4.1.3	Key components	113
4.1.3.1	Objective lens	113
4.1.3.2	Tube lens	114
4.1.3.3	Eyepiece	115
4.1.3.4	Light source	116
4.1.3.5	Illumination system	117
4.1.3.6	Imaging detector.....	118
4.2	Objective Lenses	118
4.2.1	Introduction	118
4.2.2	Aberration characteristics of objective lenses	120
4.2.2.1	Spherical aberration.....	120
4.2.2.2	Coma	121
4.2.2.3	Field curvature and astigmatism.....	121
4.2.2.4	Distortion.....	121
4.2.2.5	Chromatic aberrations	122
4.2.3	Design techniques.....	122
4.2.3.1	Aplanatic surfaces	122
4.2.3.2	Liquid immersion	125
4.2.3.3	Correction of field curvature	125
4.2.4	Objective lens design.....	128
4.2.4.1	Low-power achromats	129

4.2.4.2	Medium-power achromats	129
4.2.4.3	High-power achromats	130
4.2.4.4	Oil-immersion objective lenses	132
4.2.4.5	Plan achromats	133
4.2.4.6	Plan apochromats	135
4.2.4.7	Long-working-distance objective lenses	135
4.2.4.8	Objectives with correction lenses	139
4.2.4.9	Reflective objective lenses	140
4.3	Tube Lens	145
4.4	Eyepiece	147
4.5	Illumination System	152
4.5.1	Bright-field illumination	153
4.5.1.1	Critical illumination	153
4.5.1.2	Köhler illumination	156
4.5.2	Dark-field illumination	160
References	163
 Chapter 5 Fluorescence Imaging		165
5.1	Introduction to Fluorescence	165
5.1.1	Fluorescence process	165
5.1.2	Excitation and emission spectra	167
5.1.3	Intrinsic and extrinsic fluorescence	169
5.2	Fluorescence Imaging Techniques	171
5.2.1	Single-band fluorescence imaging	171
5.2.2	Multicolor fluorescence imaging	172
5.2.3	Hyperspectral fluorescence imaging	173
5.2.4	Fluorescence lifetime imaging	173
5.3	Components in Fluorescence Imaging Systems	173
5.3.1	Light source	174
5.3.2	Fluorescence filters	177
5.3.3	Illumination optics	177
5.3.4	Light collection/imaging optics	177
5.3.5	Detector	178
5.4	Fluorescence Filters	180
5.4.1	Types of fluorescence filters	180
5.4.2	Absorption filters	183
5.4.3	Thin-film filters	184
5.4.4	Specifications of thin-film filters	188
5.4.5	Filters in fluorescence imaging	190
5.4.5.1	Requirements for filters in fluorescence imaging	190
5.4.5.2	Filter selection	191
5.5	Fluorescence Imaging Systems	196
5.5.1	Lensless fluorescence imaging systems	196

5.5.2	Microimaging systems.....	199
5.5.3	Macrofluorescence imaging systems.....	203
5.5.4	Scanning fluorescence imaging systems.....	205
5.5.5	Fiber-optic fluorescence imaging systems.....	207
5.5.6	Multicolor fluorescence imaging systems.....	213
5.5.7	Hybrid imaging systems.....	215
5.6	Optical Designs of Fluorescence Imaging Systems.....	217
5.6.1	System consideration.....	217
5.6.2	Illumination system design.....	218
5.6.3	Detection systems.....	219
5.6.3.1	General requirements.....	219
5.6.3.2	Optical Materials.....	220
5.6.3.3	Design of a fluorescence microscope objective lens.....	225
	References.....	228

Chapter 6 Polarization Imaging..... 233

6.1	Basics of Polarized Light.....	233
6.2	Polarized-Light Interactions with Tissues.....	243
6.3	Polarization Imaging Systems in Biomedical Imaging.....	246
6.3.1	Conventional polarization imaging.....	246
6.3.2	Orthogonal polarized spectral imaging.....	247
6.3.3	Polarization ratio imaging.....	248
6.3.4	Spectral polarization difference imaging.....	249
6.3.5	Polarized light spectroscopy.....	252
6.4	Polarization Elements.....	253
6.4.1	Polarizers.....	255
6.4.1.1	Absorptive polarizers.....	255
6.4.1.2	Beamsplitting polarizers.....	258
6.4.1.3	Birefringent polarizers.....	261
6.4.1.4	Wire-grid polarizers.....	261
6.4.1.5	Circular polarizers.....	263
6.4.1.6	Special polarizers.....	264
6.4.2	Retarders.....	265
6.4.2.1	Birefringent retarders.....	266
6.4.2.2	Achromatic retarders.....	269
6.4.2.3	Variable retarders.....	270
6.4.3	Selection of polarization elements.....	270
6.4.3.1	Polarizers.....	271
6.4.3.2	Retarders.....	272
6.5	Polarization Effects in Optical Systems.....	273
6.5.1	Polarization effects of refraction/reflection.....	274
6.5.2	Polarization effects of coating.....	281
6.5.3	Birefringence.....	288

6.6	Optical Design and System Analysis	292
6.6.1	General considerations	292
6.6.2	Polarization ray tracing.....	294
	References	304
Chapter 7 Confocal Imaging		311
7.1	Introduction	311
7.1.1	Principle of confocal imaging	311
7.1.2	Resolution.....	313
7.2	Basic Components and Requirements	316
7.2.1	Light source.....	317
7.2.2	Illumination system	319
7.2.3	Objective lens	320
7.2.4	Beamsplitter.....	320
7.2.5	Pinhole.....	321
7.2.6	Pinhole lens	323
7.2.7	Scanner	324
7.2.7.1	Galvanometer scanner	324
7.2.7.2	Polygonal scanner.....	325
7.2.7.3	Acousto-optical deflector	325
7.2.7.4	MEMS scanner	326
7.2.7.5	Rotating wedge scanner.....	326
7.3	Confocal Scanning Systems	327
7.3.1	Stage scan	328
7.3.2	Beam scan.....	330
7.3.2.1	Introduction	330
7.3.2.2	Beam scan with a Nipkow disk	332
7.3.2.3	Beam scan with a spatial light modulator.....	337
7.3.2.4	Beam scan with scan mirrors.....	337
7.3.3	Relay systems	342
7.4	Optical Design of the Objective Lens	349
7.4.1	Aberrations in the objective lens	349
7.4.2	Design considerations of the objective lens	352
7.4.4	Design examples.....	355
7.4.4.1	Objective lenses for microarray scanners.....	355
7.4.4.2	Achromatic objective lens in the near-IR	356
7.4.4.3	Objective lens for UV and visible light.....	357
7.5	Fiber-Optic Confocal Imaging Systems	357
7.5.1	Single-fiber confocal imaging	359
7.5.2	Fiber-bundle confocal imaging.....	365
7.5.2.1	Introduction	365
7.5.2.2	Optical design.....	369
	References	373

Chapter 8 Endoscope Optics	379
8.1 Introduction	379
8.2 Basic Optics for Endoscopes	380
8.2.1 Illumination and imaging optics	381
8.2.2 Rigid endoscopes	383
8.2.3 Flexible endoscopes	385
8.2.4 Video endoscopes	386
8.3 Relay Lenses	387
8.3.1 Conventional relay systems	389
8.3.2 Hopkins rod-lens relay	391
8.3.3 Variations of the Hopkins rod lens	392
8.4 Objective Lenses	398
8.4.1 Introduction	398
8.4.2 Objective lenses with a fixed field of view	402
8.4.3 Objective lenses with variable fields of view	411
8.4.4 Design options	413
8.4.5 Prisms in endoscopic objective lenses	417
8.5 Illumination Systems	420
8.5.1 Light source	420
8.5.2 Fiber illumination systems	420
8.5.3 Illumination systems with fiber bundles and lenses	422
8.5.4 Illumination with LEDs	428
8.6 Wireless Endoscopes	430
8.6.1 Introduction	430
8.6.2 Imaging systems	433
8.6.3 Illumination systems	437
References	442
<i>Appendix: Lens Prescriptions</i>	445
<i>Index</i>	483

Preface

Biomedical optics is a rapidly growing area of research that has passed the “tipping point” for technology transfer, with significant momentum toward commercialization. Optical engineering is one of the keys in the development and commercialization of new imaging technologies, encompassing the selection of optical configurations, light sources, optical components, detectors, illumination and imaging system designs, and testing. Designing an efficient system for biomedical imaging requires a solid understanding of special requirements of optical systems for biomedical imaging and optical components used in the systems.

This book is developed from the SPIE course “Optical Design for Biomedical Imaging” that I have been teaching since 2008. It can be used as a reference for students, researchers, and engineers in biomedical fields. Readers can learn about the fundamentals of biomedical optics (light and tissue interaction, effect of tissue on optical systems), optical component specification and selection, light sources, detectors, various optical systems, and specific requirements for biomedical imaging. Readers will also learn how to design and model illumination and imaging systems for biomedical applications.

Chapter 1 gives an introduction to optical signals in biomedical optical imaging, tissues in optical systems, and biomedical optical imaging techniques. Readers without a background in optical design can gain some fundamental knowledge on optical systems and design from Chapter 2, which covers paraxial optics, optical properties and radiometry of optical systems, aberration fundamentals, aberration corrections, performance evaluations of optical systems, and basic optical design. Chapter 3 is dedicated to optical fibers in biomedical imaging, including the optics and optical properties of fibers and fiber bundles, fiber optic illumination, and imaging systems. Chapters 4–8 discuss microscope optics, fluorescence imaging, polarization imaging, confocal imaging and endoscope optics.

Each chapter starts with principles of imaging techniques and then discusses configurations, light sources, detectors, key optical components, and imaging and illumination system design with some design examples. Some other imaging techniques, such as optical coherence tomography, are not covered here, although they may be included in future editions.

Acknowledgments

This book is dedicated to my parents who have always loved me unconditionally. I am grateful to my wife Yiying, my son Derek, and my daughter Teresa, for their great support during the writing of this book.

I thank all those who have taught me about optical design, particularly Xiaocai Wang at the Institute of Optics & Electronics of China, Prof. Jose Sasian at the College of Optical Sciences, University of Arizona, and Dr. David Kessler at Eastman Kodak Co. I thank Ligan Wang at Semrock, Inc. for many helpful discussions on thin-film filters. I also thank many of my colleagues, particularly Victor Wong, Junwon Lee, Josh Cobb, Michael Harrigan, and Joseph Bietry for useful discussions on a variety of optically related problems. I would like to especially thank Dr. James Milch at Carestream Health, Inc. for his support in writing this book.

Rongguang Liang
December 2010

List of Acronyms and Abbreviations

AOD	acousto-optical deflector
AOTF	acousto-optical tunable filter
APD	avalanche photodiode
AR	antireflection
BFL	back focal length
CCD	charge-coupled device
CLC	cholesteric liquid crystal
CMOS	complimentary metal-oxide semiconductor
COC	cyclo olefin copolymer
COP	cyclic olefin polymer
CPC	compound parabolic concentrator
CW	continuous wave
Cy5.5	indocyanine dye
DAPI	4'-6-diamidino-2-phenylindole
DFOPS	dark-field orthogonal polarized spectral (imaging)
DMD	digital micromirror device
DNA	deoxyribonucleic acid
DOE	diffractive optical element
DOF	depth of focus
EEM	excitation-emission matrix
EFL	effective focal length
EMCCD	electron-multiplying charge-coupled device
FAD	flavin adenine dinucleotide

FITC	fluorescein isothiocyanate
FMN	flavin mononucleotide
FFL	front focal length
FOV	field of view
FWHM	full width at half maximum
GFP	green fluorescent protein
GRIN	gradient index
HBO (lamps)	mercury high-pressure arc lamps
HXP (lamps)	metal halide lamps
IBS	ion-beam sputtering
ICG	indocyanine green dye
ICCD	intensified CCD
IR	infrared
LCD	liquid crystal device
LCTF	liquid crystal tunable filter
LSA	longitudinal spherical aberration
MEMS	microelectromechanical system
MPE	multiphoton excitation
MTF	modulation transfer function
MRI	magnetic resonance imaging
NA	numerical aperture
NDDH	nicotinamide adenine dinucleotide
OCT	optical coherence tomography
OPS	orthogonal polarized spectral (imaging)
PC	polycarbonate
PCF	photonic crystal fiber
PE	phycoerythrin
PEM	piezoelastic modulator
PGC	protected-graft copolymer
PMMA	polymethyl methacrylate

PMT	photomultiplier tube
PSF	point spread function
PVA	polyvinyl alcohol
QE	quantum efficiency
RFP	red fluorescence protein
RMS	root mean square
ROI	region of interest
SHG	second harmonic generation (imaging)
SNR	signal-to-noise ratio
SPDI	spectral polarization difference imaging
TAC	triacetyl cellulose
TIR	total internal reflection
TSRLM	tandem-scanning reflected-light microscope
UV	ultraviolet
XBO (lamps)	xenon arc lamps

Chapter 1

Introduction to Biomedical Optical Imaging

1.1 Introduction

Biomedical optical imaging is one of the most rapidly growing and most relied-upon tools in health care for diagnosis and treatment of human diseases. Optical imaging with human eye visual examination was the first method ever used in disease detection and diagnosis.

Optical imaging overcomes some of the deficiencies found in the currently used medical techniques of x-ray, ultrasound, and magnetic resonance imaging (MRI), including:

- Harmful ionizing radiation, in the case of x-ray imaging,
- Lack of specific chemical and dynamic information from MRI, and
- Low resolution.

Modern biomedical optical imaging utilizes optical contrasts, such as the difference in light absorption, transmission, reflection, scattering, and fluorescence between the region of interest (ROI) and the surrounding area, providing both structural and functional information.

The following are some of the benefits offered by optical contrasts:¹

- Ionizing radiation does not exist.
- Optical spectra, based on absorption, fluorescence, or Raman scattering, provide information on cellular processes and tissue chemistry as well as detailed chemical and dynamic information.
- Optical absorption provides contrast for functional imaging.
- Optical scattering spectra provide information about the size distribution of optical scatter.
- Optical polarization provides information about structurally anisotropic tissue components.
- Optical properties of targeted contrast agents provide contrast for the molecular imaging of biomarkers.

- Optical imaging permits simultaneous detection of multiple contrast agents.

Using the spatial and temporal variations in the optical properties of tissues and cells, various types of biomedical optical imaging techniques have been developed with applications from fundamental research to clinical diagnosis.

The key advantages of these imaging techniques are:

- High resolution,
- Noninvasive procedures,
- Real-time capability,
- A wide range of imaging fields, from less than 100 nm to the entire human body,
- The ability to provide *in vitro*, *in vivo*, and *ex vivo* imaging,
- The ability to combine with other imaging techniques,
- Endoscopic compatibility to reach many organs and tissue sites for imaging,
- The ability to obtain tissue-section images and construct 3D images, and
- Point monitoring and imaging implementation.

1.2 Optical Signals in Biomedical Optical Imaging

1.2.1 Optical properties of tissues

Tissues are inhomogeneous materials. They behave like any bulk medium in which light propagation produces absorption, scattering, refraction, or reflection. Optical properties of tissues are characterized by the wavelength-dependent refractive index n , absorption coefficient μ_a , and scattering coefficient μ_s .

1.2.1.1 Refractive index

The refractive index n is a fundamental property of homogeneous media. For a tissue that is heterogeneous, the effective refractive index is generally approximated as the averaged value of the refractive indices of its constituents. The refractive index of the tissue determines the speed of light in the tissue; any change in the refractive index will introduce variations in refraction, reflection, and scattering.

Water makes up a large portion of most tissues; therefore, its refractive index ($n = 1.333$ at 587.6 nm) can be used as the minimum value for fluids and soft-tissue constituents. All tissue components have a refractive index larger than water.¹⁻³ For example, extracellular fluids have a refractive index of 1.35–1.36; for cytoplasm it is 1.36–1.375, for nuclei 1.38–1.41, for mitochondria and organelles 1.38–1.41, and for melanin 1.6–1.7. The volume-averaged refractive index of most biological tissues falls within 1.34–1.62, which is greater than the refractive index of water. For hard tissues, such as enamel, the refractive index is approximately 1.62.

Refractive index is a very important parameter in designing optical systems for biomedical imaging. Any deviation from the design value introduces some aberrations, particularly spherical aberration.

1.2.1.2 Absorption and absorption coefficients

Light absorption occurs as a result of the excitation of atoms or molecules into higher energetic states by photons. Absorption in tissues is dominated by protein, deoxyribonucleic acid (DNA), and other molecules in the ultraviolet (UV) region; by water in the infrared (IR) region; and by hemoglobin and melanin in the visible region. In the red to near-IR spectrum (from 600–1300 nm), absorption is minimal. This region is called the diagnostic and therapeutic window. At the short-wavelength end, the wavelength range (“window”) is bound by the absorption of hemoglobin. At the IR end of the window, penetration is limited by the absorption properties of water.²

The absorption coefficient μ_a is defined as the probability of photon absorption in a medium per unit path length.¹ For most biological tissues, μ_a is between 0.1 and 1 cm^{-1} . The reciprocal of μ_a is mean free path l_a :

$$l_a = \frac{1}{\mu_a}. \quad (1.1)$$

The intensity of light after traveling distance z in the tissue is given by Beer’s law:

$$I(z) = I_0 e^{-\mu_a z}, \quad (1.2)$$

where I_0 is the light intensity at $z = 0$.

The absorption coefficient of a tissue varies for different wavelengths. Each molecule or atom has its own characteristic absorption spectrum.

1.2.1.3 Scattering and scattering coefficients

Light scattering occurs in tissues that have a spatial variation in the refractive index n . Photons are scattered most strongly by a structure whose size matches the wavelength of the illumination light and whose refractive index differs from that of the surrounding medium.

When inelastic Raman scattering is not considered, light scattering is a process involving a photon elastically bouncing off of a microscopic obstacle. The photon is not destroyed in the process; only its direction of propagation is changed. The scattering properties of tissue can be characterized by Rayleigh or Mie scattering. Because scattering events redirect the photon paths randomly, the process causes image blur and reduces contrast in an imaging system.

The scattering coefficient μ_s is defined as the probability of photon scattering in a medium per unit path length. For most tissues, μ_s is between 50 cm^{-1}

(prostate) and 1000 cm^{-1} (tooth enamel). The scattering mean free path l_s is defined as the reciprocal of μ_s :

$$l_s = \frac{1}{\mu_s}. \quad (1.3)$$

The intensity of unscattered light after traveling distance z in the tissue is approximated by Beer's law:

$$I(z) = I_0 e^{-\mu_s z}, \quad (1.4)$$

where I_0 is the light intensity when it enters the tissue. The scattering coefficient is also wavelength dependent but usually much less so than the absorption coefficient.

1.2.1.4 Anisotropy

A scattering event causes the path of the photon to be changed by an angle θ , as shown in Fig. 1.1. The scattering function $p(\theta)$ describes the probability of a photon scattering into a unit solid angle oriented at an angle θ relative to the photon's original trajectory, and it has a unit of sr^{-1} . When $p(\theta)$ is constant for all θ , scattering is isotropic; otherwise, scattering is anisotropic.

As shown in Fig. 1.1, the component of the new path that is aligned in the forward direction is $\cos(\theta)$. The anisotropy g is defined as the expectation value of $\cos(\theta)$:

$$g = \langle \cos(\theta) \rangle = \int_0^\pi p(\theta) \cos(\theta) 2\pi \sin(\theta) d\theta. \quad (1.5)$$

The anisotropy is dimensionless—it is a measure of the amount of forward direction retained after a single scattering event. The value of g varies in the range from 0 to 1: $g = 0$ corresponds to isotropic scattering, and $g = 1$ to total forward scattering. For most biological tissues, g is ≈ 0.9 .

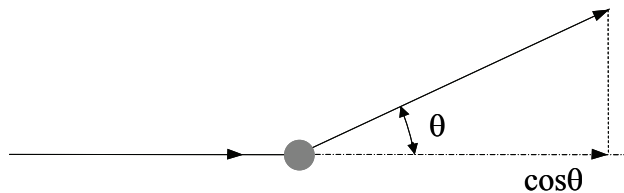


Figure 1.1 A scattering event causes a deflection at angle θ from the original forward trajectory.

1.2.1.5 Reduced scattering coefficient and penetration depth

Because not all scattering events change the path by a large angle, the reduced scattering coefficient μ'_s is introduced to represent the probability per unit distance that a photon will lose track of its original direction:

$$\mu'_s = \mu_s(1 - g). \quad (1.6)$$

On the basis of diffusion theory, the penetration depth δ of light in tissue is determined by the absorption and reduced scattering coefficients:⁴

$$\delta = \frac{1}{\sqrt{3\mu_a(\mu_a + \mu'_s)}} = \frac{1}{\sqrt{3\mu_a[\mu_a + (1 - g)\mu_s]}}. \quad (1.7)$$

In the diagnostic and therapeutic window (600–1300 nm), the penetration depth is longer for the longer wavelength because the scattering power decreases as wavelength increases.

1.2.2 Optical signals

1.2.2.1 Reflection, absorption, and scattering

The interaction between light and tissue can be classified as a reflection, scattering, and absorption process, as shown in Fig. 1.2. The optical properties of tissue, which include refractive index, scattering, and absorption coefficients, govern the interaction of light with tissue. When light is directed at tissue, a large percentage reflects or scatters off the tissue surface and maintains its polarization property. At normal incidence, approximately 4–7% of the incident light at all wavelengths reflects at the tissue surface because of the large change in the refractive index between the air and the tissue. This component accounts for the “glare” from the tissue surface and is the major visual cue by which surface texture is judged. Reflection without scattering at the tissue surface is termed specular reflection and does not change the wavelength and polarization state of the incident light. Most of the light, with one or two scattering events at the tissue surface or close to the tissue surface, also preserves its wavelength and polarization state.

The remainder of the light will penetrate the tissue and go through multiple scattering events. The light is scattered diffusely throughout the tissue before it is absorbed by the tissue components or emerges from the tissue. In special cases when the tissue is thin, a small percentage of the light may pass through the tissue without any scattering. Scattering causes a change in the direction of light propagation, and is the physical process by which the light can be detected through the tissue surface. The backscattered light carries few or no visual cues related to the surface texture, yet it carries all of the visual cues related to the internal tissue structure. After many scattering events, the scattered light loses its

polarization signature. Therefore, the backscattered light from inside the tissue is generally unpolarized.

Light absorption is fundamental to diagnostic and therapeutic applications. Without absorption, there is no energy transfer to the tissue, and the tissue is left unaffected. A transition between two energy levels of a molecule that is well defined at specific wavelengths can serve as a spectral fingerprint of the molecule for diagnostic purposes. Absorption serves as a mechanism of optical contrast for imaging. Absorption of energy is the primary mechanism that allows light to produce physical effects on tissue for treatment purposes.²

Within the diagnostic and therapeutic window, scattering is dominant over absorption. There are two types of photon scattering in tissue: elastic and inelastic. If the photon is scattered elastically, it changes its direction of propagation. Elastic scattering can reveal the size distribution of the scatterers. Inelastic scattering is a fundamental scattering process in which some energy of the photon is lost or gained; therefore, incident and scattered photons have different frequencies. In Raman scattering, which is inelastic, the molecule is excited into a virtual state. Raman scattering can reveal the specific chemical

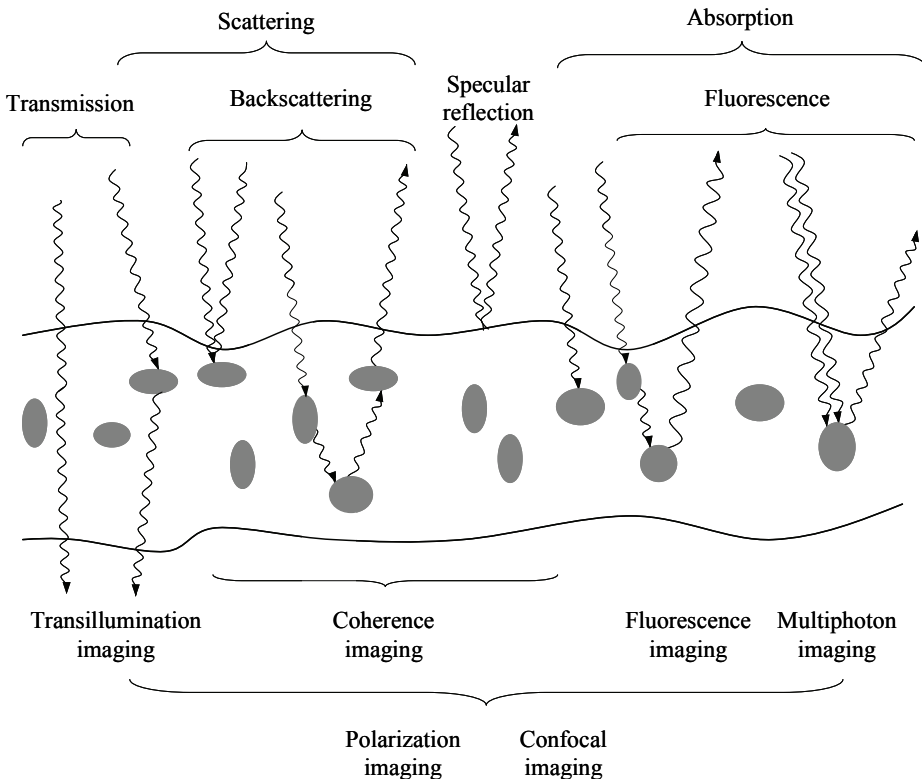


Figure 1.2 Interaction between light and tissue, and the biomedical optical imaging systems utilizing optical contrast mechanisms.

composition and molecular structure of biological tissue. It provides valuable information on the chemical and structural changes occurring as a result of disease.

Light scattering depends on the size, morphology, and structure of the components in the tissues. It also depends on the wavelength and polarization of the light. Any variation in these components as a result of disease will affect the properties of the scattered light, such as the scattering spectrum, polarization status, and coherence of the scattering light, thus providing a means for diagnosis. Scattering signals can be used to determine optimal light dosimetry and provide informative feedback during therapy.

Scattering signals are used for both spectroscopic and imaging applications. The spreading of a beam results in a loss of its initial directionality, reducing the signal in a spectroscopic system and the resolution in an imaging system.

1.2.2.2 Polarization effects

Depolarization of light inside the tissue originates from the tissue birefringence and scattering. Many biological tissues are optically anisotropic, showing a wide variety of polarization properties: linear birefringence, optical activity, and diattenuation. These properties result primarily from the linear anisotropy of fibrous structures, which is called birefringence of form, or the intrinsic anisotropic character of the tissue components or metabolic molecules, such as collagen, muscle fiber, keratin, and glucose. A large variety of tissues, such as the cornea and retina of the eye, sclera, tendon, cartilage, testis, muscle, nerve, bone, tooth, myelin, etc., exhibit birefringence of form.^{5,6}

Depolarization of light inside tissues is fairly complicated because both the birefringent and scattering effects can change the polarization state of light. Polarization gating can provide a novel contrast mechanism for tissue imaging and spectroscopy. By measuring the polarization degree of the scattered light, information about the structure of a tissue and birefringence of its components can be obtained.

Depolarization taking place during light propagation through the tissue depends not only on the type of tissue but also on the light wavelength. Therefore, the polarization properties of light can be employed as a selector of the photons coming from different depths in the tissue.

1.2.2.3 Fluorescence

When the tissue is illuminated, one photon is absorbed and excites an electron of a molecule or atom to a higher energy level. The energy is stored in the form of an excited state of the electron, which is not stable and will give back the energy in the form of thermal, kinetic, or photonic energy or a combination of these energies. Typically, the emitted photons carry less energy and thus have longer wavelengths than the photons originally absorbed. This process is called fluorescence. The molecule or atom involved in this process is called a

fluorophore. The average time between excitation and emission is called the fluorescence lifetime, which has a value on the order of nanoseconds.

The fluorophore can be endogenous or exogenous. When a fluorophore is a substance naturally contained in a tissue, its emission is called autofluorescence to distinguish it from the fluorescence generated by an exogenous fluorophore. Autofluorescence is produced by fluorophores in the tissue, including fluorophores from tissue matrix molecules and intracellular molecules like collagen, elastin, keratin, and nicotinamide adenine dinucleotide (NADH).

Fluorescence spectra provide detailed information on fluorescence molecules, such as their conformation, binding sites, and interaction within cells and tissues.⁴ Autofluorescence spectra provide a method of optically probing for the detection of cancer, metabolic changes, and so on. The fluorescence emission spectrum is specific for all fluorophores and is often used in fluorescence diagnostics.

Instead of one-photon absorption as discussed above, two or more photons of low energy can be absorbed simultaneously to produce an excited electron of high energy, which is then given back as one photon having a higher energy and lower associated wavelength than the exciting photons. This process is called multiphoton excitation (MPE), which is a nonlinear optical effect requiring very high light-energy densities that happens only in the focus of the imaging lens when a high-powered laser with ultrashort pulses is used. MPE has the advantage that fluorescence light contains information only on the focal region, eliminating the influence of multiple scattering and loss in contrast by (auto)fluorescence from surrounding regions.

1.2.2.4 Diseases and the optical properties of tissues

One of the reasons that biomedical imaging can be used to detect disease is that diseases change the optical properties of the tissue, namely, the refractive index, light scattering, and absorption.⁷ Diseases also change blood concentration, nuclear size distribution, epithelial thickness, and collagen content.

In healthy tissue, the amount of light reflected, backscattered, or transmitted from the tissue, the light's polarization properties, and the autofluorescence are different from those in diseased tissues. This difference provides a contrast mechanism for disease detection and diagnosis using optical imaging techniques. For example, changes in fluorescence reflect a complex mixture of alterations to intrinsic tissue fluorophore distribution and tissue structure. Fluorescence imaging can be used as a noninvasive tool for the detection of alterations in the structural and chemical compositions of cells; these alterations may indicate the presence of diseased tissue.

1.2.3 Tissues in optical systems

When imaging through tissue, it must be understood that tissues are not flat; they are pliable and can move voluntarily and involuntarily.⁸ When force is applied, the tissues will deform. This fact must be taken into consideration when

designing a high-performance imaging system because the movement of tissue will change the working distance and curvature of the tissue surface.

Because of its inhomogeneous properties, a tissue in the imaging path scatters the light like a diffuser, reducing image resolution. Tissues also distort the wavefront like a phase element because of its corrugated surface and the heterogeneity of the refractive index. The spatial frequency of corrugations determines the angular effects on the wavefront passing through the tissue. High-frequency corrugation causes high-angle scattering loss, but it preserves the mean wavefront. On the other hand, low-frequency corrugation introduces low-angle scattering and aberrates the mean wavefront. The effect of surface corrugation on an imaging system can be minimized by using an immersion liquid with a refractive index close to that of the tissue. With an immersion liquid filling the valleys of the tissue, the scattering at the tissue/liquid interface is reduced significantly, and the corrugated tissue/air interface is replaced by a flat liquid/air interface with minimal wavefront deformation.

With the exception of the wavefront deformations discussed above, when taking images by focusing the light into the tissue, the tissue will introduce spherical aberration due to the mismatch or the variation of refractive indices. When designing an imaging system, this effect must be considered by either treating the tissue as a flat plate with an appropriate refractive index and thickness or by using an index-matching liquid.

When focusing the light inside the tissues for imaging, the peak irradiance will be reduced, and the size of the focused spot will be increased by the tissue scattering caused by the heterogeneity of the refractive indices. The peak irradiance is reduced rapidly inside the tissue, and the speed largely depends on the scattering coefficient μ_s and the absorption coefficient μ_a . The scattering coefficient μ_s and anisotropy g govern the increase of the spot size. Figure 1.3 shows the simulated irradiance profiles of a focused spot as it moves from the tissue surface into the tissue. In this simulation, an index-matching immersion liquid is used between the objective lens and the tissue. Figure 1.3(b) is the normalized irradiance when the light is focused on the tissue surface. The spot size is very small and is determined by the focused lens. When the light is focused inside the tissue, the irradiance drops quickly and the full width at half maximum (FWHM) of the irradiance profile increases significantly, as shown in Figs. 1.3(c) and 1.3(d), due to scattering and absorption. The irradiance profiles show that an imaging system can achieve full resolution at the tissue surface, but the contrast and resolution are reduced rapidly as the focused spot moves into the tissue.

1.3 Biomedical Optical Imaging Techniques

A number of biomedical optical imaging and spectroscopic techniques, based on the optical signals discussed above, have been developed and used in fundamental research and clinical applications. This book will focus on only the imaging techniques.

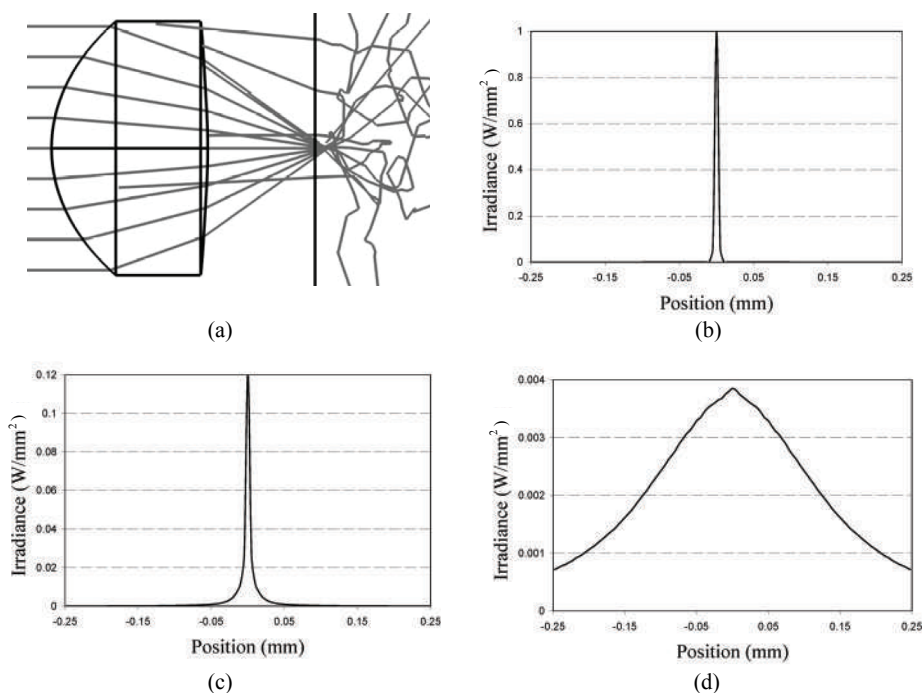


Figure 1.3 (a) Ray diagram of a focused spot inside a tissue, (b) normalized irradiance profile when the focused spot is at the tissue surface, (c) irradiance profile when the focused spot is inside the tissue but close to the tissue surface, and (d) irradiance profile when the focused spot is deep inside the tissue. The irradiances in (c) and (d) are normalized by the peak of the irradiance when the focused spot is at the tissue surface.

Figure 1.2 shows some selected optical imaging techniques and the contrast signals they utilize. The imaging techniques can be roughly classified, based on contrast mechanisms and imaging methods, as transillumination, reflectance, fluorescence, spectral, spatial filtering, coherence gating, polarization gating, time gating, and frequency-domain imaging, as well as nonlinear optical imaging systems. Some imaging systems can be placed in more than one category. For example, confocal imaging is treated as reflectance imaging as well as spatial filtering imaging. Some imaging systems also utilize more than one contrast mechanism; for example, fluorescence confocal imaging systems use a fluorescence signal and a spatially filtered backscattered photon.

Transillumination imaging utilizes a spatial variation of absorption and scattering in tissues. As the light propagates through a tissue, the transmitted light comprises three components: unscattered, weakly scattered, and multiple-scattered light. The characteristic absorption of light can be used to detect any change in absorption.

Reflectance imaging collects the backscattered light and is the most common imaging system in biomedical imaging. Both the backscattered and specularly

reflected photons are collected and contribute to the final images without filtering.

Fluorescence imaging is widely used in biomedical imaging. Part of the emitted fluorescence from the tissue is captured and contributes to the final images. The excitation light must be blocked by using optical filters to a sufficiently low level. Fluorescence imaging provides a comprehensive and detailed probing of the structure and dynamics of biological specimens of widely varying dimensions. Nonlinear optical methods, such as multiphoton imaging and label-free second harmonic generation (SHG) imaging, have also recently emerged as extremely useful biomedical imaging techniques.

Because the optical properties of the tissues are a function of wavelength, the transmitted and backscattered photons are wavelength dependent. The fluorescence signal is also wavelength dependent. Spectral imaging provides spectral information on the spatial variation of fluorescence or reflectance spectra. In fluorescence imaging, spectral imaging also allows for the simultaneous use of more than one fluorescence marker.

Spatial filtering relies on the fact that diffuse photons that undergo multiple scattering are more spread out from the focal point. It uses a small aperture to enhance image contrast by rejecting the off-axis diffuse photons. Spatial filtering is the fundamental basis of confocal imaging. The confocal aperture provides depth discrimination in reflection and fluorescence imaging.

Coherence gating is another method to separate the diffuse photons from others with less scattering. It is a reflection-imaging technique similar to ultrasound imaging, except that a low-coherence light wave is used. Coherence gating's selectivity of the backscattered light from the ROI is achieved by using the interference between the backscattered light and the reference beam. The axial (depth) resolution is defined by the coherence length of the light source; the shorter the coherence length, the better the depth resolution. The lateral resolution is determined by the spot size of the focusing optics.

A polarization-gated imaging system uses polarized light in the illumination and uses an analyzer in the detection path to separate the transmitted or backscattered photons with different polarization states, therefore providing structural and depth information.

A time-gated imaging system utilizes a train of short laser pulses as the illumination source. The transmitted or backscattered photons pass through an optical gate that opens and closes to detect different types of photons. Synchronization can be achieved by using a reference optical pulse that controls the opening and closing of the optical gate.

In a frequency-domain imaging system, the tissue is illuminated with an intensity-modulated light from a continuous-wave (CW) laser, or with the harmonic content of a pulsed-laser system. The alternating current (ac) modulation amplitude and phase shift of the transmitted or backscattered signal are measured using methods such as heterodyning or homodyning.

References

1. L. V. Wang, *Biomedical Optics: Principles and Imaging*, Wiley-Interscience, Hoboken, NJ (2007).
2. J. Mobley and T. Vo-Dinh, "Optical properties of tissue," Chapter 2 in *Biomedical Photonics Handbook*, T. Vo-Dinh, Ed., CRC Press, Boca Raton, FL (2003).
3. F. B. Bolin, L. E. Preuss, R. C. Taylor, and R. J. Ference, "Refractive index of some mammalian tissues using a fiber optic cladding method," *Appl. Opt.* **28**, 2297–2303 (1989).
4. B. C. Wilson and S. L. Jacques, "Optical reflectance and transmittance of tissue: principles and applications," *IEEE J. Quantum Electron.* **26**, 2186–2199 (1990).
5. V. Tuchin, *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis*, 2nd ed., SPIE Press, Bellingham, WA (2007). [doi: 10.1117/3.684093].
6. V. V. Tuchin, L. V. Wang, and D. A. Zimnyakov, *Optical Polarization in Biomedical Applications*, Springer, New York (2006).
7. E. Salomatina, B. Jiang, J. Novak, and A. N. Yaroslavsky, "Optical properties of normal and cancerous human skin in the visible and near infrared spectral range," *J. Biomed. Opt.* **11**(6), 064026 (2006). [doi: 10.1117/1.2398928].
8. J. Zavislan, *Notes of Optical Engineering for Biomedical Optics*, Optics Summer School, University of Rochester (2005).

Chapter 2

Introduction to Optical Design

2.1 Introduction to Optical Design

In general, optical design includes a broad spectrum of topics, including geometrical optics, image formation, aberration theory, lens design, and image quality evaluation. Each topic is worth an entire book. This chapter will discuss only the necessary background on optical design. For more detailed discussions on optical design, please refer to Refs. 1–17.

2.1.1 Paraxial optics

This section will discuss the properties of the optical system in the region close to the optical axis called the paraxial region. In a practical system, most of the rays come from outside the paraxial region, but the basic properties of an optical system described in the paraxial region are applicable outside that region as well. The ideal behavior of an optical system in the paraxial region was first developed by Gauss, so it is referred to as “Gaussian optics” in many optical references.

Figure 2.1 is a schematic diagram of a simplified optical system showing the first and last surfaces. An optical system typically comprises one or more optical elements. The first-order properties of an optical system can be completely characterized by cardinal points, which are focal points (F, F'), principal points (P, P'), and nodal points (N, N'). When both the object and the image are in air, the nodal points coincide with the principal points.

The focal points of a lens system are the points at which an incident ray parallel to the optical axis of the optical system crosses the axis after passing through the system. The front focal point has the magnification of infinity, while the magnification of the rear focal point is zero. Nodal points have the unique property that a line passing through the front nodal point in the object space will emerge from the rear nodal point in the image space having the same angle with respect to the optical axis. The nodal points are thus characterized by unit angular magnification.

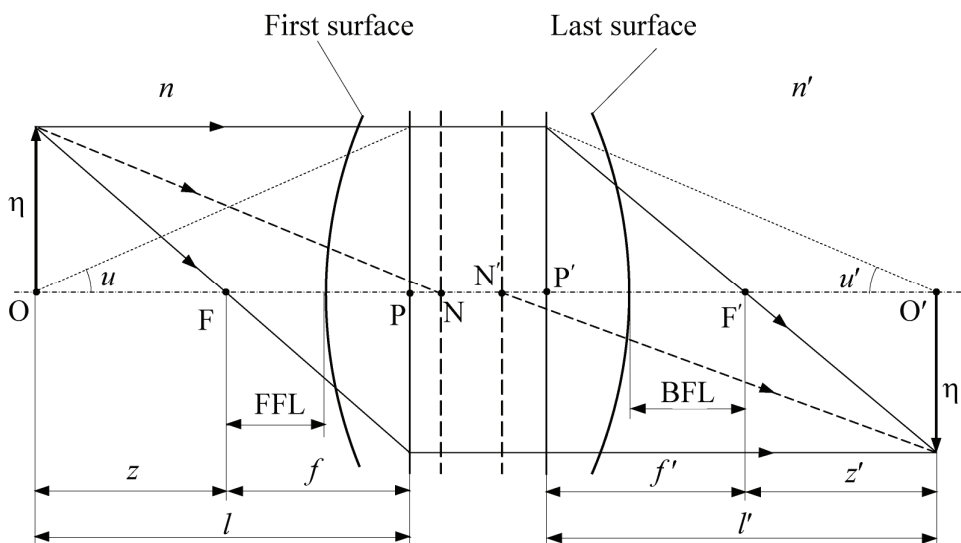


Figure 2.1 Cardinal points of an optical system.

The planes related to the cardinal points are the focal planes, principal planes, and nodal planes. The principal planes are two hypothetical planes in an optical system on which all of the refraction can be considered to take place. They are conjugates of unit magnification: one plane is the image of the other plane. For a given set of lenses and separations, the principal planes are fixed and do not depend on the object's position O . The principal points are the points where the principal planes cross the optical axis. Note that the principal planes, and therefore the principal points, can fall outside of the lens.

The principal planes are important in that a complex system can be treated as a thin lens provided the conjugate distances are measured from the principal planes. The front and rear focal lengths (f and f') are defined as the distances from the front and rear principal planes to the respective focal points. Similarly, the object and image distance (l and l') are the distances from the object and image planes to their respective principal planes. The back focal length (BFL) is the distance from the last surface of the system to the back focal point F' , and the front focal length (FFL) is the distance from the first surface to the front focal point F . The working distance is defined as the distance from the first lens surface to the object, which is $(FFL + z)$ in the figure.

The field of view (FOV) of an optical system is the diameter of the object or the maximum angular size of the object. Image height is also used to define the FOV. For a rectangular image format, the diagonal FOV is usually used.

Figure 2.1 shows three principal rays (solid lines) that are commonly used to define the first-order optical properties of an optical system. These are the ray from the object point through the nodal point, the ray from the object point parallel to the optical axis, and the ray from the object point through the focal point.

Throughout this book, the distances above the optical axis and the distances from the cardinal points to the right are positive, while the distances below the optical axis and the distances from the cardinal points to the left are negative.

2.1.2 Newtonian and Gaussian equations

Newtonian equations characterize the imaging relationship from the object plane to the image plane when the axial locations of the conjugate planes are measured relative to the respective focal points. When the refractive indices in the object and image spaces are assumed to be equal to 1, the Newtonian equations are

$$m = \frac{\eta'}{\eta}, \quad (2.1)$$

$$z = -\frac{f}{m}, \quad (2.2)$$

$$z' = -mf', \quad (2.3)$$

$$zz' = ff', \quad (2.4)$$

where m is the lateral magnification, η and η' are the object and image heights, respectively, and z and z' are the distances from the object and image planes to their respective focal points.

Gaussian equations describe the image mapping when the locations of the conjugate object and image planes are measured from their respective principal planes. The following Gaussian equations are frequently used:

$$m = \frac{l'}{l}, \quad (2.5)$$

$$l = \frac{-(1-m)}{m} f, \quad (2.6)$$

$$l' = (1-m)f', \quad (2.7)$$

$$\frac{f'}{l'} + \frac{f}{l} = 1. \quad (2.8)$$

The longitudinal magnification \bar{m} is the magnification along the optical axis; it relates the distances between pairs of conjugate planes:

$$\bar{m} = m^2. \quad (2.9)$$

This means that the longitudinal magnification is positive and that the object and image always move in the same direction.

2.1.3 Stop and pupil

As shown in Fig. 2.2, the ray bundle from the object point is defined by apertures in the optical system. An aperture stop is the aperture of one component in the optical system that limits the bundle of light from an on-axis object point that can propagate through the system. An aperture stop can be a lens aperture or a separate physical aperture in the optical path.

The image of the aperture stop in the object space is called the entrance pupil, which limits the cones of light entering the system. The image of the aperture stop in the image space is the exit pupil, which defines the cone of the light exiting the system.

The marginal and chief rays are two special tangential rays that define the properties of the object, pupil, and image. As shown in Fig. 2.2, the marginal ray starts at the on-axis object point, goes through the edge of the aperture stop, and intersects the optical axis. It defines the pupil size and image location. The marginal ray height and marginal ray angle are denoted by y and u , respectively. The chief ray starts at the edge of the object and goes through the center of the aperture stop. It defines the pupil location and image height. The chief ray height and chief ray angle are denoted by \bar{y} and \bar{u} , respectively.

For a complex imaging system, the marginal and chief rays can cross the optical axis more than once. At each intersecting point of the chief ray on the optical axis where the pupil or aperture stop is located, the size of the pupil or aperture is determined by the marginal ray height in that plane. The intermediate

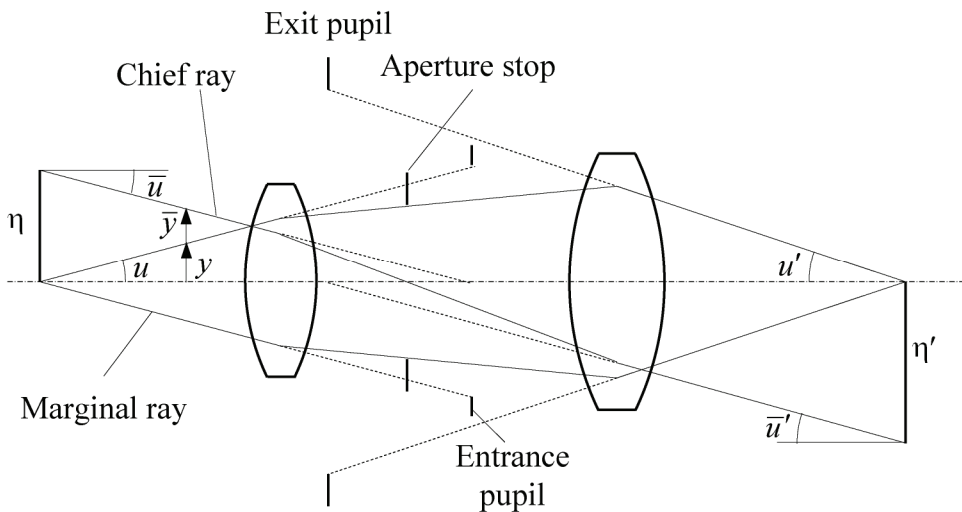


Figure 2.2 Aperture stop, pupils, and chief and marginal rays in an optical system.

image locations are the intersecting points of the marginal rays on the optical axis. The image height is determined by the chief ray height in the plane.

The aperture stop defines the ray bundle from the on-axis object point. As shown in Fig. 2.3, other elements may block the ray from the off-axis point. This is called vignetting. Vignetting causes a decrease in the brightness toward the rim and should be minimized. However, in some situations, vignetting may be desired to block some off-axis rays to improve image quality.

When two optical systems—for example, the microscope objective lens and the eyepiece—are cascaded, the exit pupil of the first must match the entrance pupil of the second in location and size. Otherwise, not all of the light from the first system can go through the second system, and the final image does not appear uniformly illuminated.

Field stops limit the FOV. A field stop must be in a plane conjugate with the object or image plane. In many situations, the sensor itself is the field stop. The images of a field stop in the object and image spaces are called the entrance and exit windows, respectively.

As shown in Fig. 2.4(a), when the aperture stop is at the rear focal point, the entrance pupil is at infinity. The system is telecentric in the object space. If the aperture stop is at the front focal point, as shown in Fig. 2.4(b), the exit pupil is at infinity, and the system is telecentric in the image space. If the aperture stop is placed at the common focal point of the front and rear lens groups, as shown in Fig. 2.4(c), both the entrance and exit pupils are at infinity, and the system is then called double telecentric. The advantage of telecentricity is that the magnification is constant for the defocused object in the object-space telecentric system, and the defocused image plane or detector will not change the image height in the image-space telecentric system.

Telecentricity is a required feature for many biomedical imaging systems, such as confocal imaging, microscopic imaging, and optical coherence tomography (OCT) imaging systems. The object at different depths will have the same magnification, leading to more-accurate 3D images.

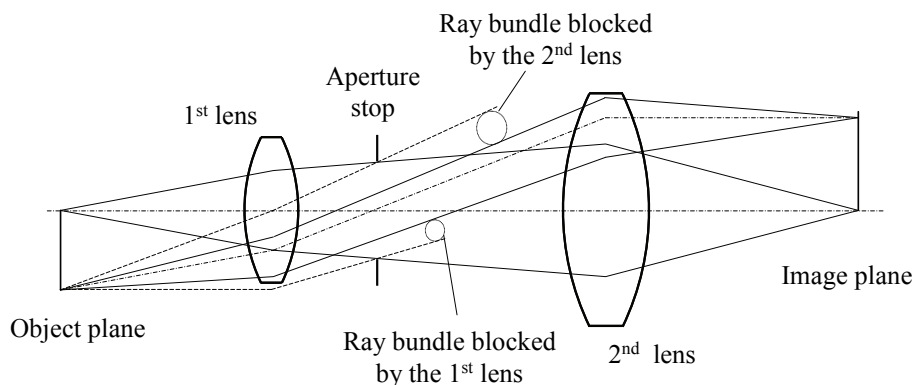


Figure 2.3 Schematic diagram of vignetting.

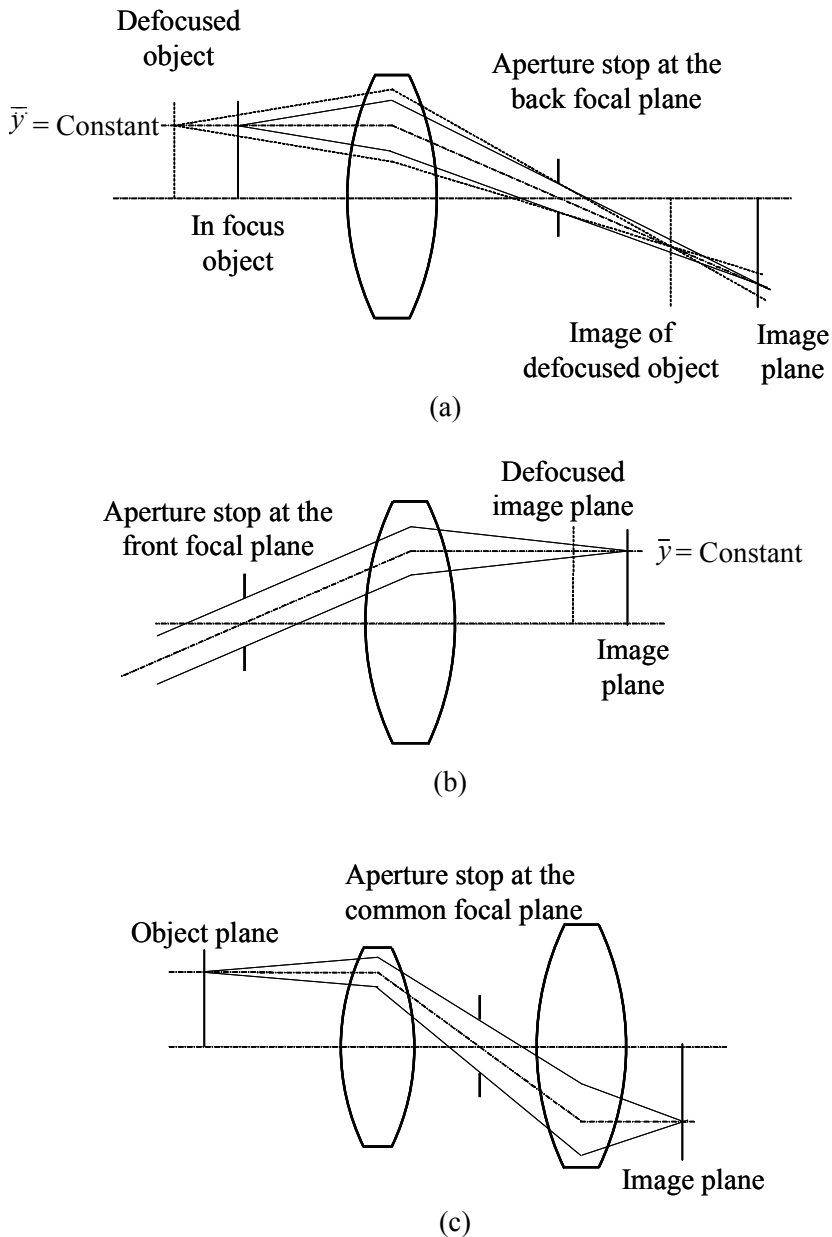


Figure 2.4 Telecentric systems: (a) object-space telecentric system, (b) image-space telecentric system, and (c) double telecentric system.

2.2 Optical Properties

2.2.1 Numerical aperture

The numerical aperture (NA) of an optical system is defined as the product of the refractive index n and the sine of its acceptance angle u :

$$\text{NA} = n \sin u, \quad (2.10)$$

and it is the most important quantity of an optical system; it determines the resolution, throughput, and depth of field/focus. The larger the NA is, the higher the resolution and the throughput.

The F -number ($f/\#$) is an alternative quantity to specify the aperture of the system. It is the ratio between the focal length f and the diameter of the exit pupil D_{ex} :

$$f/\# = \frac{f}{D_{\text{ex}}}. \quad (2.11)$$

For an infinite conjugate system, NA in the image space and $f/\#$ are closely related:

$$f/\# = \frac{1}{2\text{NA}}. \quad (2.12)$$

The working $f/\#$ describes the image-forming cone for a finite conjugate system and is defined as

$$f/\#_w = (1 - m)f/\#. \quad (2.13)$$

2.2.2 Depth of focus/field

For each imaging system, there is an acceptable image blur that is determined by the detector resolution or the overall system resolution requirement. Depth of focus (DOF) is a measurement of how much the sensor plane can be shifted from the nominal image position of an object before the resulting blur exceeds the acceptable blur criterion. It is calculated as a geometrical tolerance of the image plane without consideration of diffraction or aberration.

As shown in Fig. 2.5, the DOF is determined by the blur diameter criterion B and the system aperture D_{ex} :

$$\text{DOF} = \pm B \frac{L}{D_{\text{ex}}} = \pm B(f/\#) = \pm \frac{B}{2\text{NA}}. \quad (2.14)$$

For a diffraction-limited imaging system, the DOF corresponding to an optical path difference (OPD) of $\pm\lambda/4$ is

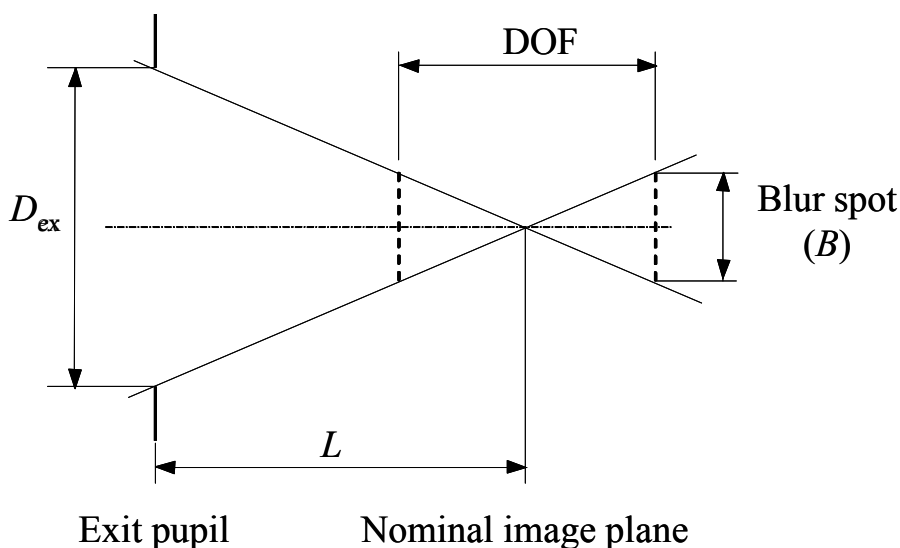


Figure 2.5 Depth of focus.

$$\text{DOF} = \pm \frac{\lambda}{2(\text{NA})^2} = \pm 2\lambda(f/\#)^2. \quad (2.15)$$

The depth of field is the amount by which the object may be shifted before the acceptable blur is produced. It can be calculated from the depth of focus in the image space.

2.2.3 Lagrange invariant

The Lagrange invariant H is a measure of the light propagating through an optical system. It is defined as

$$H = n\bar{u}y - nu\bar{y}, \quad (2.16)$$

where y and u are the marginal ray height and angle, respectively, and \bar{y} and \bar{u} are the chief ray height and angle in any arbitrary plane.

The Lagrange invariant is an invariant quantity throughout the entire system, not just at conjugate planes. At an image plane, the Lagrange invariant becomes

$$H = nu\bar{y}. \quad (2.17)$$

At a pupil plane, it is

$$H = n\bar{u}y. \quad (2.18)$$

The system throughput, or étendue, is proportional to the square of the Lagrange invariant.

2.3 Radiometry of Optical Systems

2.3.1 Basic concepts

Radiometry studies electromagnetic radiation and its transfer from one surface to another. It characterizes the propagation of radiant energy through an optical system and estimates how much source radiant power is transmitted through an optical system and is collected at the detector surface. For a more detailed discussion of this subject, refer to Refs. 15 and 16.

The basic unit of radiometry is the watt (W); all other units are defined in terms of watt, area, and solid angle. The solid angle Ω is defined as the area on a sphere of radius R subtended by the solid angle divided by R^2 :

$$d\Omega = \frac{dA}{R^2}. \quad (2.19)$$

The dimensionless unit of a solid angle is the steradian (sr), with 4π sr in a full sphere. As shown in Fig. 2.6, for an arbitrary surface area not on the surface of a sphere, the solid angle is the ratio of the projected area on a sphere and the square of the radius of that sphere. For a right circular cone of half-angle θ , the solid

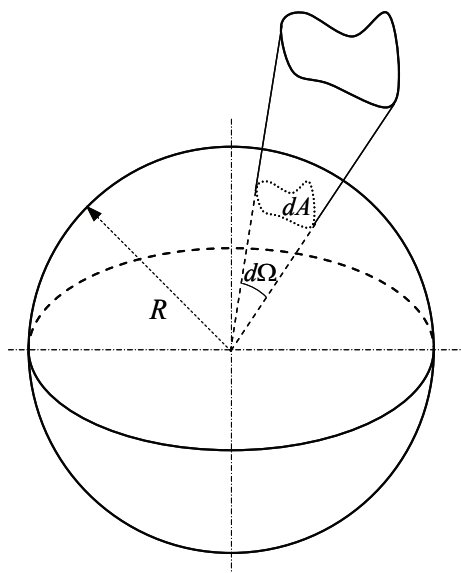


Figure 2.6 Definition of a solid angle.

angle is


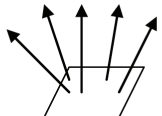
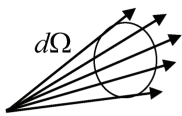
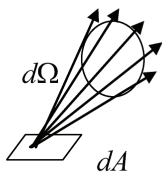
$$\Omega = 2\pi(1 - \cos\theta) = 4\pi \sin^2\left(\frac{\theta}{2}\right). \quad (2.20)$$

If the area A is at an angle ϕ from the observer's plane, the quantity $A \cos\phi$ is referred to as the projected area. The solid angle weighted by $\cos\phi$ is called the projected solid angle.

Table 2.1 contains radiometric terminology, definitions, and units. Exitance and irradiance are related by the reflectance ρ of the surface:

$$M = \rho E. \quad (2.21)$$

Table 2.1 Definitions of radiometric quantities.

Quantity	Definition	Units	
Power (Φ)	Radiant power or flux	W(J/s)	
Irradiance	Power incident per unit area upon a surface $E = \frac{d\Phi}{dA}$	W/m ²	
Radiant exitance	Power emitted per unit area from a surface $M = \frac{d\Phi}{dA}$	W/m ²	
Radiant intensity	Power radiated per unit solid angle $I = \frac{d\Phi}{d\Omega}$	W/sr	
Radiance	Power radiated per unit solid angle per unit projected area of the source, or power radiated per unit projected area per unit solid angle $L = \frac{d^2\Phi}{d\Omega dA \cos\theta}$	W/(sr·m ²)	

The exitance of a Lambertian source is related to its radiance L by π :

$$M = \pi L . \quad (2.22)$$

For an imaging system working on axis with a small NA, the flux through the system is

$$\Phi = LA\Omega . \quad (2.23)$$

$A\Omega$ is the system throughput, also known as the geometrical extent or étendue. It is proportional to the square of the Lagrange invariant:

$$A\Omega = \frac{\pi^2 H^2}{n^2} . \quad (2.24)$$

This quantity is also invariant.

One of the most important theorems in radiometry is the conservation of radiance. The conservation of radiance means that the radiance of an image formed by a lossless optical system is equal to the radiance of the object multiplied by the transmission of the system. More generally, the radiance divided by the square of the refractive index is the invariant quantity.¹

2.3.2 Radiometric approximations

Lambert's cosine law, the inverse-square law, and the cosine-fourth falloff law are three well-known laws in radiometry. Many extended sources of radiation obey Lambert's cosine law, which states that the radiant intensity emitted from a Lambertian source decreases with the cosine of the angle θ from the normal of the surface:

$$I = I_0 \cos \theta , \quad (2.25)$$

where I_0 is the intensity at normal. Although the emitted radiation per steradian falls off with $\cos\theta$ in accordance with Lambert's law, the projected area falls off at exactly the same rate. Therefore, the radiance of a Lambertian surface is constant with respect to θ .

The inverse-square law states that the irradiance from an isotropic point source varies inversely with the square of the distance from the source:

$$E = \frac{I \cos \theta}{d^2} , \quad (2.26)$$

where d is the distance from the light source to the detector, as shown in Fig. 2.7(a), I is the intensity of the light source, and θ is the tilt angle of the receiver relative to the light propagation direction.

When an extended Lambertian light source with an area A illuminates a flat surface with distance d , as shown in Fig. 2.7(b), the irradiance on the flat surface is

$$E = \frac{LA \cos^4 \theta}{d^2}. \quad (2.27)$$

The image irradiance varies across the FOV in proportion to the fourth power of the cosine of the field angle.

2.3.3 Radiometry of optical systems

As shown in Fig. 2.8, the power Φ from an area A on the object captured by the lens is

$$\Phi = LA\Omega = LA \frac{\pi D_L^2}{4l^2}, \quad (2.28)$$

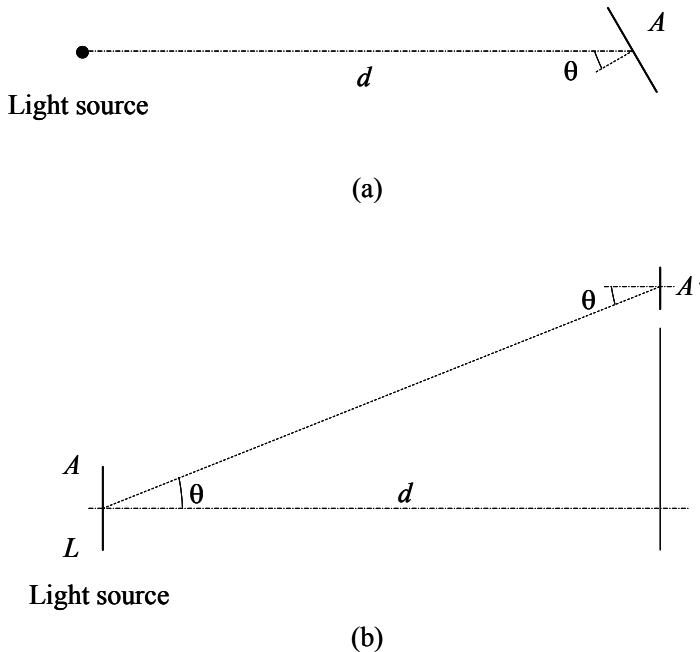


Figure 2.7 (a) Illustration of the inverse-square law. (b) Illustration of the cosine-fourth falloff law.

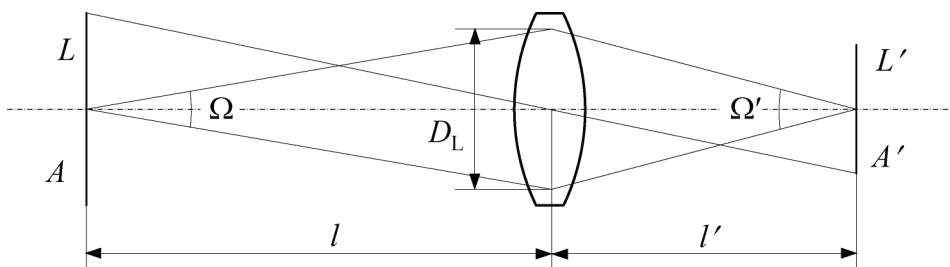


Figure 2.8 Radiation transfer in an optical system.

where D_L is the diameter of the lens and l is the working distance from the object to the lens. Thus, the flux collected by a lens is proportional to the square of the NA. All of the power is transferred to the image, assuming the system transmission is 1.

For the optical system in Fig. 2.8, we have

$$l = \frac{l'}{m}, \quad A = \frac{A'}{m^2}, \quad l = \frac{-(1-m)f}{m}. \quad (2.29)$$

Therefore,

$$\Phi = \frac{\pi L A' D_L^2}{4(1-m)^2 f^2} = \frac{\pi L A'}{4(1-m)^2 (f/\#)^2}. \quad (2.30)$$

The image plane irradiance is

$$E' = \frac{\Phi}{A'} = \frac{\pi L}{4(1-m)^2 (f/\#)^2} = \frac{\pi L}{4(f/\#_w)^2} = \pi L (\text{NA})^2. \quad (2.31)$$

This equation is known as the camera equation, relating the image irradiance to the scene radiance.

2.4 Introduction to Aberrations

When discussing the optical properties of an optical system in the previous section, it had been assumed that the rays propagate through the optical system according to the paraxial prediction. However, real rays do not refract according to the linear, first-order Snell's law but according to the nonlinear Snell's law.

The sine function in Snell's law can be expanded in an infinite Taylor series. When only the first term is used, it is called the first-order or paraxial approximation and is valid only when θ is small, which means that the rays are close to the optical axis. For real rays, θ is a function of the ray intersection height and surface curvature. Rays from the same object point but with different

angles no longer intersect at the same point in the image space. This phenomenon is called monochromatic aberration.

All optical materials have dispersion, which is a variation of the refractive index with wavelength. Thus, light with different wavelengths will be focused at different locations because the power of the lens depends on the refractive index. This phenomenon is called chromatic aberration.

Exact ray tracing is a rigorous way to analyze real lens surfaces and system aberrations. We can consider the aberrations either in terms of the geometrical ray intersection errors (ϵ_x, ϵ_y) at the image plane or as the optical path length difference $W(x_p, y_p)$. The first method to describe aberrations is called ray aberration, and the second is called wavefront aberration.

The ray aberration is defined as the distance between the reference image point and the corresponding real image point. The reference image point is defined by the intersection of the paraxial chief ray and the paraxial image plane. The ray aberration is usually measured as the longitudinal and transverse ray aberrations, as shown in Fig. 2.9. The longitudinal aberration ϵ_z is the axial

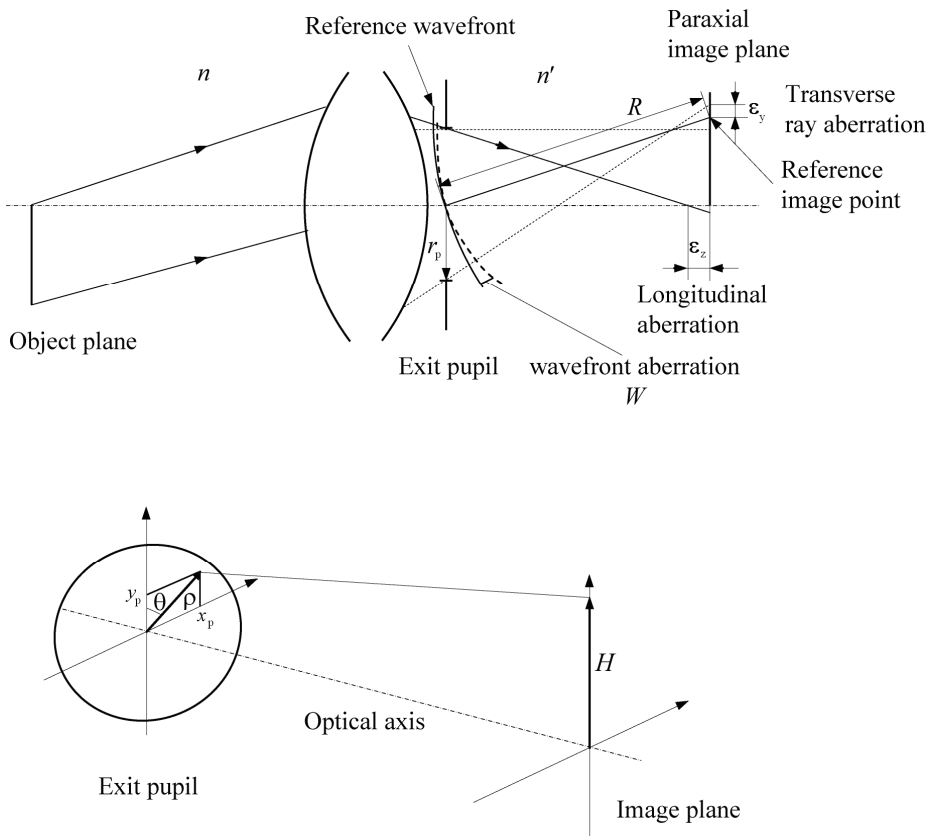
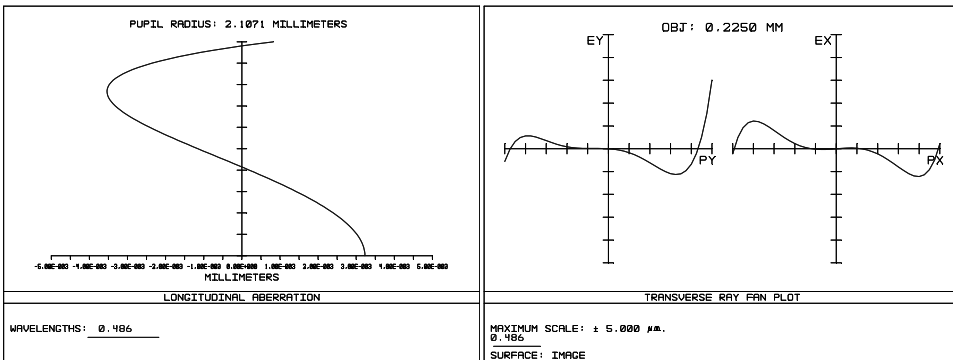


Figure 2.9 Transverse ray, longitudinal, and wavefront aberrations.

displacement from the nominal of the axial intersection point. It is measured as the distance to the paraxial image plane along the optical axis as a function of pupil height. Figure 2.10(a) is a typical plot of longitudinal aberration. One of the advantages of longitudinal aberration is its direct relation to the paraxial image location.

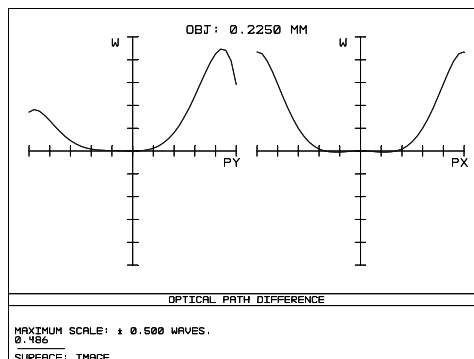
Transverse ray aberrations are lateral displacements of a ray intersection point from a reference point on the paraxial image plane. For a rotationally symmetric system, two special sets of rays are used for aberration analysis: tangential rays intersecting the pupil at $x_p = 0$ and sagittal rays at $y_p = 0$. Transverse ray aberrations are represented as the tangential ray fan ϵ_y as a function of y_p for $x_p = 0$, and the sagittal ray fan ϵ_x as a function x_p for $y_p = 0$. Figure 2.10(b) is a plot of transverse ray aberration.

The wavefront aberration $W(x_p, y_p)$ is defined as the difference in optical path length between the actual wavefront and the reference sphere or reference wavefront centered on the reference image point. Figure 2.10(c) is the plot of wavefront aberration corresponding to the transverse ray aberration in Fig. 2.10(b).



(a)

(b)



(c)

Figure 2.10 Aberration plots: (a) longitudinal aberration, (b) transverse ray aberration, and (c) wavefront aberration.

Longitudinal and transverse ray aberrations provide a complete and powerful method to describe the aberrations of an optical system. They are preferentially used in designing and characterizing optical systems. In recent years, wavefront aberrations have been used frequently in aberration analysis because of the following advantages:

- The effect of diffraction can be included in the calculation of wavefront aberration, and wavefront aberration is the essential input in calculating the diffraction image.
- Wavefront aberrations can be measured accurately by interferometric methods.
- The wavefront aberration of an optical system is simply the sum of the single-surface contributions.

Given that the rays are normal to the wavefronts, the transverse ray and wave aberrations are related in that the transverse ray aberration is the derivative of the wave aberration with respect to the exit pupil coordinates:

$$\epsilon_x = -\left(\frac{R}{r_p}\right)\left[\frac{\partial W(x_p, y_p)}{\partial x_p}\right], \quad (2.32a)$$

$$\epsilon_y = -\left(\frac{R}{r_p}\right)\left[\frac{\partial W(x_p, y_p)}{\partial y_p}\right], \quad (2.32b)$$

$$\frac{R}{r_p} = \frac{-1}{n'u'} = 2(f/\#), \quad (2.32c)$$

where r_p is the physical radius of the exit pupil, R is the radius of the reference sphere, x_p and y_p are the normalized pupil coordinates, and ϵ_x and ϵ_y are the transverse ray aberrations, as shown in Fig. 2.9.

For a rotationally symmetric optical imaging system, the wavefront aberration can be expressed in a power series as a function of normalized image height H , azimuth angle θ , and normalized pupil height ρ :

$$\begin{aligned} W = & W_{020}\rho^2 + W_{111}H\rho\cos\theta + W_{040}\rho^4 + W_{131}H\rho^3\cos\theta + W_{222}H^2\rho^2\cos^2\theta \\ & + W_{220}H^2\rho^2 + W_{311}H^3\rho\cos\theta + \text{fifth-order terms.} \end{aligned} \quad (2.33)$$

The first two terms in Eq. (2.33) originate from the change of the reference image point, which is the center of the reference sphere. The first term, $W_{020}\rho^2$, is the result of an axial displacement of the reference image point and is called defocus. It can be canceled by proper focusing. The second term, $W_{111}H\rho\cos\theta$, is related

to a displacement of the reference image point perpendicular to the optical axis. It corresponds to a tilt of the reference sphere and is linear in the field variable H ; it describes the difference between the paraxial magnification and the actual magnification of the system, and it can be cancelled by the proper choice of magnification.

In an optical system with defocus W_{020} , the actual image plane is away from the paraxial image plane. Defocus is not field dependent, and it is not considered a real aberration. It can be used to balance other aberrations for better image quality. The transverse ray aberration of defocus is

$$\varepsilon_x = -2 \frac{R}{r_p} W_{020} \rho \cos \theta = -2 \frac{R}{r_p} W_{020} x_p, \quad (2.34a)$$

$$\varepsilon_y = -2 \frac{R}{r_p} W_{020} \rho \sin \theta = -2 \frac{R}{r_p} W_{020} y_p. \quad (2.34b)$$

Figure 2.11 shows the plots of transverse ray aberration and wavefront aberration of a perfect system with defocus only.

The next five terms in Eq. (2.33) represent the five primary monochromatic aberrations: spherical aberration, coma, astigmatism, field curvature, and distortion. Since they all have the power sum of four and their derivatives have the power sum of three, they are called third-order aberrations.

2.4.1 Spherical aberration $W_{040} \rho^4$

Spherical aberration is due to the variation in focal length with the radial zone in the pupil. It causes a transverse ray aberration that varies as the third

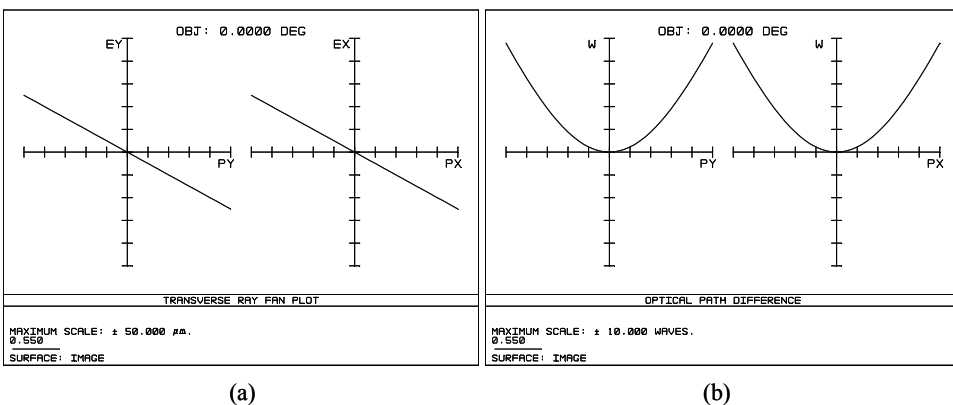


Figure 2.11 (a) Transverse ray and (b) wavefront aberrations in the presence of defocus.

power of the pupil height ρ and is independent of the field parameter H . As shown in Fig. 2.12, the rays with different pupil heights are focused toward different locations other than the paraxial image plane.

The magnitude of spherical aberration is proportional to the cube of the ray height at the pupil plane. When the spherical aberration is measured along the optical axis, it is called longitudinal spherical aberration (LSA), the longitudinal ray error ε_z of which is approximated as

$$\varepsilon_z = -16(f/\#)^2 W_{040} y_p^2. \quad (2.35)$$

Generally, spherical aberration is measured as a transverse ray aberration:

$$\varepsilon_x = -4 \frac{R}{r_p} W_{040} \rho^3 \sin \theta, \quad (2.36a)$$

$$\varepsilon_y = -4 \frac{R}{r_p} W_{040} \rho^3 \cos \theta. \quad (2.36b)$$

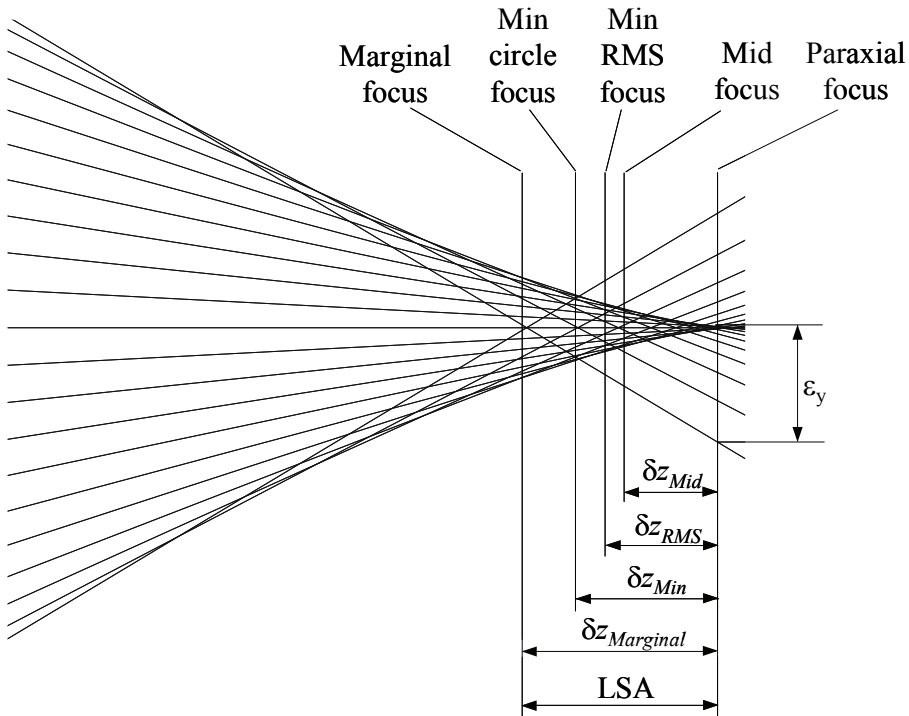


Figure 2.12 Caustic of spherical aberration and aberration balance between spherical aberration and defocus.

The ray fans are

$$\varepsilon_x = -4 \frac{R}{r_p} W_{040} x_p^3, \quad (2.36c)$$

$$\varepsilon_y = -4 \frac{R}{r_p} W_{040} y_p^3. \quad (2.36d)$$

A spherical aberration with a negative sign is called undercorrected, which means the marginal ray intersects the optical axis in front of the paraxial image plane.

For an optical system with spherical aberration, the best image quality is not obtained at the paraxial image plane but between the marginal focus and the paraxial focus. Several focus criteria provide a balance between spherical aberration and defocus: Mid focus is the center between the paraxial focus and the marginal focus, min RMS is the minimum root-mean-square (RMS) spot size, and min circle is the smallest circle containing all rays. The amount of defocus at different planes is as follows:¹⁷

$$\text{LSA} \approx -16(f/\#)^2 W_{040}. \quad (2.37)$$

$$\text{Mid focus:} \quad W_{020} = -W_{040}, \quad \delta z_{\text{Mid}} = 0.5 \text{LSA}. \quad (2.38a)$$

$$\text{Min RMS:} \quad W_{020} = -1.33W_{040}, \quad \delta z_{\text{rms}} = 0.67 \text{LSA}. \quad (2.38b)$$

$$\text{Min circle:} \quad W_{020} = -1.5W_{040}, \quad \delta z_{\text{Min}} = 0.75 \text{LSA}. \quad (2.38c)$$

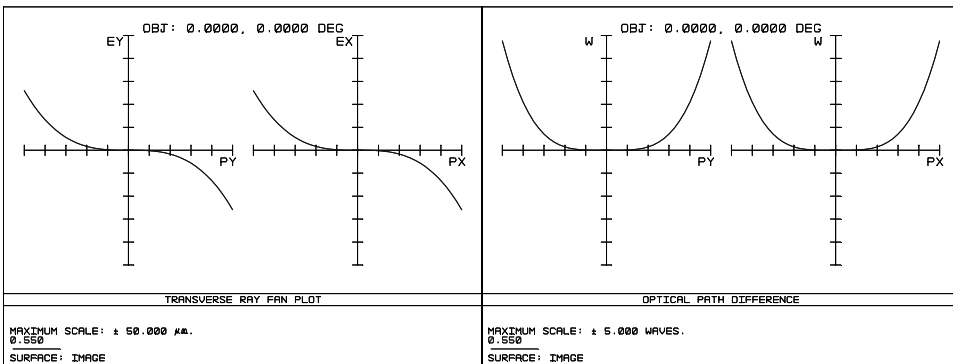
$$\text{Marginal focus:} \quad W_{020} = -2W_{040}, \quad \delta z_{\text{Marginal}} = \text{LSA}. \quad (2.38d)$$

The transverse ray and wavefront aberrations in the presence of spherical aberration without and with compensating defocus are plotted in Figs. 2.13(a) and 2.13(b), respectively.

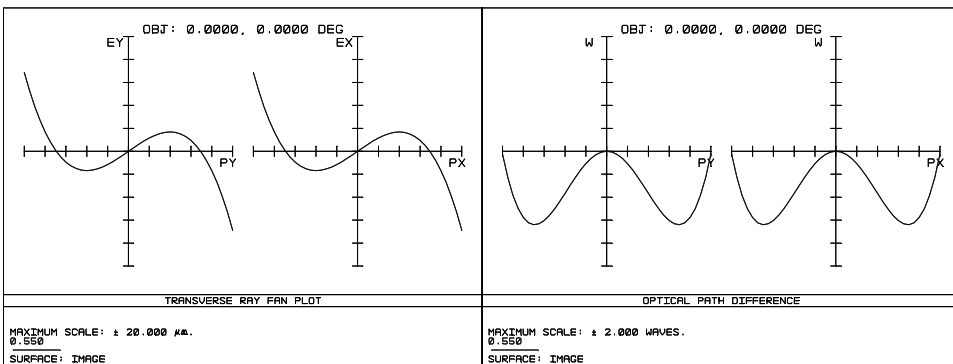
Spherical aberration is the only third-order monochromatic aberration that is seen on axis. It is dependent on aperture and does not have through-focus symmetry. Doubling the aperture results in 16 times the wavefront aberration.

2.4.2 Coma $W_{131} H \rho^3 \cos \theta$

Coma is an off-axis aberration due to the variation of image location and size with the zonal radius in the pupil for an off-axis object point. It can also be regarded as the variation of magnification with pupil position since the size of the ring increases as the radial zone moves out to the edge of the pupil. Rays that



(a)



(b)

Figure 2.13 Plots of transverse ray and wavefront aberrations in the presence of spherical aberration (a) without and (b) with compensating defocus.

transmit through the lens through different zones of the aperture stop cross the image plane at different heights from the optical axis. As shown in Fig. 2.14, only ray pairs from opposite sides of the zone intersect at the same point on the paraxial plane; each annular zone in the exit pupil maps to a displaced ring of light in the paraxial plane, resulting in a comet-like image blur.

In a system with coma, most of the light energy is concentrated in the vicinity of the chief ray. Coma is an odd aberration, linearly proportional to the FOV and proportional to the square of the aperture. The sizes of the rings and their displacement from the Gaussian image point increase with the radius of the zone in the pupil. This causes blurring in the image plane for off-axis object points. A lens with considerable coma may still produce a sharp image in the center of the field, but the image would become increasingly blurred toward the edges.

The transverse ray aberration is

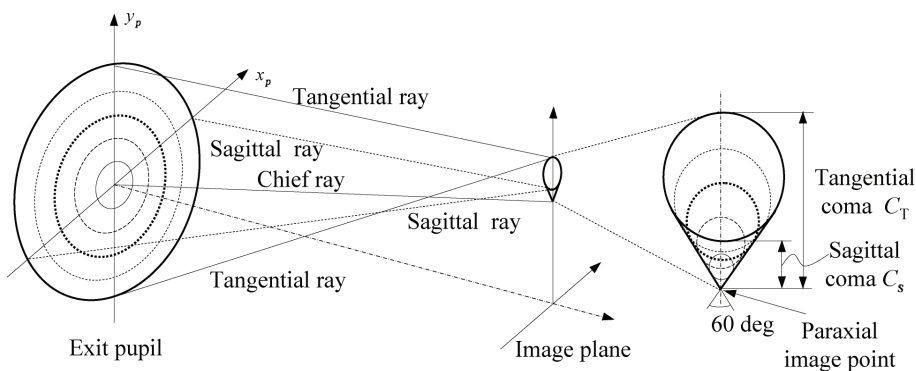


Figure 2.14 Schematic ray path for primary coma.

$$\epsilon_x = -\frac{R}{r_p} W_{131} H \rho^2 \sin 2\theta, \tag{2.39a}$$

$$\epsilon_y = -\frac{R}{r_p} W_{131} H \rho^2 (2 + \cos 2\theta). \tag{2.39b}$$

The ray fans are

$$\epsilon_y = -3 \frac{R}{r_p} W_{131} H y_p^2, \tag{2.39c}$$

$$\epsilon_x = 0. \tag{2.39d}$$

The transverse ray and wavefront aberrations in the presence of coma are plotted in Fig. 2.15.

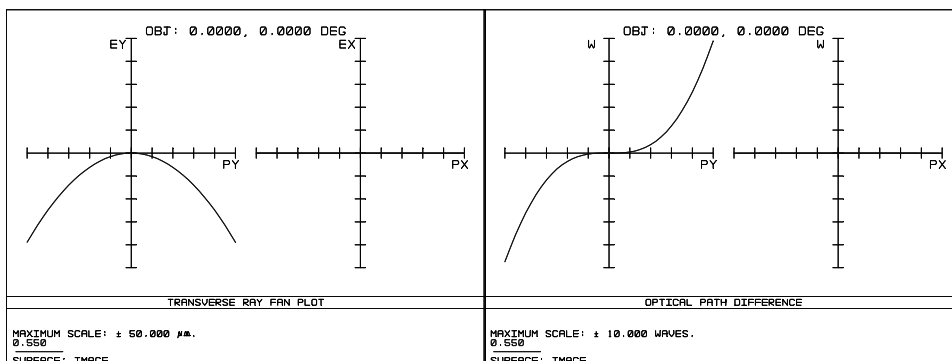


Figure 2.15 Plots of transverse ray and wavefront aberrations in the presence of coma.

Tangential coma C_T , defined as the distance from the chief ray intersection point to the tangential ray intersection point, and sagittal coma C_s , defined as the distance from the chief ray intersection point to the sagittal ray intersection point, are another two measures of coma. Tangential coma is always 3 times sagittal coma.

2.4.3 Astigmatism $W_{222}H^2\rho^2\cos^2\theta$

In the presence of astigmatism, rays in the meridional and sagittal planes are not focused at the same distance from the lens. Astigmatism can be treated as the variation of focus with pupil orientation.

When the light cone from an off-axis point enters a lens surface obliquely, its footprint is elliptical. It has more power in the tangential direction than in the sagittal direction, causing the tangential ray fan to focus closer to the lens than does the sagittal ray fan. As shown in Fig. 2.16, rays in the tangential plane focus along the line that is perpendicular to the tangential plane. Rays in the sagittal plane focus farther from the lens, along the line perpendicular to the sagittal plane. The length of the focused line is approximately $8(f/\#)W_{222}H^2$. Between the astigmatic foci, the image of a point is blurred with the shape of an ellipse or circle. The smallest size of an image blur is halfway between two astigmatic foci, and the blur is a circle with a diameter of $4(f/\#)W_{222}H^2$.

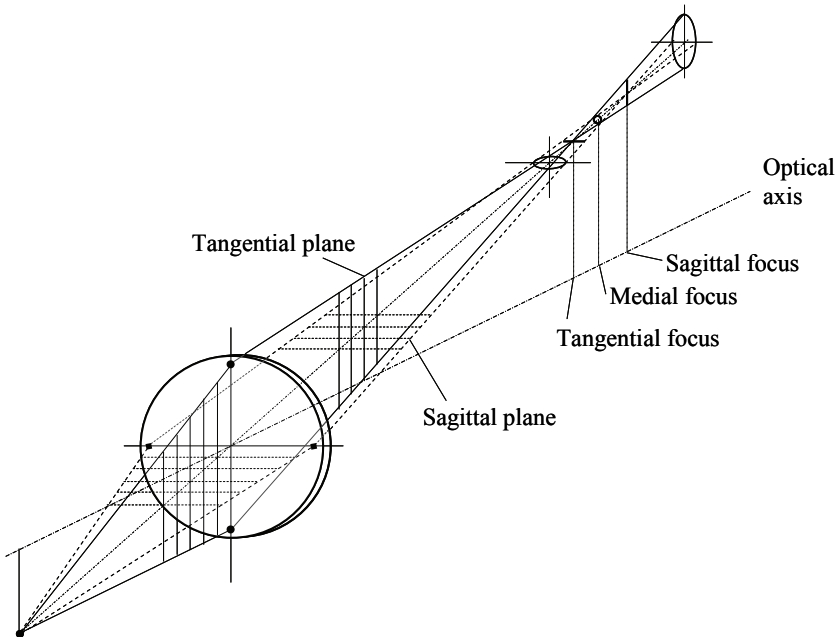


Figure 2.16 Astigmatic focal lines and astigmatism.

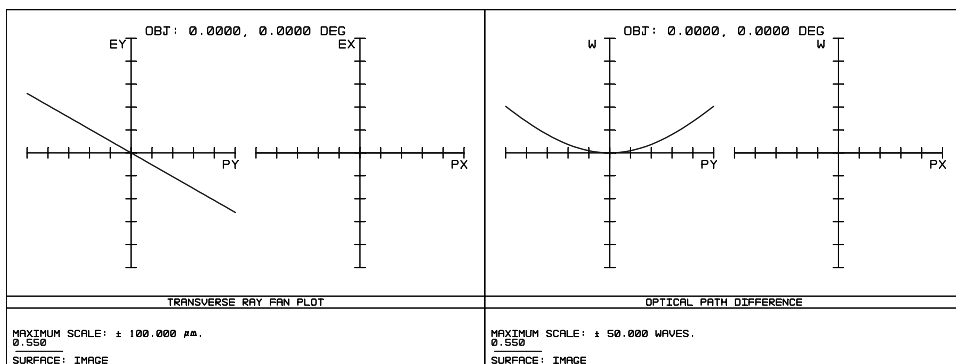
Astigmatism is defined as the variation of focus with the azimuthal angle in the pupil plane. It is linearly proportional to the lens aperture and to the square of the field angle. There is no astigmatism on axis. The transverse ray aberration with astigmatism only is

$$\epsilon_x = 0, \tag{2.40a}$$

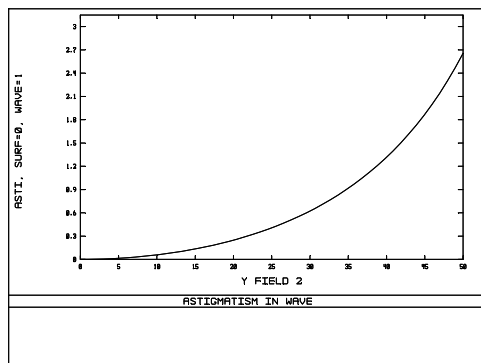
$$\epsilon_y = -2 \frac{R}{r_p} W_{222} H^2 y_p. \tag{2.40b}$$

Figure 2.17(a) plots the transverse ray and wavefront aberrations of an optical system with only astigmatism present. Astigmatism can also be plotted as a function of the field, like longitudinal aberration. Figure 2.17(b) shows longitudinal astigmatism in waves. Using the sagittal focus as the reference plane, longitudinal astigmatism is¹⁷

$$\delta z \approx -8(f/\#)^2 W_{222} H^2. \tag{2.41}$$



(a)



(b)

Figure 2.17 (a) Plots of transverse ray and wavefront aberrations in the presence of astigmatism; (b) plot of longitudinal astigmatism.

2.4.4 Field curvature $W_{220}H^2\rho^2$

Even in the absence of astigmatism, a lens surface has a tendency to image a flat object to a curved surface rather than to a flat surface. This effect is called field curvature and is illustrated in Fig. 2.18, where the aperture stop is located at the center of the surface curvature. The astigmatism is zero because of the concentric surface and chief ray position.

Field curvature is a field-dependent longitudinal focal shift that varies with the square of the field angle. It is closely related to astigmatism. When the optical system has no astigmatism, the sagittal and tangential images coincide with a surface called the Petzval surface. For a single surface, the radius of the Petzval surface is determined by its radius r and the refractive indices before and after refraction, n and n' , respectively:

$$r_p = \frac{-nr}{n' - n}. \quad (2.42)$$

For an optical system with m surfaces, the Petzval curvature is⁶

$$\frac{1}{r_p} = -n'_m \sum_i \frac{n'_i - n_i}{n'_i n_i r_i}, \quad (2.43)$$

where n_i and n'_i are the refractive indices of the media in the object side and image side of surface i with a radius of r_i . For an optical system consisting of thin lenses in the air, the Petzval curvature is

$$\frac{1}{r_p} = -\sum_i \frac{1}{n_i f_i}, \quad (2.44)$$

where f_i is the focal length of the i th lens.

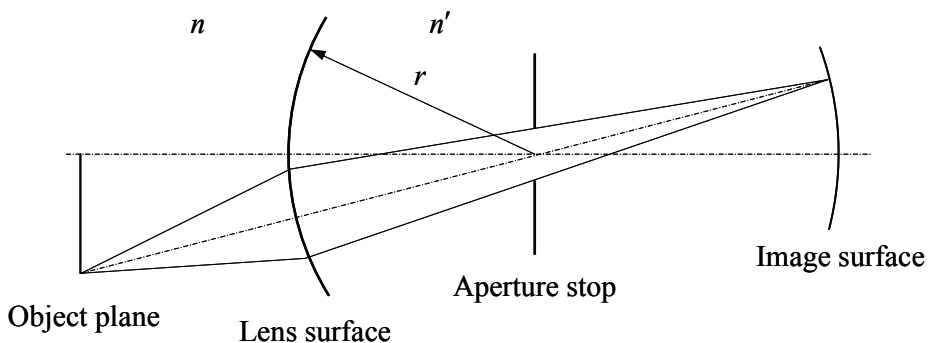


Figure 2.18 Illustration of field curvature.

The Petzval surface is not a good image surface; it is the fundamental field curvature of the system. The Petzval curvature is inversely proportional to the product of the refractive index of the lens and its focal length. If there are many components in an optical system, the Petzval curvature is the sum of the Petzval curvature contributions from all lenses. When the optical system has astigmatism, both the sagittal and tangential image planes are closer to the lens than is the Petzval surface, and the tangential image is three times farther from the Petzval surface than the sagittal image.

The transverse ray aberration of the system with field curvature is

$$\varepsilon_x = -2 \frac{R}{r_p} W_{220} H^2 x_p, \quad (2.45a)$$

$$\varepsilon_y = -2 \frac{R}{r_p} W_{220} H^2 y_p. \quad (2.45b)$$

Field curvature is often plotted as longitudinal aberration as a function of the field. Figure 2.19 shows four image surfaces that are equally spaced. A medial surface can be obtained by balancing astigmatism and field curvature.

2.4.5 Distortion $W_{311} H^3 \rho \cos \theta$

Distortion occurs because the real ray bends more than the paraxial ray. It is a field-dependent magnification error that varies as the third power of the FOV.

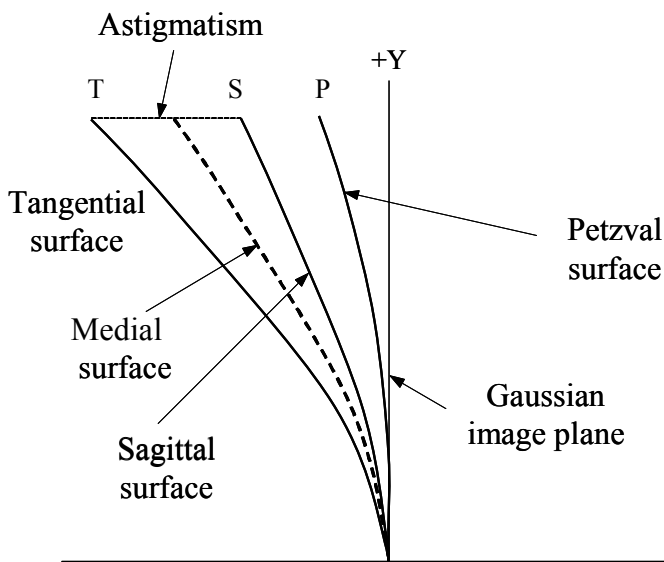


Figure 2.19 Longitudinal field curvature and astigmatism.

This aberration is odd given that it is proportional to an odd power of field H . Distortion is the variation of tilt with the field; it does not blur the image. Points are still imaged into points; however, straight lines are not imaged into straight lines.

The amount of distortion is generally calculated as a percentage of the paraxial image height. Distortion is defined as

$$\text{Distortion} = \frac{y - y_{\text{para}}}{y_{\text{para}}} \cdot 100\%, \quad (2.46)$$

where y is the image height in the paraxial image plane and y_{para} is the paraxial image height.

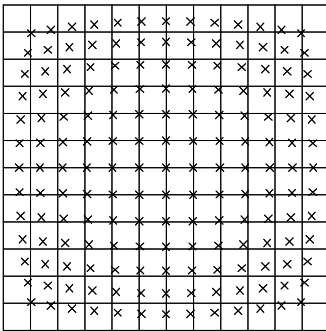
The transverse ray aberration with only distortion is

$$\varepsilon_x = 0, \quad (2.47a)$$

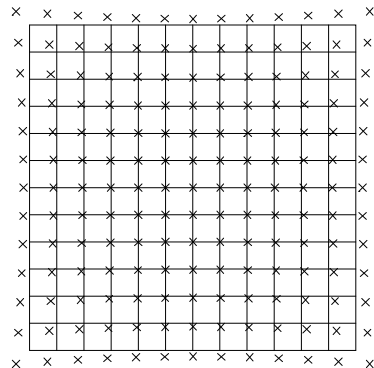
$$\varepsilon_y = -\frac{R}{r_p} W_{311} H^3. \quad (2.47b)$$

The transverse ray fan for distortion is constant with respect to y_p .

When the magnification decreases with the off-axis distance of the object, the distortion is negative, and each image point moves radially toward the center of the image; this is called barrel distortion. When the magnification increases with the off-axis distance of the object, the distortion is positive, and each image point is displaced radially outward from the center; this distortion is called pincushion distortion. Figure 2.20 shows barrel and pincushion distortion.



(a)



(b)

Figure 2.20 Grid distortion: (a) barrel distortion and (b) pincushion distortion.

2.4.6 Chromatic aberrations

Due to material dispersion, the focal length of a lens is a function of wavelength because the focal length is dependent on the refractive index. A lens without correction of this error will focus the light from the same point with different wavelengths at different locations, as shown in Fig. 2.21. These chromatic errors are called chromatic aberrations. Chromatic aberrations are first-order aberrations and are present in any refractive optical system.

There are two types of chromatic aberrations (1) chromatic variations in paraxial image forming properties of a system, and (2) wavelength dependence of monochromatic aberrations. The first type of chromatic aberration is generally of the most interest.

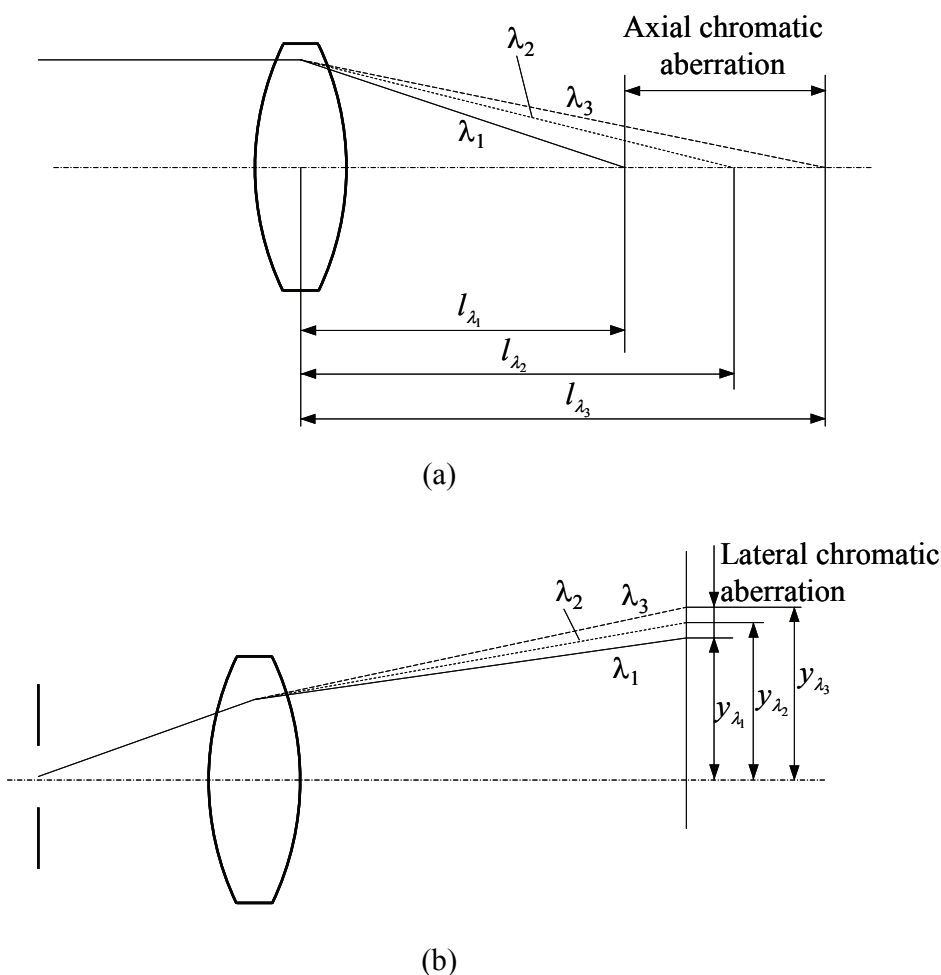


Figure 2.21 Chromatic aberrations: (a) axial chromatic aberration and (b) lateral chromatic aberration. In this figure, $\lambda_1 < \lambda_2 < \lambda_3$.

The variation of focal length with wavelength is called axial (longitudinal) chromatic aberration. This is a chromatic aberration of the marginal ray and thus of the image position. Axial chromatic aberration causes light of different wavelengths to focus at different planes, resulting in blur of the image in the paraxial image plane. The variation of magnification with wavelength is called lateral chromatic aberration. This is chromatic aberration of the chief ray and thus of the image size. Lateral chromatic aberration causes light of different wavelengths to focus at different heights in the paraxial image plane.

Axial chromatic aberration is defined as the axial distance between two image points for different wavelengths, as shown in Fig. 2.21(a):

$$\delta z_\lambda = l_{\lambda_3} - l_{\lambda_1}. \quad (2.48)$$

It is linearly proportional to the pupil height and can also be expressed as wavefront aberration $\delta_\lambda W_{020} \rho^2$.

Lateral chromatic aberration is defined as the difference of image heights for different wavelengths [as shown in Fig. 2.21(b)].

$$\varepsilon_{y_\lambda} = y_{\lambda_3} - y_{\lambda_1}. \quad (2.49)$$

It is linearly proportional to the FOV. Expressed as wavefront aberration, lateral chromatic aberration is $\delta_\lambda W_{111} H \rho \cos \theta$.

When chromatic aberrations are corrected for wavelengths λ_1 and λ_3 , the middle of the spectrum, λ_2 , behaves differently. The chromatic aberration that remains after the primary chromatic aberrations are corrected is called the secondary chromatic aberration. The secondary chromatic aberration must be corrected for some high-performance imaging systems used in the broadband spectrum.

2.5 Seidel Aberrations

The surface contribution to the primary aberrations can be calculated by tracing two paraxial rays, the chief and marginal rays, as shown in Fig. 2.22. The derivation was first performed by L. Seidel in 1856, and the corresponding aberration coefficients are called Seidel aberrations. Seidel aberrations have five terms (S_I , S_{II} , S_{III} , S_{IV} , and S_V) for monochromatic aberrations and two terms (C_I and C_{II}) for chromatic aberrations.

To make the equations simple, the following two paraxial quantities are defined for each surface of the system:

$$A = n(y'c + u) = n'(y'c + u') = ni = n'i', \quad (2.50a)$$

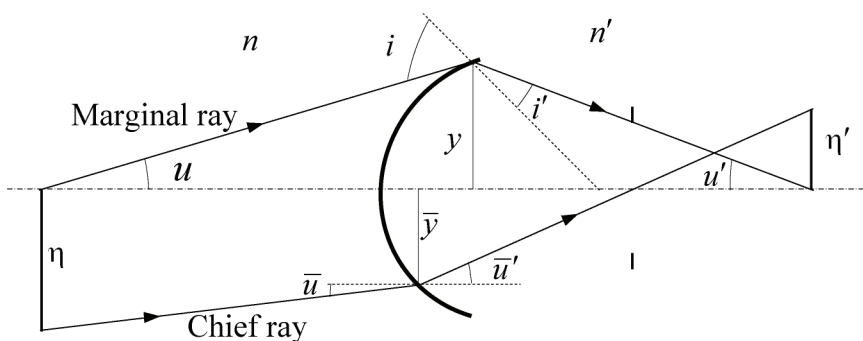


Figure 2.22 Chief ray and marginal ray to calculate the Seidel-aberration terms.

$$\bar{A} = n(\bar{y}c + \bar{u}) = n'(\bar{y}'c + \bar{u}') = n\bar{i} = n'\bar{i}'. \quad (2.50b)$$

As shown in Fig. 2.22, all quantities with an overbar refer to the chief ray, and those without an overbar refer to the marginal ray. Here n and n' are the refractive indices; i and i' are the paraxial incidence angles; u and u' are the paraxial ray angles before and after refraction, respectively; y is the paraxial incidence height at the surface; and c is the surface curvature.

The Seidel sums are defined as follows:¹³

$$S_{\text{I}} = -\sum A^2 y \Delta \left(\frac{u}{n} \right), \quad (2.51a)$$

$$S_{\text{II}} = -\sum A \bar{A} y \Delta \left(\frac{u}{n} \right), \quad (2.51b)$$

$$S_{\text{III}} = -\sum \bar{A}^2 y \Delta \left(\frac{u}{n} \right), \quad (2.51c)$$

$$S_{\text{IV}} = -\sum H^2 c \Delta \left(\frac{1}{n} \right), \quad (2.51d)$$

$$S_{\text{V}} = -\sum \frac{\bar{A}}{A} \left[H^2 c \Delta \left(\frac{1}{n} \right) + \bar{A}^2 y \Delta \left(\frac{u}{n} \right) \right], \quad (2.51e)$$

$$C_{\text{I}} = \sum A y \Delta \left(\frac{\delta n}{n} \right), \quad (2.51f)$$

$$C_{\text{II}} = \sum \bar{A}y\Delta\left(\frac{\delta n}{n}\right), \quad (2.51g)$$

where

$$\delta\left(\frac{u}{n}\right) = \frac{u'}{n'} - \frac{u}{n}, \quad (2.52a)$$

$$H = nu\eta, \quad (2.52b)$$

$$\Delta\left(\frac{\delta n}{n}\right) = \frac{n-1}{nv}, \quad (2.52c)$$

$$v = \frac{n_d - 1}{n_F - n_c}. \quad (2.52d)$$

The above Seidel sums relate to the wavefront aberration coefficients as

$$W_{040} = \frac{1}{8}S_{\text{I}}, \quad (2.53a)$$

$$W_{131} = \frac{1}{2}S_{\text{II}}, \quad (2.53b)$$

$$W_{222} = \frac{1}{2}S_{\text{III}}, \quad (2.53c)$$

$$W_{220} = \frac{1}{4}S_{\text{IV}}, \quad (2.53d)$$

$$W_{311} = \frac{1}{2}S_{\text{V}}, \quad (2.53e)$$

$$\delta_\lambda W_{020} = \frac{1}{2}C_{\text{I}}, \quad (2.53f)$$

$$\delta_\lambda W_{111} = \frac{1}{2}C_{\text{II}}. \quad (2.53g)$$

2.6 Performance Evaluations

During the process of optimizing an optical system in the design stage and again after assembling, the system's performance must be evaluated frequently. Except for the transverse ray and wavefront aberrations discussed in Sec. 2.4, other image quality criteria are also commonly used. These criteria include spot diagram, modulation transfer function (MTF), encircled energy, RMS spot radius, RMS wavefront error, and Strehl ratio. Each of these criteria relates in different ways to the image quality of the system.

It is important to note that the image quality of an optical system is dependent not only on the optics but also on the sensor, display device, and other components in the system. For example, when the human eye is the final sensor, the system requirement on defocus and field curvature may be less stringent because of the eye's natural ability to accommodate defocus and field curvature. However, digital imaging sensors whose photosensitive elements are flat cannot compensate for such phenomena.

2.6.1 Transverse ray aberration

Figure 2.23 shows two typical plots of transverse ray aberrations as a function of the pupil coordinate. Note that no skew rays are considered in the plots. The aberration characteristics and dependence on pupil height and wavelength can easily be observed and analyzed. For example, Fig. 2.23(a) shows that the system has a reasonable amount of lateral chromatic aberration, while Fig. 2.23(b) illustrates that the lateral chromatic aberration is well controlled but the axial chromatic aberration needs to be improved. Other aberrations, such as spherical aberration and coma, are also clearly shown in both plots.

Generally, only the maximum aberration values are related to image quality, but the shapes and slopes of the aberration curves provide significant information on aberration characteristics. For example, a slope not only provides the values of the aberrations but also indicates the existence of higher-order aberrations.

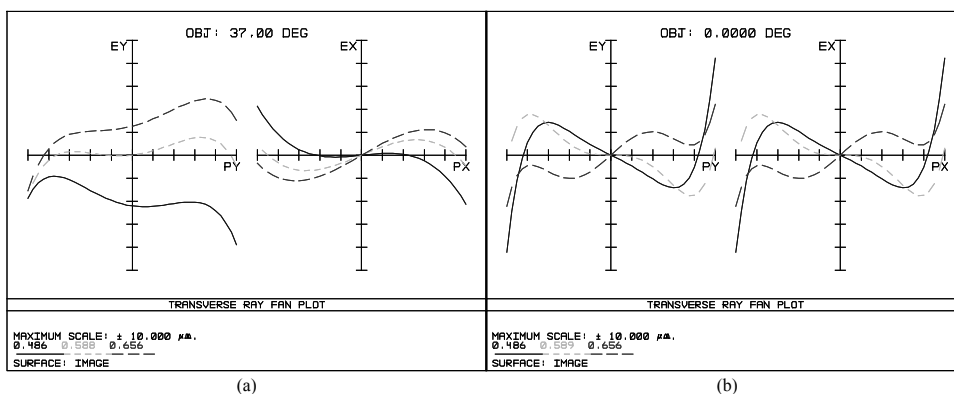


Figure 2.23 Plots of transverse ray aberrations.

Ray fan plots are very useful for assessing the existence and magnitude of most aberrations present in an optical system. Some guidelines for identifying the existence of aberrations from ray fan plots are as follows:

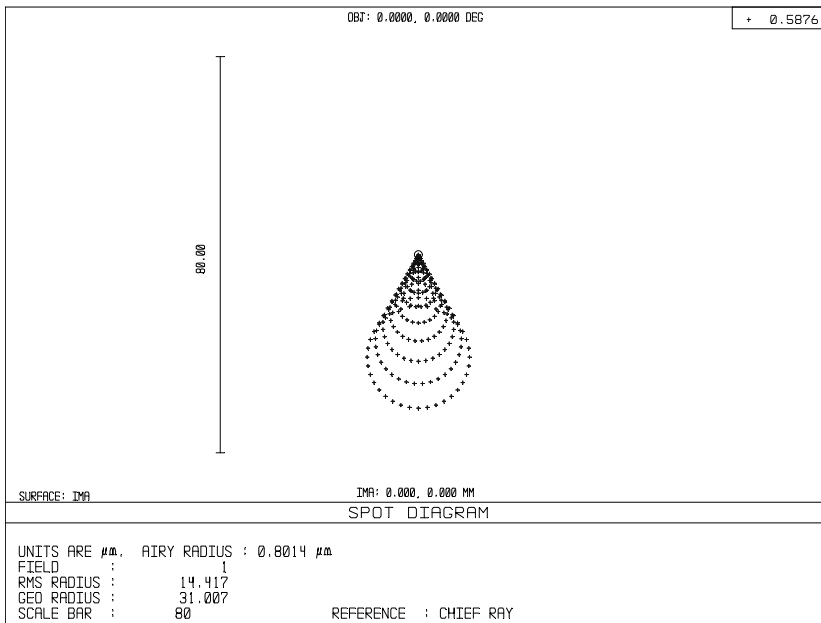
- Spherical aberration: The curve is not a linear function of the pupil height at $H = 0$.
- Coma: The intersected point of the line connecting the two ends of the tangential ray fans at the vertical axis is not at the origin. In other words, $\varepsilon_y(y_p = 1) \neq -\varepsilon_y(y_p = -1)$.
- Astigmatism: The slopes of the tangential and sagittal ray fans at the origin are different.
- Field curvature: The slopes through the origin at $H = 0$ and $H = 1$ are different.
- Defocus: The primary wavelength has a slope through the origin at $H = 1$.
- Axial chromatic aberration: The slopes for the two end wavelengths are different.
- Lateral chromatic aberration: The curves do not cross at the origin of the tangential field at $H = 1$.

2.6.2 Spot diagram

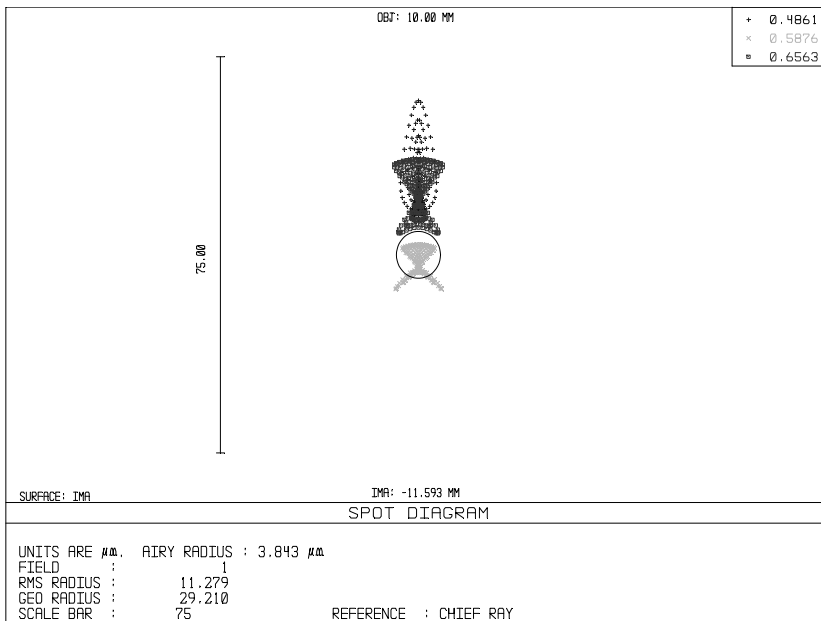
A spot diagram is another means of describing transverse ray aberrations. It is a plot of the ray intersection points relative to the reference image point at the image plane when rays are traced from a single object point through a uniform grid at the entrance pupil. Each ray corresponds to the same amount of energy. A spot diagram gives a visual representation of the image of an object point. Figure 2.24(a) illustrates the spot diagram for a lens with coma only. The shape of the spot diagram shows this characteristic. However, it is difficult to distinguish typical aberrations from spot diagrams of a system that has a combination of various aberrations. For example, Fig 2.24(b) displays the spot diagram of an off-axis point. It shows some lateral chromatic aberration and coma but no additional information on the other aberrations.

Through-focus spot diagrams show the characteristics of the aberrations and give the estimated size of the image blur. Figure 2.24(c) is a through-focus spot diagram for a lens with astigmatism only, giving information on the two focused lines. Figure 2.24(d) is another through-focus spot diagram for a lens with spherical aberration only. The figure clearly shows that the best focus is not at the paraxial focal plane, but rather is $\sim 50 \mu\text{m}$ closer to the lens.

A spot diagram does not include the diffraction effect, so it does not give a correct figure of merit for diffraction-limited systems. A better measure of spot size is the RMS spot size, which will be discussed in Sec. 2.6.5. When working with pixilated sensors, the RMS spot radius is very useful for checking whether the image of an object point falls within a pixel.

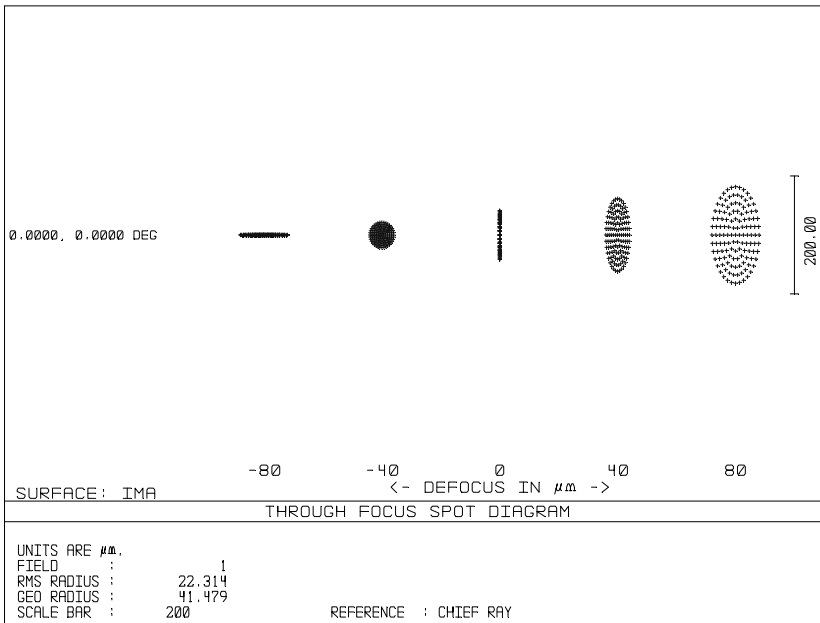


(a)

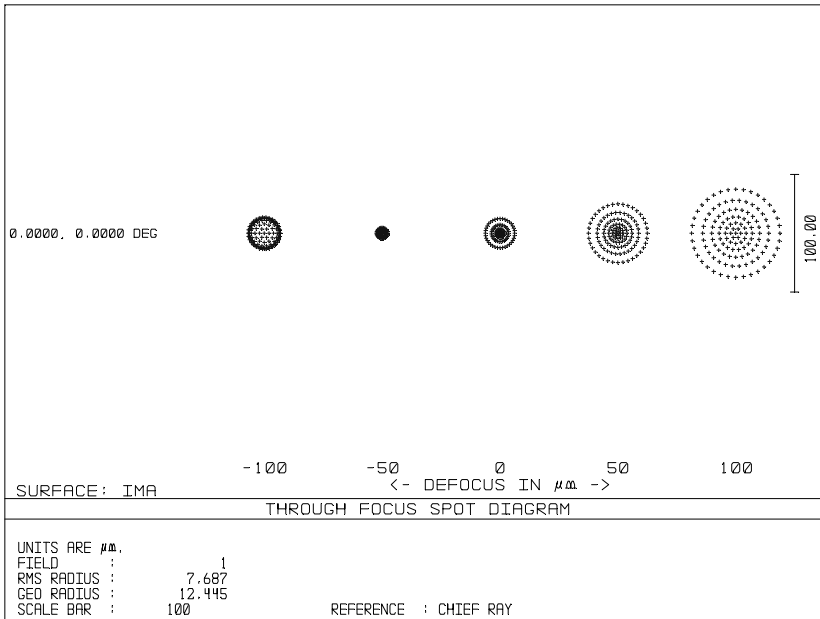


(b)

Figure 2.24 Spot diagrams: (a) spot diagram for a system with coma only and (b) off-axis spot diagram on the paraxial image plane. (Continued on next page).



(c)



(d)

Figure 2.24 (continued) Spot diagrams: (c) through-focus spot diagram for a system with astigmatism only, and (d) through-focus spot diagram for a system with spherical aberration only.

2.6.3 Encircled/ensquared energy

Encircled or ensquared energy is the percentage of total energy enclosed as a function of distance from either the chief ray or the centroid of the image of an object point. The details of intensity distribution in the point spread function (PSF) are ignored.

The advantage of using encircled/ensquared energy is that a single number can be obtained. This number is the diameter of a circular area or the dimension of a square area on the image surface containing a specified fraction of the energy. Such a number is extremely useful for digital imaging systems using a charge-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) sensor and confocal imaging system. The ensquared energy function is appropriate for an imaging system with square pinholes or square detection pixels. As a rule of thumb, 80% of the energy from an object point should fall within one pixel or pinhole for an imaging system with good image quality.

Figure 2.25 shows an example of encircled energy for a microscope objective lens with an NA of 0.65 and a magnification of 40. The diffraction-limited distribution is plotted for reference. For an on-axis point, 80% of the energy is within a circular aperture with a radius of $0.85 \mu\text{m}$, while an aperture with a radius of $2.5 \mu\text{m}$ is required to contain 80% of the total energy for an off-axis point.

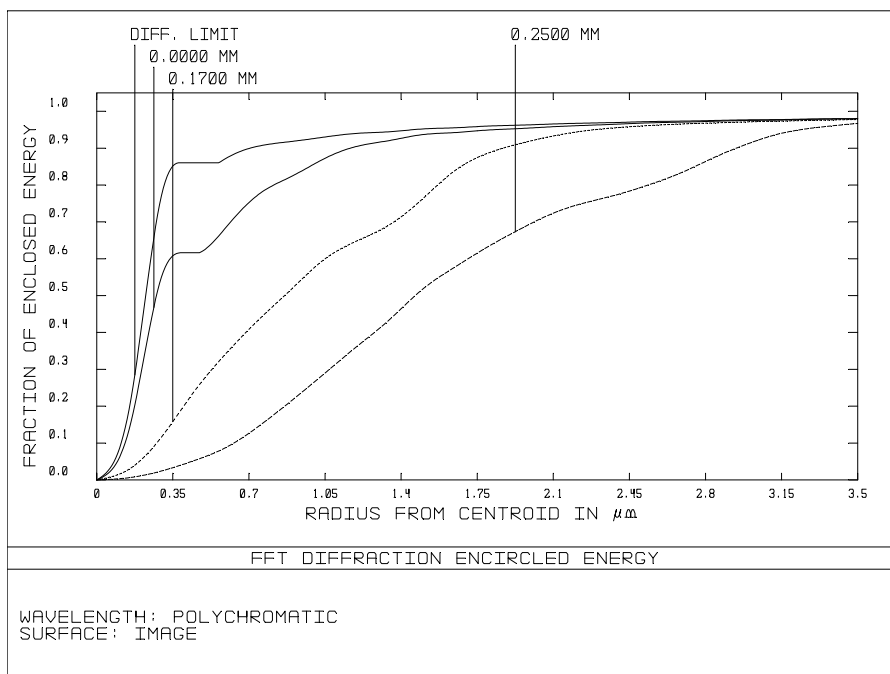


Figure 2.25 Encircled energy for a 40 \times /0.65 microscope objective lens.

2.6.4 Modulation transfer function

Aberration plots provide a good indication of the lens performance, but some aberrations may not be easy to measure. Having a measurable, objective performance criterion to evaluate the optical system is preferable.

The modulation transfer function (MTF) is the most widely used criterion; MTF describes the ability of an optical system to transfer object contrast to an image. It represents the transfer of modulation from the object to the image by the lens as a function of spatial frequency. MTF is defined as the ratio of the image contrast to the object contrast and is usually given by the number of cycles per mm or line pairs per mm, where one cycle corresponds to one line pair (a combination of one bright and one dark line). The cutoff frequency, where the MTF goes to zero, is

$$f_{\text{cutoff}} = \frac{1}{\lambda(f/\#)}. \quad (2.54)$$

MTF is perhaps the most comprehensive performance criterion of an imaging system. One advantage of MTF is that it can be cascaded by simply multiplying the MTFs of two or more components. As a rule of thumb, the MTF for a good imaging system should be larger than 0.5 at the Nyquist frequency. For digital imaging systems where CCD or CMOS sensors are used as the detectors, the Nyquist frequency is equal to $1/[2(\text{pixel size})]$.

Figure 2.26 is the MTF plot of a $4\times/0.1$ microscope objective lens. The separation of the tangential and sagittal MTFs is due to the field-dependent aberration astigmatism.

2.6.5 Root-mean-square wavefront error/spot radius

The RMS wavefront error is an assessment of image quality using the deviation of the wavefront in the exit pupil from the reference sphere. The RMS wavefront error W_{rms} is defined as the square root of the difference between the average of the squared wavefront deviations and the square of the average of the wavefront deviation:

$$W_{\text{rms}} = \sqrt{\langle \Delta W^2 \rangle - \langle \Delta W \rangle^2}. \quad (2.55)$$

When W_{rms} is smaller than $1/14$, the system is treated as a diffraction-limited system.

The RMS spot radius is another quantity that is commonly used in evaluating system performance. It is defined as the square root of the averaged squared distance from the intercept point (x_c, y_c) of the chief ray to the intercept point (x, y) of each ray:

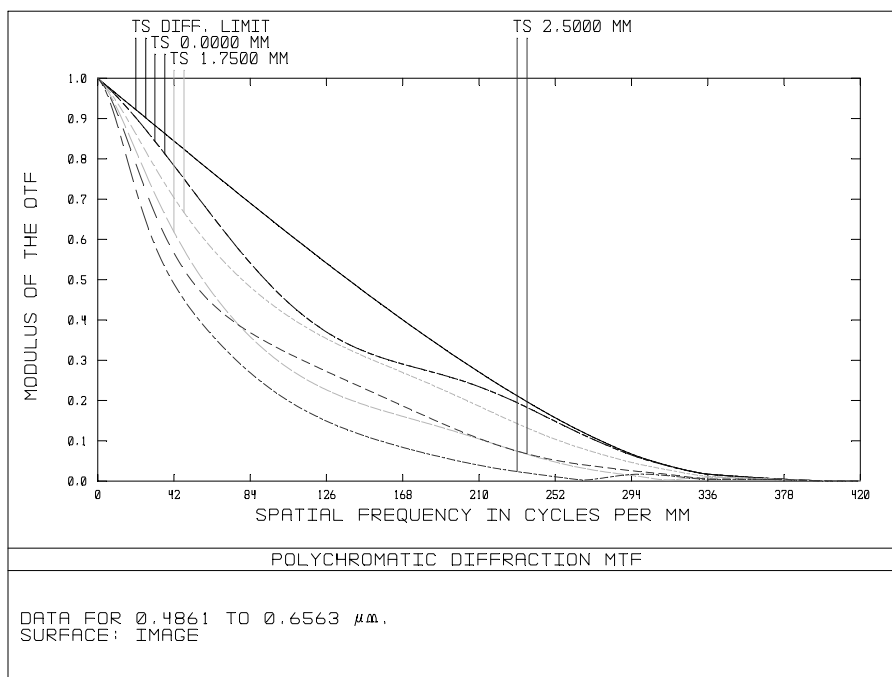


Figure 2.26 MTF plot of a 4x/0.1 microscope objective lens.

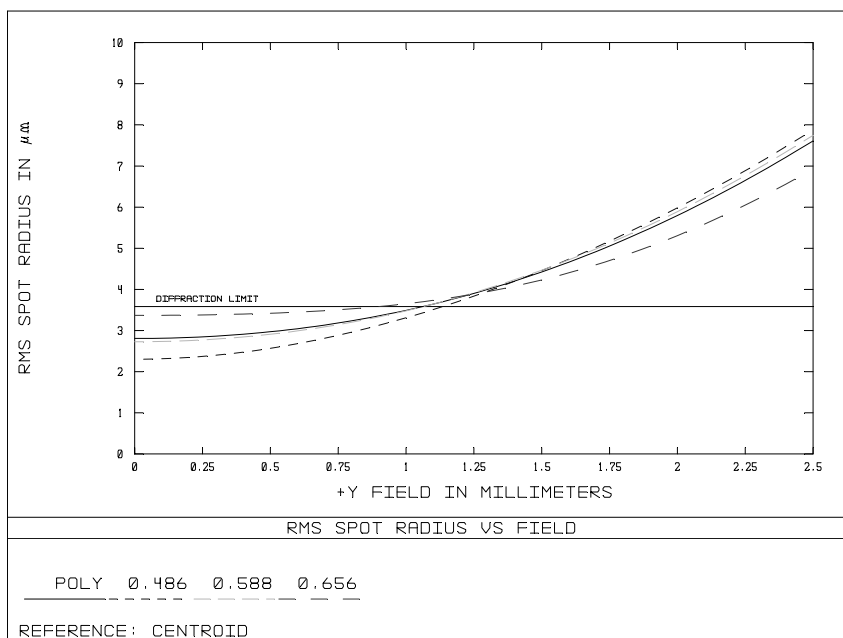
$$r_{\text{rms}} = \left[\sum_{1}^n \frac{(x - x_c)^2 + (y - y_c)^2}{n} \right]^{\frac{1}{2}}. \quad (2.56)$$

The RMS spot radius is a geometric assessment of image quality. When working with pixilated sensors, it is very useful for checking whether the image of an object point falls within a pixel.

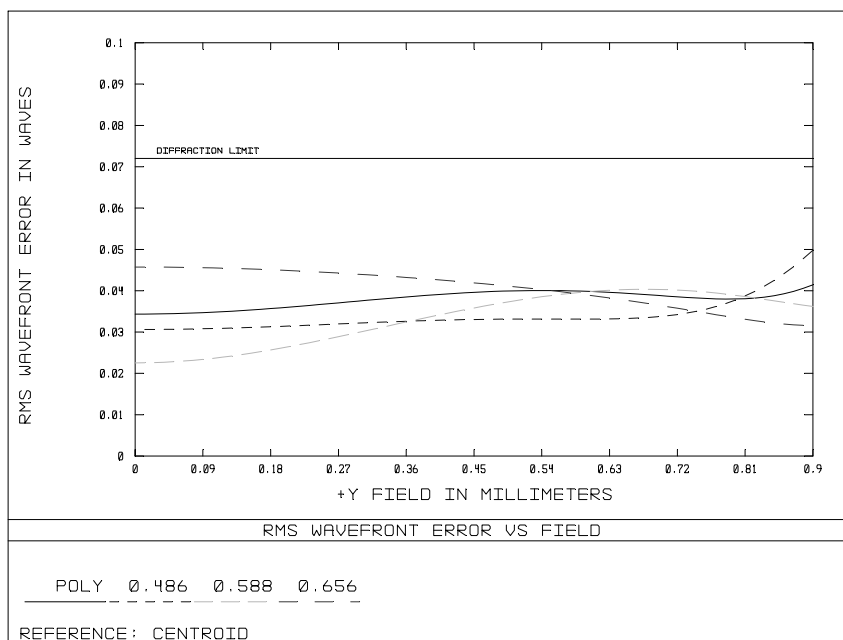
The RMS spot radius is suitable for evaluating systems with large aberrations, and RMS wavefront error is suitable for evaluating systems with small aberrations. These values are frequently used in optical designs to check field-dependent performance. Figure 2.27 shows RMS spot radius and RMS wavefront error of two objective lenses as a function of field. Figure 2.27(a) shows that the 4x/0.1 objective lens has an RMS spot size smaller than the Airy radius within 40% of the field. Figure 2.27(b) shows that the RMS wavefront error of a 10x/0.32 plan apochromatic microscope objective lens is smaller than 1/14, which means that the objective has a corrected field curvature and has a diffraction-limited performance across the entire FOV.

2.6.6 Strehl ratio

PSF is an essential quality criterion of an optical imaging system; however, the complete intensity distribution is a complicated function. The Strehl ratio is



(a)



(b)

Figure 2.27 (a) The RMS spot radius versus FOV for a 4 \times /0.1 microscope objective lens and (b) RMS wavefront error versus FOV for a 10 \times /0.32 plan apochromatic microscope objective lens.

often used as an alternative quantity to evaluate system performance. It is the reduced height of the peaked PSF intensity in the centroid of the energy for the real aberrated system referenced to the ideal aberration-free system. A Strehl ratio of 1 corresponds to an ideal system.

The Strehl ratio provides valuable estimated system performance only for systems with low wavefront aberrations. An approximate relationship between the Strehl ratio SR and the RMS wavefront error is

$$SR = e^{-\left(2\pi \frac{W_{\text{rms}}}{\lambda}\right)^2} \quad (2.57)$$

This approximation is valid for systems having a Strehl ratio greater than 0.1. The Maréchal criterion states that a system is regarded as well corrected if the Strehl ratio is ≥ 0.8 , which corresponds to an RMS wavefront error of $\lambda/14$.

Figure 2.28 shows two plots of the Strehl ratio versus FOV for the 40 \times /0.65 and 10 \times /0.32 microscope objective lenses. The plot in Fig. 2.28(a) indicates that the 40 \times /0.65 objective lens has a significant amount of field curvature since the Strehl ratio decreases quickly from the axis to the edge of the field and has a diffraction-limited performance within 40% of the field. Figure 2.28(b) shows that the Strehl ratio of the 10 \times /0.32 plan apochromatic microscope objective lens is larger than 0.8 for wavelengths of 0.48–0.65 μm , which means that field curvature is corrected satisfactorily across the entire FOV.

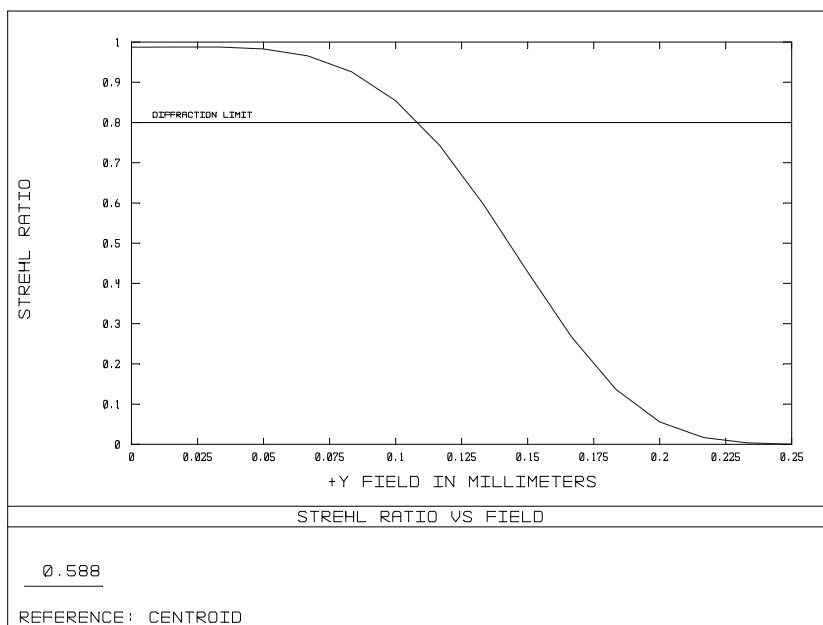
2.7 Correction of Aberrations

As discussed in Sec. 2.4, almost all optical systems have—more or less—aberrations. One of the key tasks of optical design is to control or avoid aberrations. While lens design software is currently very powerful, the basic strategies based on aberration theory still play a key role in optimizing optical systems. Table 2.2 depicts the relative effectiveness of the strategies most commonly used for correcting third-order aberrations.⁶

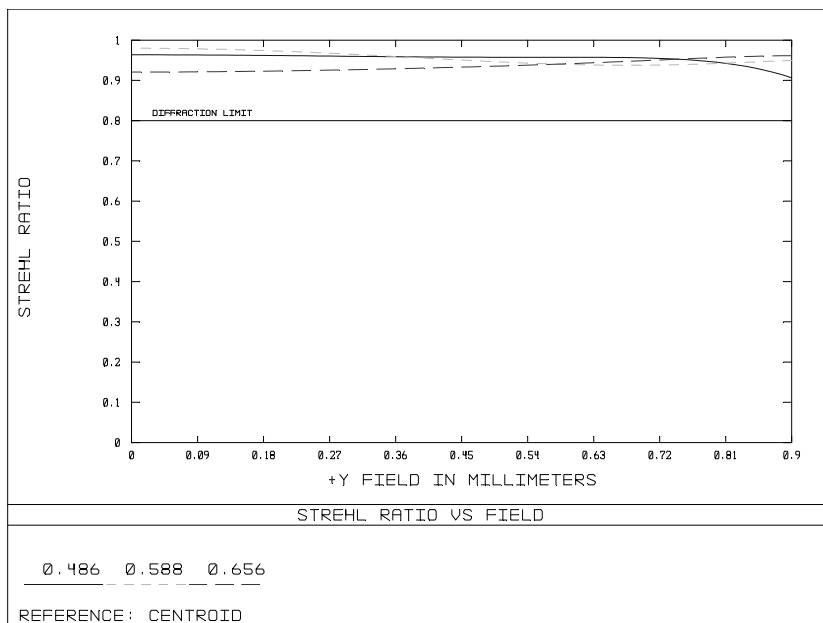
2.7.1 Lens bending

Lens bending involves varying the radii of the lens surfaces but keeping the lens power constant. The locations of the principal points change while the lens is bent into different shapes. For a thin lens with an aperture on its surface, the lens shape has a direct impact on spherical aberration and coma but not on astigmatism, field curvature, distortion, or chromatic aberration.¹⁰ When the stop is away from the lens, the lens shape has an additional impact on astigmatism and distortion. In practical systems, lens bending induces changes in lens separation.

Figure 2.29 illustrates infinite conjugate BK7 singlets with a focal length of 10 mm and an $f/\#$ of 4. The back focal length is adjusted to achieve the smallest RMS spot size possible. The plano-convex lens in Fig. 2.29(b) has the worst performance because the second surface has all of the optical power, and the



(a)



(b)

Figure 2.28 (a) Strehl ratio versus FOV for a 40 \times /0.65 microscope objective lens and (b) Strehl ratio versus FOV for a 10 \times /0.32 plan apochromatic microscope objective lens.

Table 2.2 Effectiveness of common aberration-correction strategies.

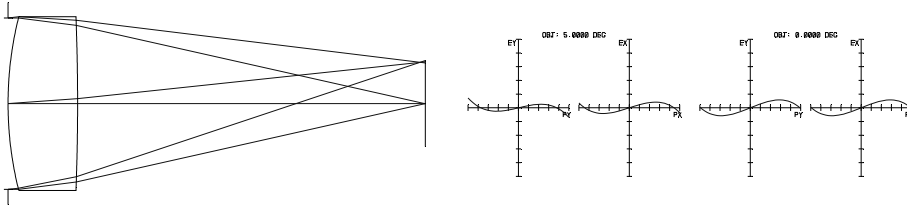
	Spherical aberration	Coma	Astigmatism	Field curvature	Distortion	Axial color	Lateral color
Lens bending	■	■	■		■		
Lens splitting	■	■	■	■	■	■	■
Aspheric surface	■	■	■		■		■
Aplanatic surface	■	■	■				
Cemented surface	■	■	■	■	■	■	■
Optical material	■		■	■	■	■	■
Stop position		■	■		■		■
Symmetry principle		■			■		■
Field lens	■	■	■	■	■	■	■

biconvex lens in Fig. 2.29(c) has the best performance because the lens power is distributed into two surfaces. The performance of the convex-plano lens in Fig. 2.29(a) is close to the biconvex lens. For this reason, the convex surface of a plano-convex focusing lens is usually faced toward the collimated beam.

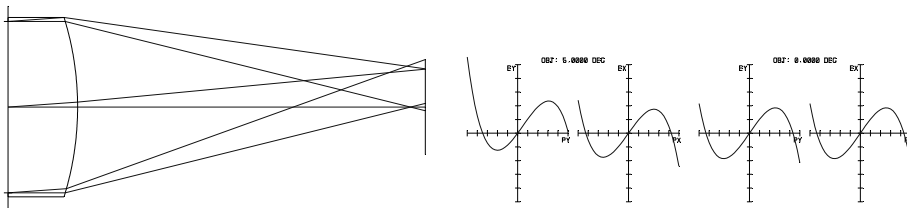
2.7.2 Lens splitting

Lens splitting separates a high-powered lens element into two or more elements. High incidence angles on the lens surfaces can be reduced significantly, which means less aberration will be introduced. By the nonlinearity of Snell's law, the smaller the incidence angle is, the less aberration will be introduced by the surface.

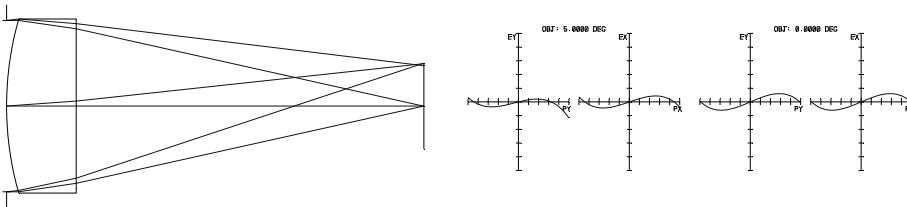
Lens splitting is an effective way to reduce spherical aberration. For a thin lens, the spherical aberration is roughly reduced by a factor of m^2 , where m is the number of splits. The effectiveness of lens splitting in reducing spherical



(a)



(b)



(c)

Figure 2.29 Lens bending for (a) biconvex, (b) plano-convex, and (c) convex-plano lenses.

aberration is clearly demonstrated in Fig. 2.30. While splitting the lens, a focal length of 10 mm and an $f/\#$ of 2 are maintained. Each plot contains three elements, but the number of singlets with optical power is 1–3. The unit of the ray fan plot is the same for all three configurations.

2.7.3 Lens compounding

Lens compounding or lens combination is a method of replacing a singlet with a cemented or air doublet to control ray paths. It is different from lens splitting in

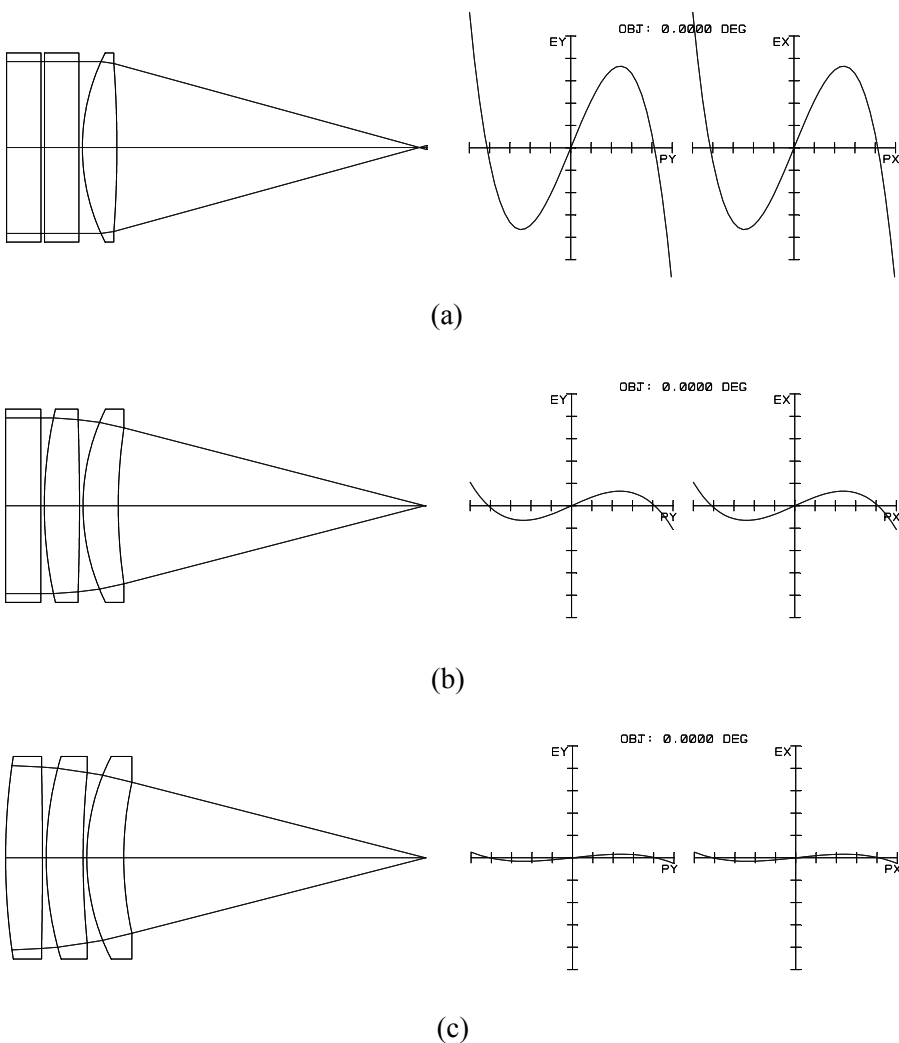


Figure 2.30 Lens splitting and its effect on spherical aberration: (a) one singlet, (b) two singlets, and (c) three singlets.

that the two new elements involved in lens compounding have powers with opposite signs.

Cemented surfaces are widely used in optical systems to control primarily spherical and chromatic aberrations. Spherical aberration can be effectively controlled by lens compounding because an additional surface can reduce the angle of incidence.

To correct the axial chromatic aberration, the total power variation should be zero:

$$\Delta\Phi = \frac{\Phi_1}{v_1} + \frac{\Phi_2}{v_2} = 0, \quad (2.58)$$

and the power of the doublet is

$$\Phi = \Phi_1 + \Phi_2, \quad (2.59)$$

where Φ_1 and Φ_2 are the optical powers and v_1 and v_2 are the Abbe numbers of the two lens elements. Therefore, the powers of the two elements are

$$\Phi_1 = \frac{1}{1 - \frac{v_2}{v_1}} \Phi, \quad (2.60a)$$

$$\Phi_2 = \frac{1}{1 - \frac{v_1}{v_2}} \Phi. \quad (2.60b)$$

Regardless of what materials are used, one lens in the doublet is positive and the other is negative. The ratio of the Abbe numbers should be different from 1. The larger the difference between two Abbe numbers, the smaller the individual lens power and the larger the lens radius. If the ratio of an Abbe number is close to 1, the element power becomes extremely large, and the cemented surface will have strong curvature, leading to large aberrations.

2.7.4 Aplanatic surface

From Eqs. (2.51a), (2.51b), and (2.51c), if a surface meets the condition

$$\Delta\left(\frac{u}{n}\right) = 0, \quad (2.61)$$

it will not introduce spherical aberration, coma, or astigmatism. This surface is called an aplanatic surface. A lens with an aplanatic surface is called an aplanatic

lens. Aplanatic lenses have been widely used in high-NA microscope objective lenses.

From the paraxial ray trace shown in Fig. 2.31, we have the object and image distances:

$$l = \frac{n' + n}{n} R, \quad (2.62a)$$

$$l' = \frac{n' + n}{n'} R. \quad (2.62b)$$

Therefore, l and l' always have the same sign. This means that an aplanatic surface cannot form a real image of a real object; either the object or the image is virtual.

Also from Eqs. (2.51a), (2.51b), and (2.51c), the other two aplanatic conditions are $h=0$ and $A=0$. The first condition means the object and image are both located at the lens surface, and the second condition means that both the object and the image are at the center of curvature.

2.7.5 Aspherical surface

Although there are several types of aspherical surfaces, the most common form of an aspherical surface is a rotationally symmetrical surface with the sag defined as

$$z = \frac{cr^2}{1 + \sqrt{1 - (1+k)c^2r^2}} + \sum a_i r^{2i}, \quad (2.63)$$

where c is the base curvature at the vertex, k is the conic constant, r is the radial coordinate, and $a_i r^{2i}$ is the higher-order aspheric term. The main feature of the additional terms, other than the radius, is that higher-order aspheric terms do not influence the paraxial imaging and can be used to control the aberrations.

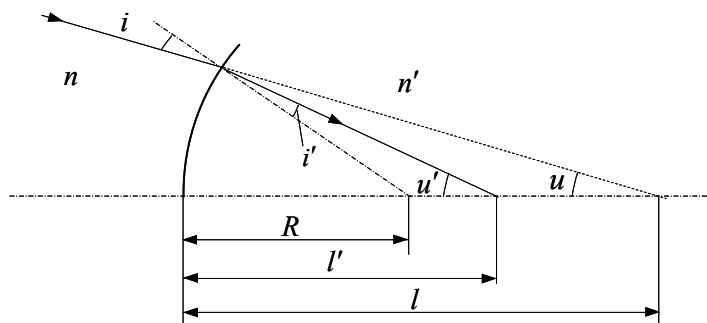


Figure 2.31 The aplanatic condition.

Generally, an aspherical surface is used to compensate for aberrations generated from other parts of the system, but sometimes it is used to avoid the occurrence of an aberration, for example, as a parabolic surface in a telescope.

A conic surface is free of spherical aberration for one particular set of conjugate points. For example, a parabolic mirror forms a perfect image of a point at infinity. When the object is axially moved away from the aberration-free conjugate point, some spherical aberration will be introduced.

Aspherical surfaces are very effective in controlling spherical aberration and distortion. To control spherical aberration that is not field dependent, the aspherical surface is generally placed close to the aperture stop. In the case of distortion, it is more effective to put the aspherical surface away from the stop because distortion is a field-dependent aberration.

Some basic guidelines in using aspherical surfaces are as follows:⁹

- Use a conic surface for correcting third-order spherical aberrations.
- To avoid artificially large aspherical coefficients, try not to use both conic k and r^4 terms simultaneously because the first term of the expansion of a conic k is r^4 .
- Use r^4 or higher-order terms rather than a conic surface on a flat or nearly flat surface.
- Use a conic surface along with higher-order terms on a curved surface.
- Optimize the optical system using spherical surfaces first, then use aspheric surfaces beginning with the conic and lower-order terms, and then gradually add higher-order terms if necessary.
- Avoid using a large number of aspherical surfaces, especially with higher-order terms, because they may beat against each other.

2.7.6 Optical materials

As Seidel equations show, all aberrations depend directly, more or less, on the refractive indices of the optical materials. The selection of optical materials has a significant impact on the aberrations. The lens surfaces generally have larger radii for the materials with higher refractive indices, therefore producing fewer aberrations. This is because the incident angles are smaller for the surfaces with larger radii.

Dispersion is critical in correcting chromatic aberrations. As discussed in Sec. 2.7.3, dispersions directly determine the powers of the individual elements in a cemented doublet and therefore the curvature of the cemented surface.

2.7.7 Stop shift

The stop shift determines the chief ray and the path of the off-axis beam through the lens, therefore affecting oblique aberrations, such as coma, astigmatism, distortion, and lateral chromatic aberration. The stop shift has no effect on spherical aberration, field curvature, or axial chromatic aberration.

As an example, Fig. 2.32 shows the impact of stop shift on coma. The off-axis spot diagrams [from left to right in Fig. 2.32(b)] change dramatically as the aperture stop moves away from the lens.

Note that changes in aberrations depend on the existing aberrations in some way. For example, if the system is free of spherical aberration, then a stop shift will not have an impact on coma. If the initial system is aberration free, then a stop shift will not introduce any aberration.

2.7.8 Symmetry principle

When a system is completely symmetrical, odd aberrations (namely, coma, distortion, and lateral chromatic aberration), which depend on an odd power of the field, are zero. The odd aberrations generated in the lens group in front of the stop have the same amount of aberrations in the rear group, but with the opposite sign; therefore, the front and rear group aberrations are canceled.

If the lens works with a magnification other than -1 , even if the configuration is exactly symmetrical, the odd terms will no longer be zero, but in most cases, they are small enough to be corrected by a slight change in the exactly symmetrical condition. In many cases, the lens configuration is quasi-symmetrical, not completely symmetrical but can nonetheless reasonably compensate for the aberrations.

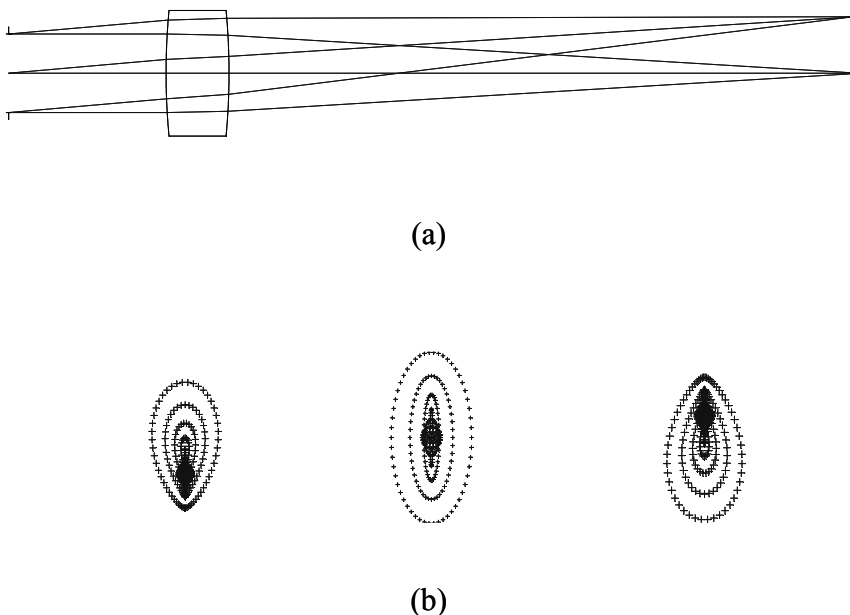
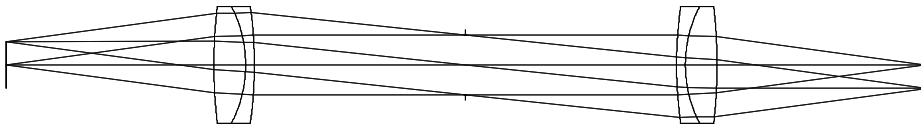
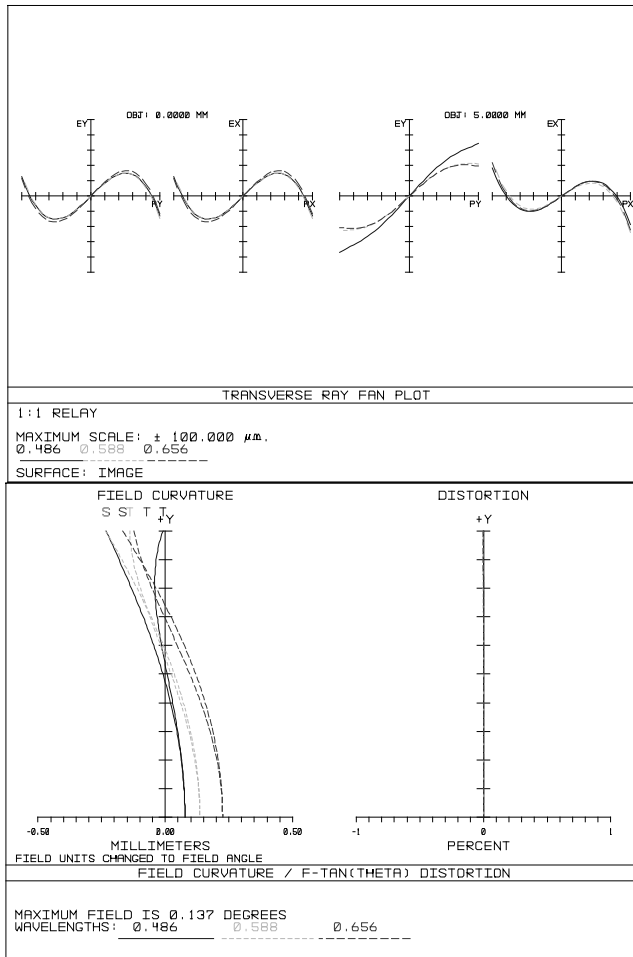


Figure 2.32 Effect of stop shift on coma: (a) lens configuration. (b) Off-axis spot diagrams when the aperture stop is 10, 15, and 20 mm from the lens.

Figure 2.33 shows a double telecentric relay lens consisting of two identical doublets. Because of its completely symmetrical layout, no coma, distortion, or lateral chromatic aberration can be found in the ray fan and distortion plots in Fig. 2.33(b).



(a)



(b)

Figure 2.33 (a) Double telecentric relay lens with unit magnification. (b) Plots of ray fan and distortion.

2.7.9 Field flatteners

Most optical systems have field curvature if no special measures are applied to control it. There are several approaches to correct field curvature.

One approach is to place a negative field lens close to the image plane or the intermediate image plane. This lens has a negative Petzval sum, and all other aberrations are small because the marginal ray height is small. The field lens is usually placed a small distance from the image surface; otherwise, dust particles and scratches on the lens surface will be in focus.

Another approach is to use a thick meniscus lens. If the lens has the same curvature on both surfaces, the Petzval sum is zero because the power of each surface is the negative of the other, but it has a positive power that is proportional to the thickness of the lens. Therefore, we can design a lens with zero power and a negative Petzval sum to reduce the inherent Petzval sum of the imaging system.

Finally, another method is to use separated lenses. When a positive lens and a negative lens with the same amount of power but opposite signs are separated by a distance, the Petzval sum is zero, but the power of the lens group is positive and is proportional to the distance between the two lenses. This principle is widely used to reduce Petzval curvature.

2.8 Optical Design

2.8.1 Determining the system specifications

The first step in lens design is to determine the system specifications, which are based on the application. The specifications include wavelength, resolution, focal length, NA or $f/\#$, FOV, image size, magnification, uniformity, performance, space constraints, image quality, and other requirements, such as environment. Generally, a first-order design can be developed to meet some of the specifications, such as the total track (the distance between the object and the image), magnification, and focal length.

2.8.2 Selecting a starting design

The next step is to select a starting design that has the potential to meet the above specifications. The proper system configuration of an optical system is the key to a successful design effort. The system configuration includes the lens format, number of elements, relative positions, and relative optical power distribution within the lens system. If no structural changes are made during the optimization process, the starting configuration determines the performance of the final system. Therefore, selecting a good starting configuration is very important, especially to a less-experienced designer. Three approaches are commonly used in selecting a starting design.

The first approach is the analytic method. For some simple imaging systems, it is possible to find a solution based on first-order properties and aberration theory. This approach is only possible for simple and known lens configurations, such as doublet, triplet, monocentric systems, and simple reflective systems. One

excellent example is the Schwarzschild microscope objective lens, which will be discussed in Chapter 4. Generally, it is very difficult to find an analytic solution for a new lens configuration.

The second approach is to start with a well-known lens format, such as Cooke triplet or double Gauss, and then add the necessary components step by step to enlarge the NA and FOV and to correct the aberrations. This approach requires a good understanding of aberration theory and design methods.

The third approach is to select an existing design that has similar specifications from a prior design, patent, or other published literature. The parameters used in selecting the design form include NA, FOV, spectral range, and performance requirements.

Figure 2.34 provides a useful guide for selecting an appropriate form on the basis of FOV and NA.⁸ It covers most, but not all, types of optical systems. The starting point should be a configuration that is inherently capable of meeting the design specifications.

After selection, the starting system must be scaled or modified to meet the first-order requirement. The system's potential for meeting the design requirements should also be quickly evaluated before proceeding with detailed optimization.

2.8.3 Optimization

2.8.3.1 Establishing variables, constraints, and merit functions

With the starting design in place, the next step is to set the variables, constraints, and merit functions for optimization. The variables include the following lens parameters: radius, lens thickness, air-gap distance, surface tilt and decenter, glass refractive index and Abbe number, and aspheric coefficients. The surface radii are the most effective parameters for correcting aberrations. The lens thicknesses and air-gap distances are less sensitive to aberration corrections; they are useful for fine correction and adjusting the design when fitting the test plate.

The constraints include items such as focal length, magnification, FOV, overall length or size of the system, working distance, ray angles, lens thickness, air-gap distance, and virtually any other system requirement.

The criteria discussed in Sec. 2.6 can be used to define the merit function for optimization. Typically, default merit functions, based on RMS spot radius, RMS wavefront error, or MTF in commercial optical design software, can meet the requirements of most applications. Other criteria related to system requirements can be added to the default merit function. A good procedure is to ensure the main targets in the beginning and then add additional targets during the optimization process.

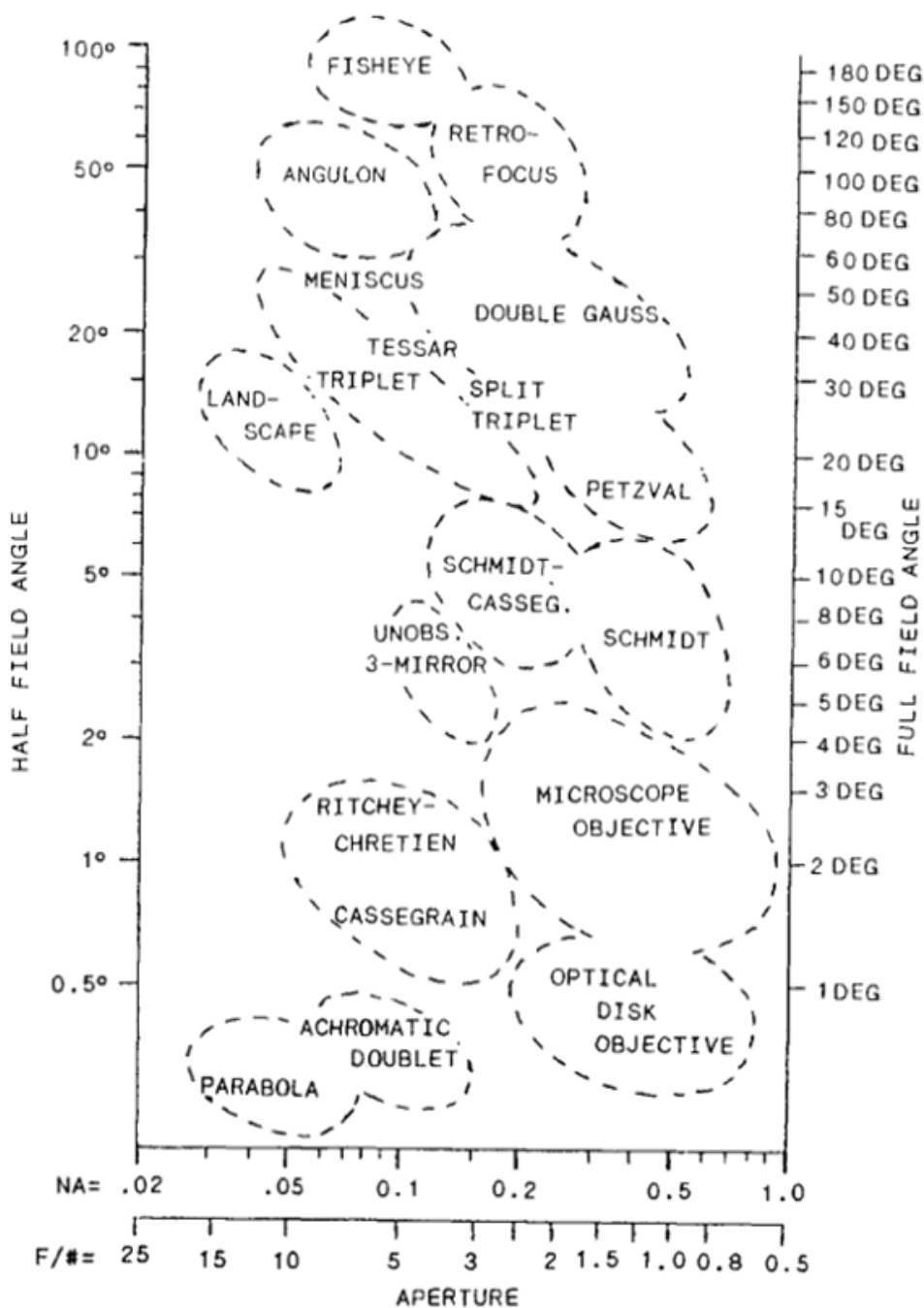


Figure 2.34 Map showing the design types that are commonly used for various combinations of NA and FOV. (Reprinted with permission from Ref. 8: © 2004 by McGraw-Hill.)

2.8.3.2 Optimization and evaluation

Optimization is not a magic black box that can provide a great system that meets the constraints and criteria in the merit function. Optimization changes only the variables involved in reaching the local or global minimum of the starting configuration. The outcome heavily depends on the starting configuration and the merit functions. Generally, more than one optimization cycle is required to reach a viable design. After each optimization cycle, it is necessary to evaluate the performance using the criteria specified for the optical system. Criteria include MTF, encircled energy, RMS spot radius, RMS wavefront error, Strehl ratio, and others.

The following is a list of standard techniques that are commonly used to improve the optimization for a fixed lens configuration:^{6,8,9,11}

- First, vary only the curvatures, and then add the thicknesses and air spaces.
- Add or reduce the variables.
- Change the weights of the fields, wavelengths, and targets.
- Avoid constraining more than two of (1) focal length, (2) object to image distance, and (3) magnification.
- Once identified, correct the troublesome aberrations as soon as possible to avoid propagation through the lens.
- Raise or lower the refractive index of the materials, maintaining element power.
- Introduce unusual partial dispersion materials if a secondary spectrum is a problem.
- Modify the merit functions to target the most troublesome aberrations by increasing weights or by adding operands.
- Significantly modify the constructional parameter and freeze it for several optimization cycles, then release it (for example, modifying a troublesome surface curvature to reduce the angle of incidence).

In most cases, simply optimizing the starting configuration cannot meet the requirements. It is necessary to make some structural changes that are based on the outcome of the optimization and to modify the merit functions. This process requires having solid knowledge of lens-design principles and experience with optical system design.

As a first step to making structural changes, it is necessary to identify the surfaces or elements with large third-order contribution or a large angle of incidence. Incidence angles are good indicators of the surfaces that need structural changes. The Seidel coefficients exactly represent primary or third-order aberrations and are also reliable indicators of a higher-order aberration. The following is a list of classic structural-change techniques:^{6,8,9,11}

- Add a lens, parallel plate, or meniscus shape in front or in back of the lens or in large air space.
- Remove any unnecessary surface or lens.
- Split an element into two (or more) elements.
- Replace a singlet with a cemented or air doublet.
- Split a cemented doublet.
- Reverse a cemented doublet or thick meniscus.
- Cement two adjacent surfaces that have nearly equal curvature.
- Insert a thick, low-power meniscus element before, after, or within the lens.
- Aspherize a surface (or surfaces).
- Add a field flattener or a concave surface close to the image.
- Shift the aperture stop.
- Add thickness to the positive-power meniscus to improve the Petzval field curvature.
- Add lenses to improve system symmetry.

Performance evaluation is an essential part of the optimization process; system performance must be evaluated after each optimization cycle.

In addition to evaluating image quality, another key step in lens design is tolerance analysis. Tolerance analysis is the error budgeting of the manufacturing and assembly of optical and optomechanical components. Sometimes, the system may need to be modified or redesigned completely in order to meet manufacturing tolerances and cost.

References

1. W. J. Smith, *Modern Optical Engineering*, 4th ed., McGraw-Hill, New York (2008).
2. R. Kingslake, *Optical System Design*, Academic Press, San Diego, CA (1983).
3. P. Mouroulis and J. Macdonald, *Geometrical Optics and Optical Design*, Oxford University Press, New York (1996).
4. H. Gross, *Handbook of Optical Systems*, vol. 1, *Fundamentals of Technical Optics*, Wiley-VCH, Berlin (2005).
5. W. Singer, M. Totzeck, and H. Gross, *Handbook of Optical Systems*, vol. 2, *Physical Image Formation*, Wiley-VCH, Berlin (2005).
6. H. Gross, H. Zügge, M. Peschka, and F. Blechinger, *Handbook of Optical Systems*, vol. 3, *Aberration Theory and Correction of Optical Systems*, Wiley-VCH, Berlin (2007).
7. H. Gross, F. Blechinger, and B. Achtner, *Handbook of Optical Systems*, vol. 4, *Survey of Optical Instruments*, Wiley-VCH, Berlin (2008).

8. W. J. Smith, *Modern Lens Design*, 2nd ed., McGraw-Hill, New York (2004).
9. R. E. Fischer, B. Tadic-Galeb, and P. R. Yoder, *Optical System Design*, 2nd ed., McGraw-Hill, New York (2008).
10. M. J. Kidger, *Fundamental Optical Design*, SPIE Press, Bellingham, WA (2001) [doi: 10.1117/3.397107].
11. M. J. Kidger, *Intermediate Optical Design*, SPIE Press, Bellingham, WA (2004) [doi: 10.1117/3.540692].
12. D. Malacara and Z. Malacara, *Handbook of Optical Design*, 2nd ed., Marcel Dekker, New York (2004).
13. W. T. Welford, *Aberrations of Optical Systems*, Adam Hilger, IOP Publishing Ltd., Bristol, UK (1986).
14. V. Mahajan, *Aberration Theory Made Simple*, SPIE Press, Bellingham, WA (1991) [doi: 10.1117/3.43000].
15. W. L. Wolfe, *Introduction to Radiometry*, SPIE Press, Bellingham, WA (1998).
16. R. W. Boyd, *Radiometry and the Detection of Optical Radiation*, John Wiley & Sons, Hoboken, NJ (1983).
17. J. E. Greivenkamp, *Field Guide to Geometrical Optics*, SPIE Press, Bellingham, WA (2004) [doi: 10.1117/3.547461].

Chapter 3

Optical Fibers in Biomedical Imaging

3.1 Introduction

Optical fibers have been used extensively in biomedical imaging for both illumination and imaging, mainly because they are thin and flexible. Optical fibers can deliver light from bulky light sources to remote locations, such as internal organs that are accessible noninvasively through natural openings or minimally invasively through tissues. In some imaging systems, the optical fibers also collect an optical signal from the measurement region and deliver it to the detector. Applications of optical fibers include light therapy, surgical instruments, diagnostic devices, endoscopy, optical spectroscopy, x-ray imaging, dental handheld devices, and many others.

Four different types of waveguides, namely, solid-core fibers, liquid light guides,¹ hollow waveguides,²⁻⁴ and photonic crystal fibers,⁵ have been developed and used for biomedical applications. Solid-core fibers and liquid light guides direct light by total internal reflection (TIR), whereas the light-guiding principle for hollow waveguides involves only reflection.

A hollow waveguide consists of a hollow tube coated internally with a thin metal layer and a dielectric layer. The light is guided by the reflection at the inner surfaces. Hollow waveguides are simple and low cost; they can transmit wavelengths over a large range and can transmit high laser power without any damage to the waveguide. The disadvantages of hollow waveguides include a large attenuation and lack of flexibility. Hollow waveguides are usually used in the mid-IR region for medical applications.

Liquid light guides are semiflexible hollow tubes filled with a liquid that is transparent in the wavelength range of interest. They are ideal for visible and UV light applications. Unlike optical fiber bundles, liquid light guides do not suffer from packing-fraction losses. Their advantage over solid-core fibers is that they are easier to manufacture. Liquid light guides have been used in some medical applications as laser- and light-beam delivery conduits.

Photonic crystal fibers (PCFs) are a new class of optical fibers based on the properties of photonic crystals. There are two types of photonic crystal fibers:

index-guiding fibers, which guide light through modified TIR, and photonic bandgap or air-guiding fibers, which guide light through the photonic bandgap effect.⁶ PCFs possess unique characteristics that ordinary optical fibers do not have. They are single mode from the UV to IR with large mode-field diameters, large NA, optimized dispersion properties, and air-core guidance.

There are two decisive advantages of optical fibers in biomedical imaging:

- (1) The large NA enables the delivery of high levels of illumination from bulky light sources to remote locations.
- (2) The flexibility of both illumination and imaging fiber bundles is ideal for procedures that are not possible with rigid endoscopes.

The field of fiber optics is very broad, from theory to manufacturing to applications. A comprehensive review is beyond the scope of this book. Instead, this book discusses the basic principles and properties of solid-core fibers, optics for fiber-optic imaging, and the application of fiber optics in biomedical imaging. The interested reader can find numerous books on optical fibers.⁵⁻¹⁰

3.2 Fundamentals of Optical Fibers

An optical fiber is a cylindrical glass or plastic waveguide that carries light along the fiber. The basic structure of an optical fiber consists of three parts: the core, the cladding, and the coating or buffer. The central core, which has a higher refractive index, is embedded into the outer cladding, which has a lower refractive index, as shown in Fig. 3.1(a). The traditional fiber, commonly called a step-index fiber, has a uniform refractive index throughout the core. When an incident ray strikes the core-cladding boundary at an angle larger than the critical angle θ_c , it is guided through the core without refraction loss, as shown in Fig. 3.1(b). Even though light can propagate along a fiber core and fibers can accept light with a larger incident angle without the cladding layer, the cladding is necessary to maintain the TIR condition along the entire fiber so as to reduce scattering loss at the cylindrical surface of the core, to protect the fiber from absorbing surface contaminants, and to add mechanical strength.

The critical angle θ_c at the core-cladding boundary is defined as the incident angle when the refractive angle in the cladding is 90 deg:

$$\sin \theta_c = \frac{n_2}{n_1}, \quad (3.1)$$

where n_1 and n_2 are the refractive indices of materials in the fiber core and the cladding, respectively. Any ray with an incident angle on the core-cladding boundary larger than the critical angle will be reflected without loss.

The maximum acceptance angle θ_a of the fiber is given by

$$n_0 \sin \theta_a = n_1 \sin \theta_1 = n_1 \cos \theta_c = n_1 \sqrt{1 - \left(\frac{n_2}{n_1}\right)^2} = \sqrt{n_1^2 - n_2^2}. \quad (3.2)$$

The NA of the fiber is defined as $n_0 \sin \theta_a$. For the fiber in air, $n_0 = 1$,

$$\text{NA} = \sqrt{n_1^2 - n_2^2}. \quad (3.3)$$

If the incident angle of the light ray is larger than the acceptance angle θ_a , the light will leak into the cladding and will not be guided through the fiber, as shown by the dashed line in Fig. 3.1(b). However, not all of the light with an incident angle less than the acceptable angle can be transmitted and can exit at the other end of the fiber. Light may be lost because of the reflection at the fiber surfaces and the absorption inside the core.

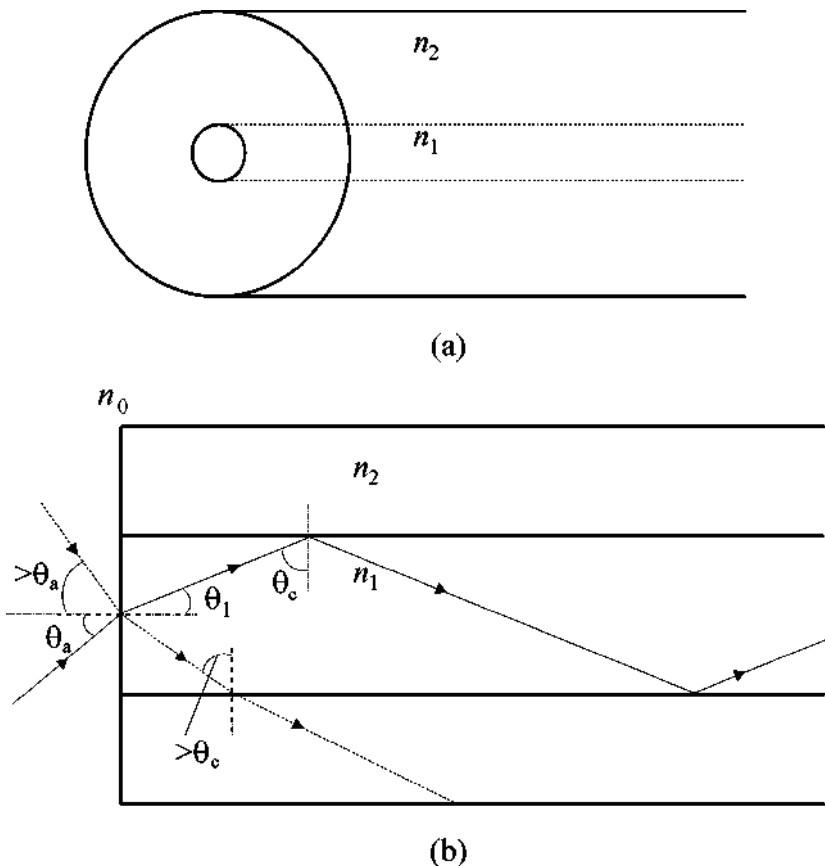


Figure 3.1 (a) An optical fiber is a cylindrical dielectric waveguide with a high-index core and low-index cladding. The buffer is not shown. (b) The trajectory of a meridional ray inside a fiber.

Two types of rays, meridional and skew, are typically used to represent light propagation along the optical fiber. As shown in Fig. 3.1(b), meridional rays pass through the optical axis of the fiber and are used to illustrate the basic transmission properties of optical fibers, such as the acceptance angle. Skew rays travel through the fiber without passing through the optical axis, as indicated in Fig. 3.2. The acceptance angle for the skew rays is larger than that of the meridional rays.⁹

Optical fibers are classified into two types by the number of modes that can be propagated along the fiber: single-mode fibers and multimode fibers. The number of modes a fiber can support is determined by its structural parameters, namely, the refractive indices of the fiber core and cladding and the radius of the fiber core. The normalized frequency, also called the V parameter, is related to the structural parameters of the fiber and is defined as

$$V = 2\pi \frac{a}{\lambda_0} \text{NA}, \quad (3.4)$$

where a is the radius of the core and λ_0 is the working wavelength in air.

When the fiber has $V < 2.405$, only the fundamental mode can exist; this fiber is called a single-mode fiber. A low V number makes a fiber sensitive to microbend loss and to absorption loss in the cladding. When $V \gg 1$, the number of mode M is approximated by

$$M \approx \frac{4}{\pi^2} V^2. \quad (3.5)$$

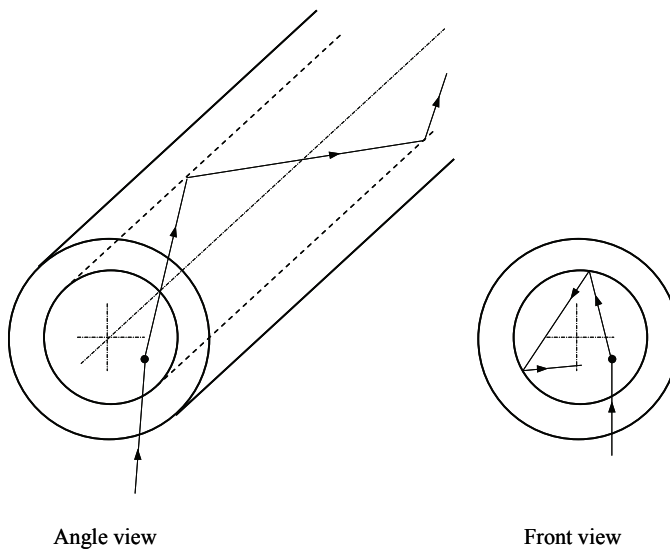


Figure 3.2 A skew ray propagating inside the fiber.

Single-mode fibers have small core diameters, typically less than 10 μm . They have lower signal loss and larger bandwidth than multimode fibers. When single-mode fibers are operated at wavelengths longer than the cutoff wavelength, light radiates into the cladding and is lost.

Single-mode fibers conserve the coherence of the propagating laser light; the output beam is a series of concentric, diverging wavefronts and can be recollimated and focused. In biomedical imaging, single-mode fibers are often used in systems based on the interferometric technique.

Multimode fibers usually have large core sizes and high NAs. The typical diameter of a fiber core is larger than 50 μm , and the NA is higher than 0.2. Light can be coupled into a multimode fiber more efficiently than into a single-mode fiber because of the larger core and higher NA. The light from a light-emitting-diode (LED) can be launched directly into a multimode fiber with reasonable coupling efficiency. Within a multimode fiber, each of the different modes still maintains coherence, but the overall output is the sum of all propagated modes. The light from the multimode fiber can neither be restored to a single concentric wavefront nor be focused to a diffraction-limited spot.

The differences among the group velocities when light propagates in a multimode fiber introduce a variety of propagation times; therefore, the light pulses are broadened as they propagate through the fiber. This effect is called modal dispersion and limits the speed at which the adjacent pulse can travel without overlapping. Modal dispersion is reduced in a graded-index fiber, whose fiber core refractive index varies from a maximum value at its center to a minimum value at the core-cladding boundary. In contrast, the fiber core in a step-index fiber has a constant refractive index. Light rays that travel farther from the fiber's axis travel a longer distance inside the multimode fiber. Since the refractive index decreases from the center of the core, the speed of light propagating in the graded-index fiber increases with the distance from the optical axis of the core. Therefore, the travel time of each mode is the same.

Most multimode graded-index fibers have a parabolic refractive-index profile. The NA decreases as a function of radial distance from the fiber center because of the decrease in refractive index. The axial numerical aperture (NA_0) is defined as

$$\text{NA}_0 = \sqrt{n_1^2 - n_2^2}, \quad (3.6)$$

where n_1 is the refractive index at the center of fiber core and n_2 is the refractive index of the fiber cladding.

Multimode graded-index fibers accept less light than multimode step-index fibers with the same core diameter, and they have a different acceptance angle across the fiber core because of the decreasing refractive index from the center to the edge of the fiber core. A multimode graded-index fiber with the same normalized frequency V as a multimode step-index fiber will have approximately one-half as many propagating modes as the step-index fiber.

Both glass and plastic materials have been used in fiber cores and cladding. Plastic fibers are usually more flexible than glass fibers and have a higher NA, a larger core diameter, and a lower cost. However, plastic fibers have higher light loss because of impurities and intrinsic absorption, particularly in the short-wavelength range. They may also show more autofluorescence and cannot withstand thermal stress as well as glass fibers.

3.3 Light Propagation in the Fiber

Light propagation along an optical fiber can be described by geometrical optics and wave optics. The geometrical optics approach, also called ray theory, is used to approximate the light acceptance and guiding properties of an optical fiber. This approach provides a clear picture of the propagation of the light in the fiber. However, this approach fails when the diameter of the fiber core is too small. The ray theory only describes the direction of a plane wave inside a fiber; it does not take into account interference between the plane waves. The wave optics approach, also called mode theory, uses electromagnetic wave behavior to describe the propagation of light inside a fiber. Mode theory will not be discussed in this book. Interested readers can refer to Refs. 7 and 10.

3.3.1 Single-mode fibers

Single-mode fibers have a very small diameter of only a few microns and a small NA. They can transmit only one mode of light and conserve its coherence. The output from a single-mode fiber is given by

$$E(r) = E_0 \exp \left[-2 \left(\frac{r}{w} \right)^2 \right], \quad (3.7)$$

where E_0 is the irradiance on the optical axis, r is the distance from the optical axis, and w is the $1/e^2$ radius of the beam profile. The output beam can be effectively collimated and focused into a small spot. Single-mode fibers are generally used in biomedical imaging systems that employ a laser as the light source.

3.3.2 Multimode fiber

Multimode fibers have a large core diameter and a large acceptance angle. When light is coupling into the multimode fiber, high-order modes will encounter more reflections and travel a longer distance per unit length. As a result of fiber defects such as impurities, sharp bends, and local distortion, the transmission losses of high-order modes are larger than those of low-order modes. Sharp bends and local distortion increase the angle of reflection, forcing the low-order modes to become higher-order modes inside the fiber. This technique is called mode mixing. Mode stripping is another technique to suppress particular modes in order to obtain a specific irradiation distribution.

When light propagates along the straight multimode fiber, the incident angle and exit angle are the same because the TIR inside the fiber does not change the ray angle relative to the optical axis of the fiber. In Fig. 3.3(a), two rays incident at different locations on the fiber core have the same incident angles; the exit rays propagate in different directions but have the same exit angle relative to the optical axis of the fiber. Therefore, for a straight fiber, the light emerging from the output surface has a cone or doughnut shape, depending on how the light is coupled into the fiber. However, when the fiber is curved, the emerging light usually has a different angle from the incident light. The angle can be smaller or larger. It is clear from Fig. 3.3(b) that the exit angles of the two rays (solid and dashed lines), which have the same incident angle at the entrance surface, have significantly different exit angles relative to the optical axis of the fiber. Figure 3.3(c) shows that rays incident at the same location with different angles reflect at various positions and distribute over the output surface. The straight fiber redistributes the light at the output surface but is not a good homogenizer because of its rotationally symmetric structure.¹¹ The light distribution on the output surface depends on the fiber length and the input ray angle. As long as the fiber is straight, the uniformity of the light distribution at the output surface oscillates as the fiber length varies. In summary, a curved fiber redistributes the light spatially and angularly; a straight fiber, on the other hand, can only redistribute light spatially.

Bends or local distortions alter ray angles, as shown in Fig. 3.3(b). Some rays may be forced into higher modes, with the possibility of leaking to the cladding, resulting in a transmission loss. Therefore, a minimum fiber bending radius is usually specified for each type of fiber.

Figure 3.4 is a simulation of the way in which a multimode fiber redistributes light spatially. The fiber used in the simulation has a core diameter of 0.05 mm. The input light has a Lambertian angular distribution over 10 deg and a uniform spatial distribution. A 0.02×0.02 mm aperture is attached to the input end of the fiber to limit the coupling area for simulation purposes only. Figures 3.4(a)–3.4(c) illustrate the irradiance distributions at the input surface, the output surface of a 500-mm straight fiber, and the output surface of a 250-mm fiber with an S-shaped curve, respectively. The white dashed circle on each plot indicates the fiber core-cladding boundary. Figure 3.4(d) plots the normalized irradiance profiles of the light distributions in Figs. 3.4(a)–3.4(c). The profiles have a smooth distribution with some statistical noise. It is clear that the straight fiber redistributes the input light but is not an effective light homogenizer. The curved fiber improves the uniformity of the irradiance at the output surface, but the falloff from the center is obvious. The uniformity of irradiance can be improved with more bends or twists.

When the fiber is long enough and has bends and twists, the spatial distribution of the light at the fiber surface of a multimode fiber is equal to the spatial distribution of the refractive index, which means it is uniform for the multimode step-index fiber, given that the refractive index at the core is a constant.¹²

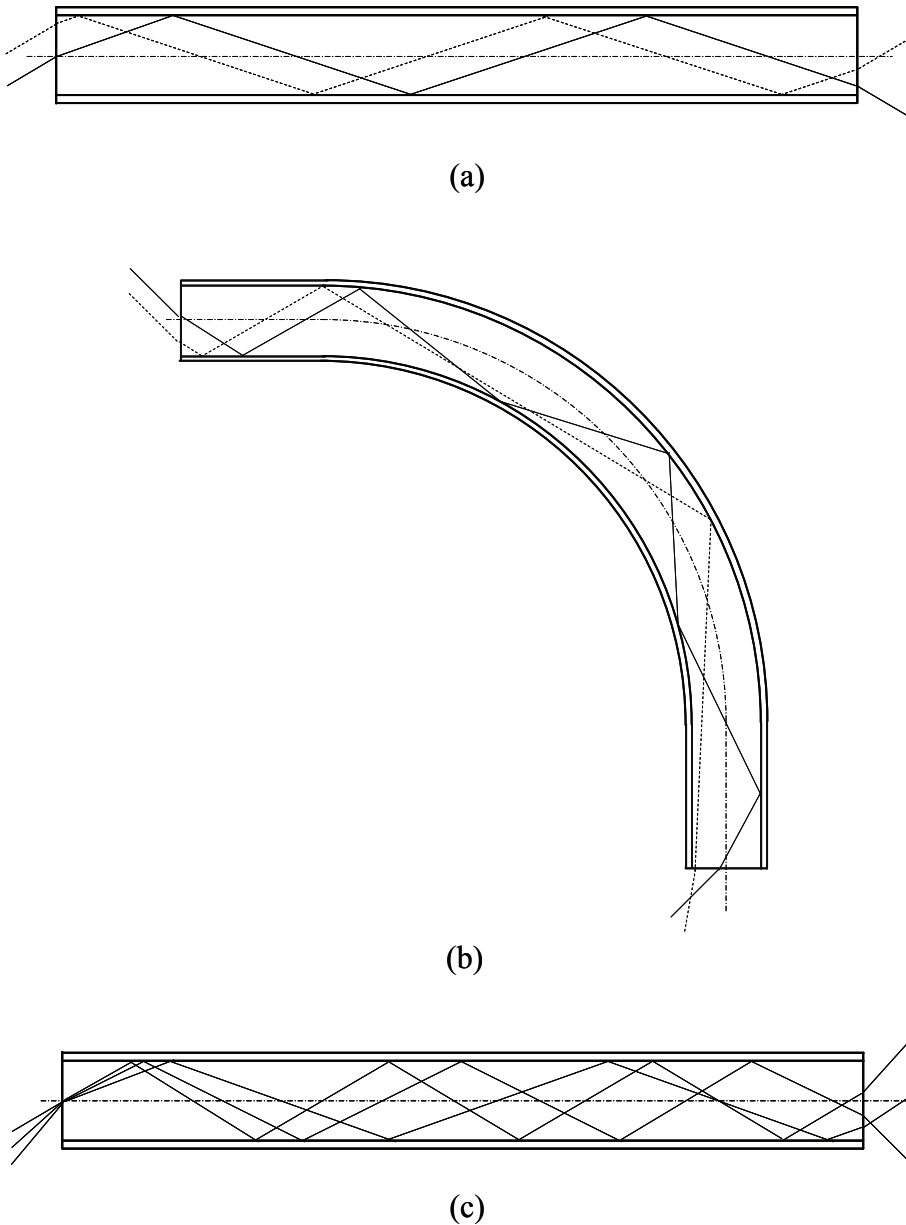


Figure 3.3 Light paths inside the multimode fiber. (a) A straight fiber changes the ray positions but not ray angles, (b) a curved fiber redistributes the light spatially and angularly, and (c) a straight fiber redistributes the light spatially but not angularly.

The angular distribution of the light from the multimode fiber depends on the modes, bends, and twists of the fiber. For a straight multimode fiber, the range of the exit angle is basically the same as the range of the incident angle at the input

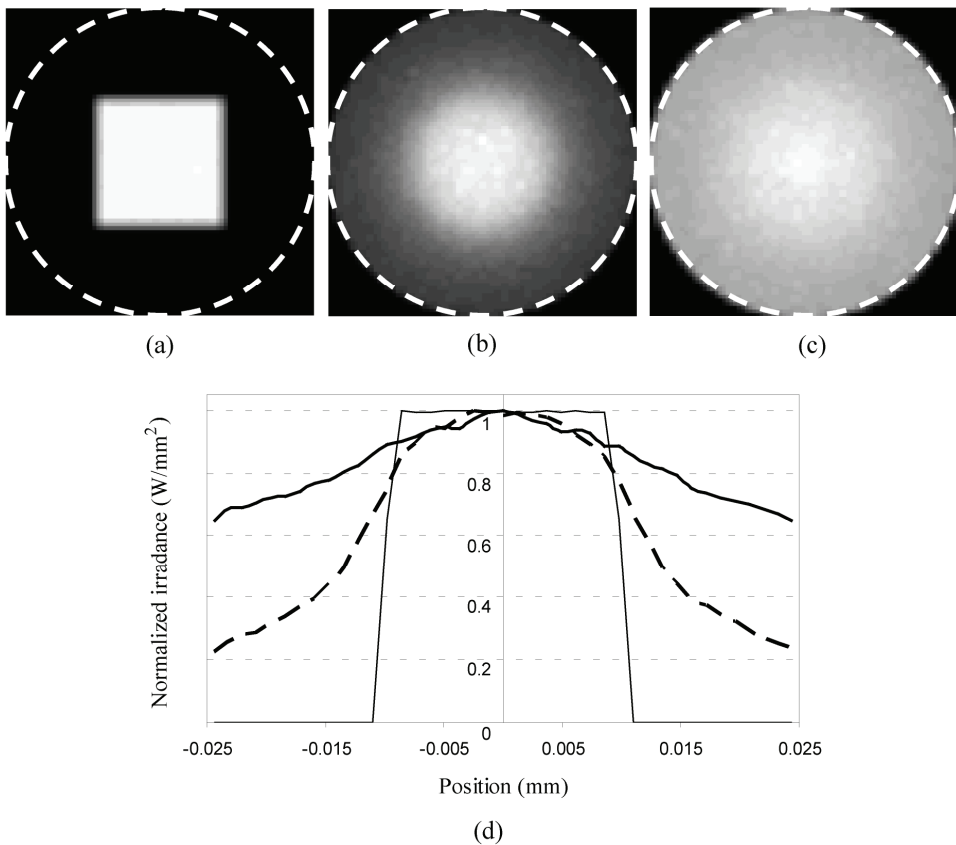


Figure 3.4 Simulation of light distribution at the output surface of a multimode fiber. (a) The irradiance at the input surface, (b) the irradiance at the output surface of a 500-mm straight fiber, (c) the irradiance at the output surface of a 250-mm fiber with an S-shaped curve, and (d) the line profiles of normalized irradiance at the input surface (thin solid line), at the output surface of a straight fiber (dashed line), and at the output surface of a curved fiber (thick solid line).

surface, and the intensity distribution is rotationally symmetric. As a result of the bends and local distortions, the exit angle of the light can be larger than that of the light coupled into the fiber but not larger than the acceptance angle of the fiber.

Figure 3.5(b) shows the normalized simulated intensity of the light emerging from a straight fiber and a fiber with two bends. The profiles have a smooth distribution with some statistical noise. To simplify the simulation and comparison, both the spatial and angular distributions of the input light are uniform, and both the cone angle and oblique incident angle are 10 deg. The thin solid line is the intensity at the input surface of a straight multimode fiber, the thick dashed line is the intensity at the output surface of a straight multimode fiber, and the thick solid line is the intensity at the output surface of a curved

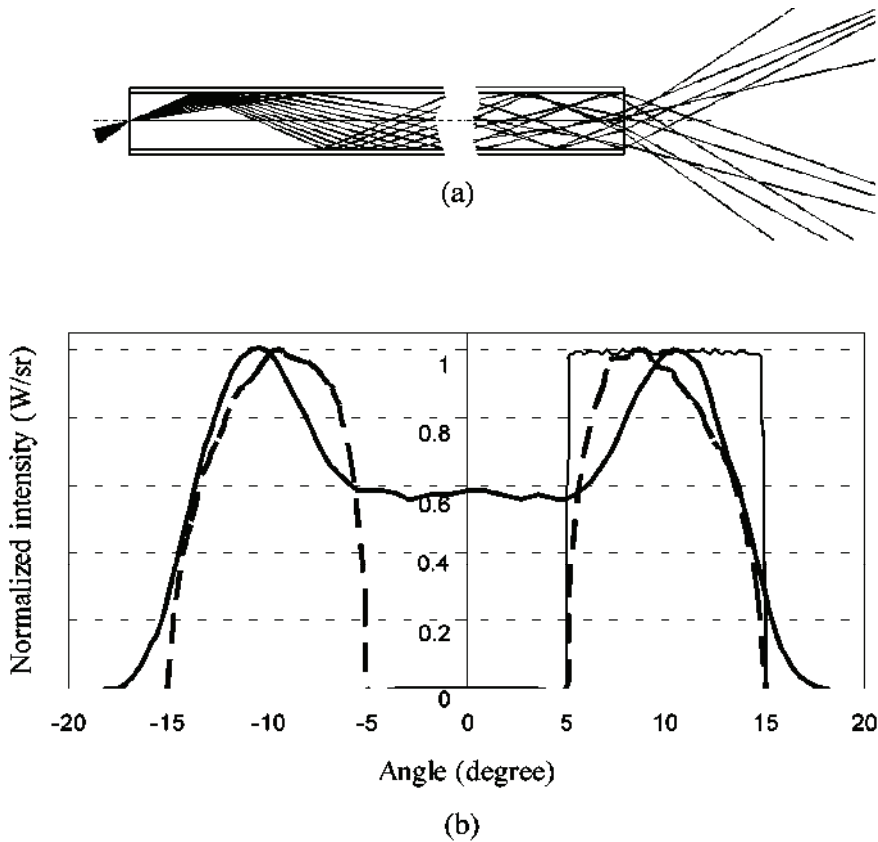


Figure 3.5 (a) A cone of light is coupled into the fiber at an oblique angle. (b) Normalized simulated intensities of the light at the input surface (thin solid line) and the output surface (thick solid line) of a straight multimode fiber and a curved multimode fiber (thick dashed line). Both the cone angle and oblique angle are 10 deg.

multimode fiber. Because of the rotational symmetric property of the fiber, the intensity profile of the output light has a doughnut shape. The straight fiber does not change the extent of the angular distribution, but the intensity profile is slightly different from the angular distribution at the input surface because of the skew rays. While spatially redistributing the light across the output surface of the fiber, the curved fiber also redistributes the light angularly, as shown in the normalized intensity of the light from the curved fiber in Fig. 3.5(b). The bent fiber converts a significant amount of the light to low-angle rays and a small portion of the light to high-angle rays. The characteristic of the angular distribution of the light from the bent fiber depends on the number of bends and the radius of each bend.

The spatial distributions of the light at the input and output surfaces as well as at a plane 2 mm from the output surface of a straight multimode fiber and a curved multimode fiber are plotted in Fig. 3.6. Because of the limited fiber length

and bends, the spatial distribution of the light at the output surface is far from uniform. The irradiance at the plane 2 mm from the output surface has a doughnut shape, corresponding to the doughnut-shaped intensity profile in Fig. 3.5(b). For a straight fiber, there is no light in the central portion, as shown in Fig. 3.6(c). The size of this central portion is determined by the angular distribution of the input light. The curved fiber distributes a certain amount of light to the central portion, as shown in Fig. 3.6(f). Adding more bends will distribute more light to the central portion. Both angular and spatial distribution of the light emerging from the output end of the fiber can be significantly improved by twisting the fiber.

Generally, the relative intensity of the light in a large-angle high-order mode is weaker than in lower-order modes. This difference exists because (1) the ray at the large angle is weaker in intensity when coupled into the fiber and (2) the optical path of the ray at a large angle is longer and the number of reflections on the core-cladding boundary is larger, producing higher absorption and reflection loss. Theoretically, the angular distribution of the emerging light from a multimode fiber can be approximated by a Gaussian function:¹²

$$I(\alpha) = I_0 \exp \left[-2 \left(\frac{\alpha}{a} \right)^2 \right], \quad (3.8)$$

where α is the angle formed with the optical axis, I_0 is the light intensity at the fiber surface when $\alpha = 0$, and a is the divergence angle of the emerging light.

When the step-index multimode fiber is long enough and has enough bends and twists, the irradiance distribution of the light is uniform at and near the fiber surface, but it quickly develops into a Gaussian distribution away from the fiber. Figure 3.7 plots the normalized irradiance at planes with different distances from the fiber surface. At the output surface, the irradiance is uniform (thin solid line), but it develops into an approximate Gaussian distribution within a few millimeters. The higher the NA, the faster the light deviates from uniform distribution. Therefore, when a multimode fiber with a large core is employed to provide uniform illumination, it must be very close to the illumination target, or else a relay system is needed to image the fiber surface to the illumination target.

3.3.3 Fiber tips

Fiber tips with geometries other than a flat surface will change the beam profile and light distribution. Figures 3.8(b)–3.8(d) show three modified fiber tips and the light distributions near them. For comparison, the standard fiber tip and light distribution are shown in Fig. 3.8(a).

In Fig. 3.8(b), a ball lens made directly from the fiber core is used to confine and focus light from the fiber core, increasing the light irradiance at a certain distance from the ball lens. Because of spherical aberration, the ball lens is more

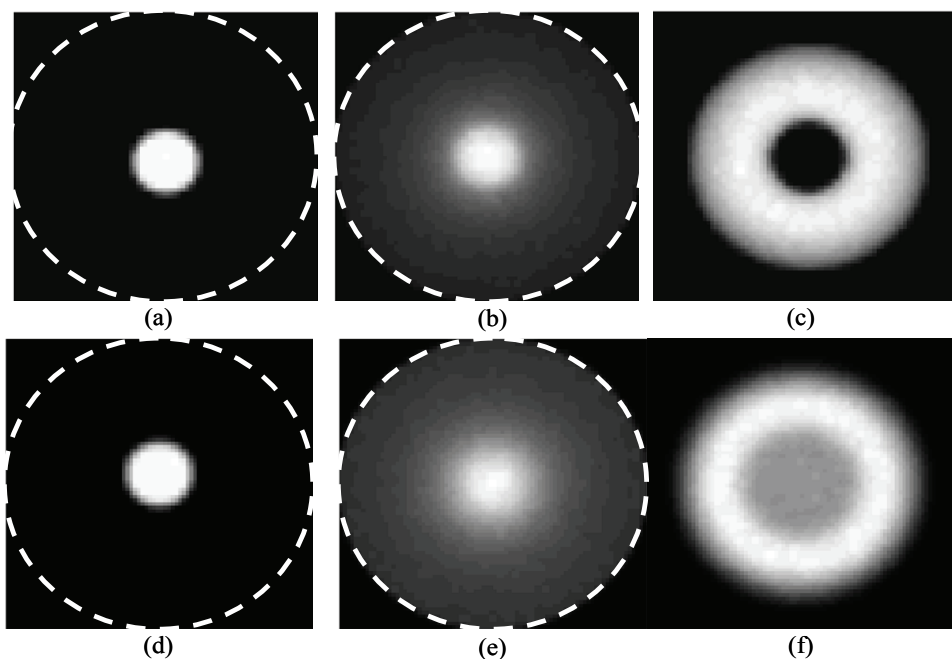


Figure 3.6 Simulated irradiance of the light at different planes for straight and curved multimode fibers with 50- μm core diameters when the light with a 10-deg cone is coupled into the fiber at an angle of 10 deg. Only the central part of the input surface is illuminated. (a), (b), and (c) show the irradiance at the input surface, output surface, and at a plane 2 mm from the output surface of a straight multimode fiber, respectively. (d), (e), and (f) show the irradiance at the corresponding planes of a curved multimode fiber. The dashed circle is the core-cladding boundary.

effective in focusing the less-divergent light into a small spot. Since the light is confined to a small region with high irradiance, and the light irradiance rapidly decreases away from the focal region, this type of fiber tip can be used in laser surgery. A lens more complex than a ball lens can also be used to meet some specific requirements, such as the size of the focused spot or the distance of the focused spot from the lens.

Figure 3.8(c) shows a fiber taper employed to increase the fluence rate at, or just in front of, the fiber tip. As the light propagates to the end of the fiber taper, the light irradiance increases, which is clear from the light distribution. The ray angle also increases gradually until it refracts out of the taper or is reflected back to the fiber and then to the light source. When the fiber taper makes contact with tissue, most of the light refracts out of the taper because the tissue–fiber surface alters the TIR condition.

As shown in the light distribution image of Fig. 3.8(c), the light emerging from the fiber taper has a large divergence angle. Therefore, the irradiance decreases rapidly from the fiber tip, minimizing damage to adjacent tissue when a fiber taper is used for tissue ablation.

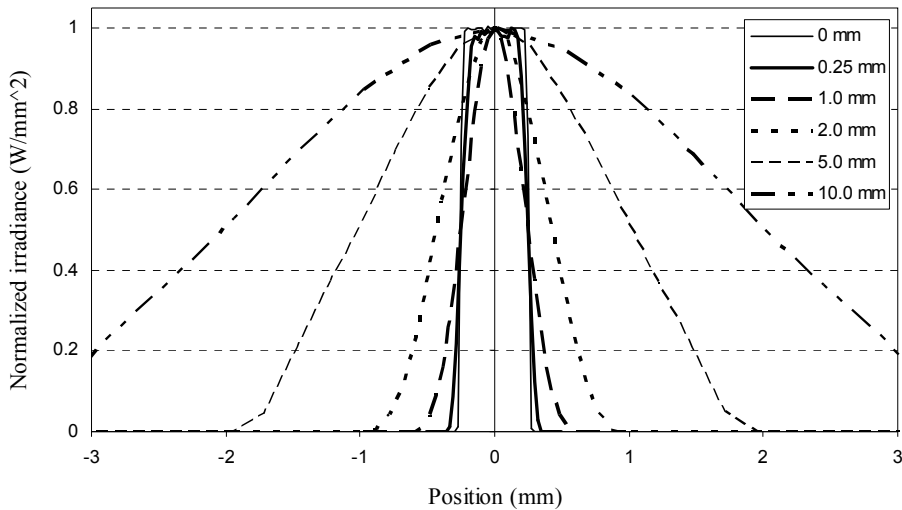


Figure 3.7 Normalized irradiance distribution of light from a multimode fiber as it propagates away from the fiber surface. The legend indicates the distance from the fiber surface.

The power density of the light at the taper depends on the diameter of the fiber core, the NA of the fiber, and the taper angle. When the taper angle becomes smaller, the power density is much higher at the tip. This type of fiber tip, made of silica or sapphire, has been used as a laser scalpel.

Figure 3.8(d) shows an angled fiber tip that is used for radial illumination to direct the light away from the optical axis of the fiber. The cylindrical wall of the fiber acts as a cylinder lens to focus the light in one direction, as shown in the irradiance image. Depending on the angle of the flat surface and the NA of the fiber, a certain percentage of the rays refracts out of the polished flat surface. Also shown in the irradiance image, some light exits the flat surface backward. Efficiency in the side illumination can be increased by applying a reflective coating on the flat surface.

3.3.4 Fiber bundles

A group of fibers can be combined with appropriate end terminations and protective sheathing to form fiber light guides and imaging fiber bundles. The fiber ends are either epoxied or hot fused together. Epoxied fiber bundles are temperature stable up to 120°C; hot-fused fiber bundles have a higher temperature resistance up to 350°C. The transmission of a hot-fused fiber bundle is higher because there is no light lost through the interstitial gaps between the fibers. Both ends of the fiber bundles are ground and polished perpendicular to the optical axis, thus enabling maximum transmission.

The fiber light guide, also called an incoherent fiber bundle, is used to deliver light from one location to another. The relative positions of the fibers at one end of the fiber bundle do not correlate with the positions at the other end. Randomization of the fibers is extremely important for fiber light guides; fibers from one end of the light guide should be thoroughly mixed in the other end to ensure a very homogeneous illumination pattern. Also, as discussed in Sec. 3.3.2, the straight multimode fiber does not change the ray angle as the ray propagates inside the fiber; the fibers in the fiber light guide are often twisted to make the angular distribution uniform at the output end. Depending on the applications, the input and output ends of a fiber light guide can be arranged in different shapes.

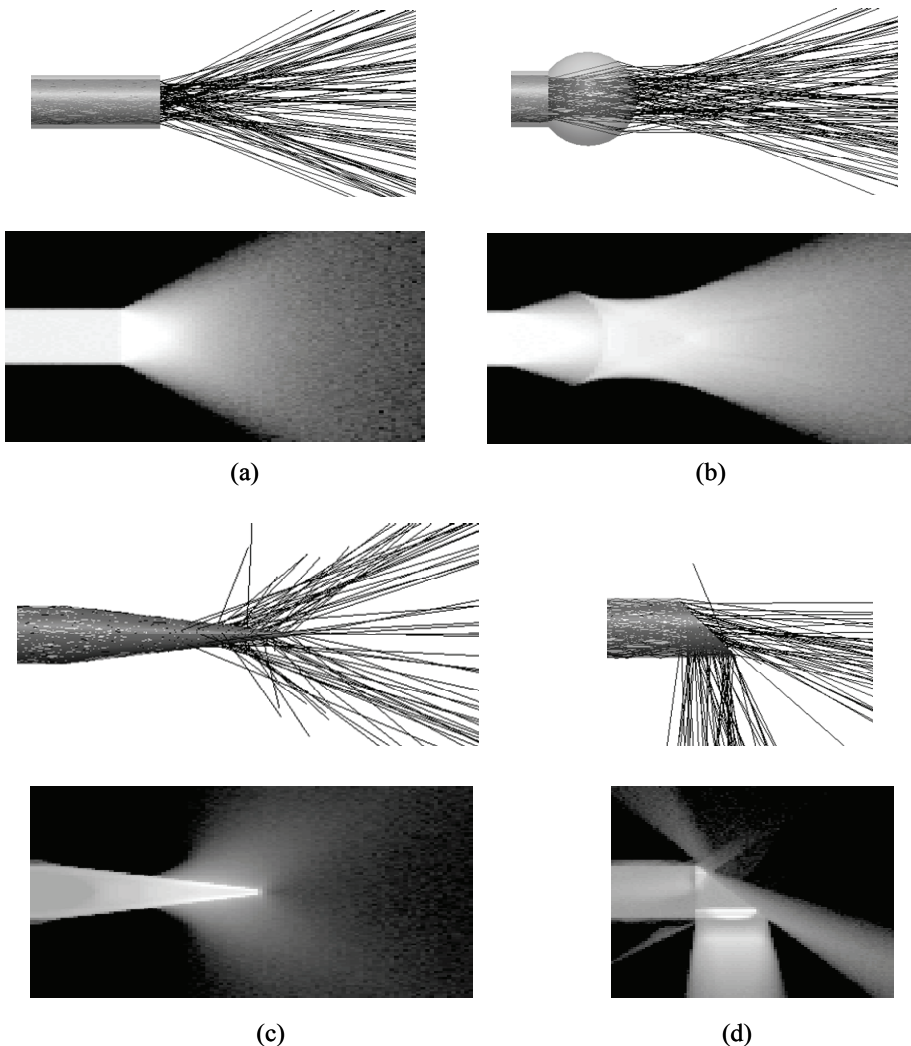


Figure 3.8 Fiber tips and the related light distributions. (a) Standard flat fiber tip, (b) fiber tip with ball lens, (c) fiber taper, and (d) fiber tip with obliquely polished surface. (There is no direct relationship to the brightness of the irradiance plots.)

In contrast, the ordering of the fibers in an imaging fiber bundle at one end must coincide with the ordering of the same fibers at the other end. This type of fiber bundle is sometimes termed a coherent fiber bundle. Coherent fiber bundles can be further classified into flexible imaging bundles and faceplates. All of the fibers in a flexible imaging bundle are epoxied or fused together through the entire length. Between the bundle's rigid ends, the fiber bundle is flexible. Faceplates are a shorter version of the imaging bundles and are usually much larger than flexible imaging bundles. Faceplates are either straight or tapered to magnify or demagnify the image. The surface of a faceplate can be polished to different shapes, such as convex, concave, and aspherical.

There are several packing configurations, such as square packing, where each fiber is surrounded by four fibers, and hexagonal packing, where each fiber is surrounded by six fibers, as shown in Fig. 3.9. The hexagonal packing in Fig. 3.9(b) is the most common as it has a higher packing fraction, which is defined as the ratio of the core area to the total area. The packing fraction of hexagonal packing is

$$\frac{\pi}{2\sqrt{3}} \left(\frac{r}{R} \right)^2, \quad (3.9)$$

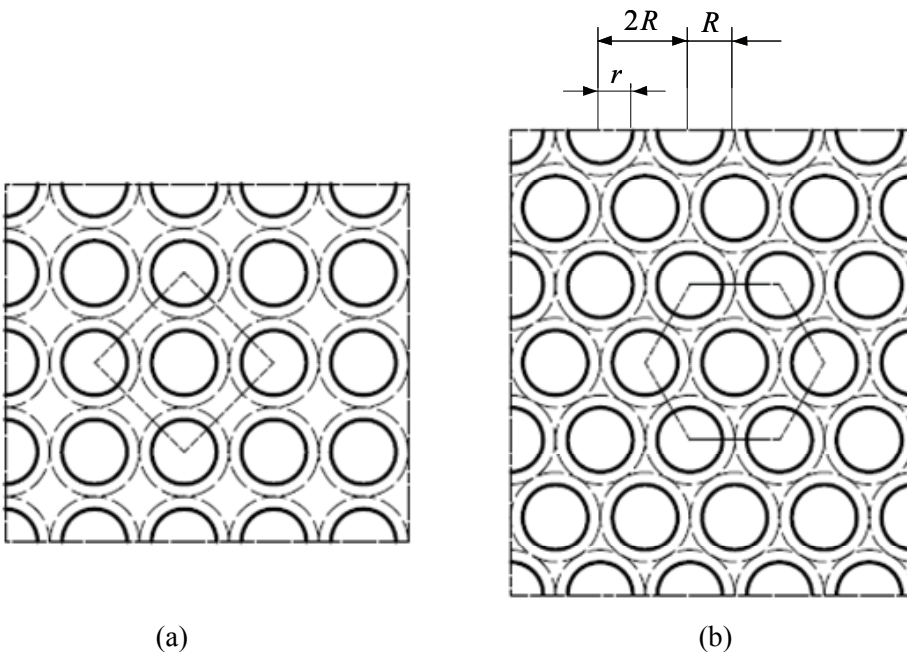


Figure 3.9 Packing configuration: (a) square packing and (b) hexagonal packing. The solid circle is the fiber core, and the dashed circle is the fiber cladding.

where r and R are the radii of the core and cladding, respectively. When a fiber core has a common cladding, R is the half distance between the centers of adjacent cores. Only light possessing an incident angle equal to or smaller than the acceptance angle of the fiber and impinging on the area of the core can be transmitted.

3.3.5 Fiber transmission

The efficiency of fiber in transmitting light is not only determined by the NA and core diameter but also limited by the following factors:

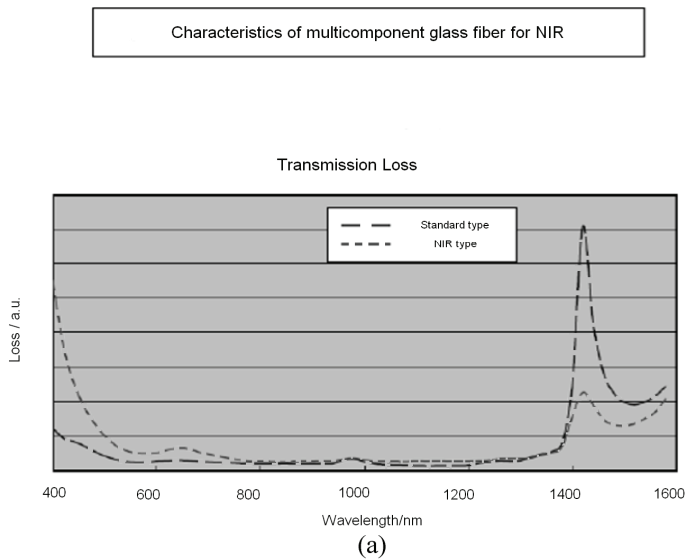
- The Fresnel loss on fiber surfaces. The loss from Fresnel reflection is approximately 4% at each end.
- The quality of the surface cleaving or polishing.
- The absorption of the light inside the core. It is mainly caused by unavoidable traces of colored metal oxides and is proportional to the total length of the ray in the fiber.
- The imperfect TIR on the core-cladding boundary. Even though the refraction and scattering loss on each TIR is very small, a measurable loss is expected over the fiber length because the number of reflections is large.

Typical spectral transmittances of commercially available fiber bundles are shown in Fig. 3.10. Both fiber light guides and imaging fiber bundles have a much higher loss in the short-wavelength range. They also have some absorption peaks in the near-IR region.

The higher absorption in the green-to-blue region is due to impurities such as Cr and Fe in the core glass, and other metals in the raw materials. Eliminating these impurities can achieve higher transmittance. However, pure materials could be too expensive for commercial fiber bundles.

For either illumination or imaging, the NA should be large in order to obtain bright images, which means that the refractive index of the core should be high. However, the higher the index, the higher the absorption loss is in the short wavelengths, resulting in yellowish images. This consideration is of particular importance in biomedical imaging, where detecting details often relies on perceiving differences in color.

The quality of the fiber bundle, such as the transmittance of light and the quality of the transmitted image, is affected by defects in the fiber bundle. Any defect larger than the resolution of the imaging system will appear in the final image. The most prominent defects include dark spots, dark lines, and disorder in the fiber alignment. These defects originate either from the manufacturing process or from repeated use. A dark spot is the most noticeable defect caused by a fracture or crack in one or more fibers in a bundle, where reduced or no light can be transmitted. A dark spot can also be caused by a solid particle inside the fiber bundle. A dark line usually appears when an array of fibers does not have



Transmission Characteristics of Fujikura's Image Fiber

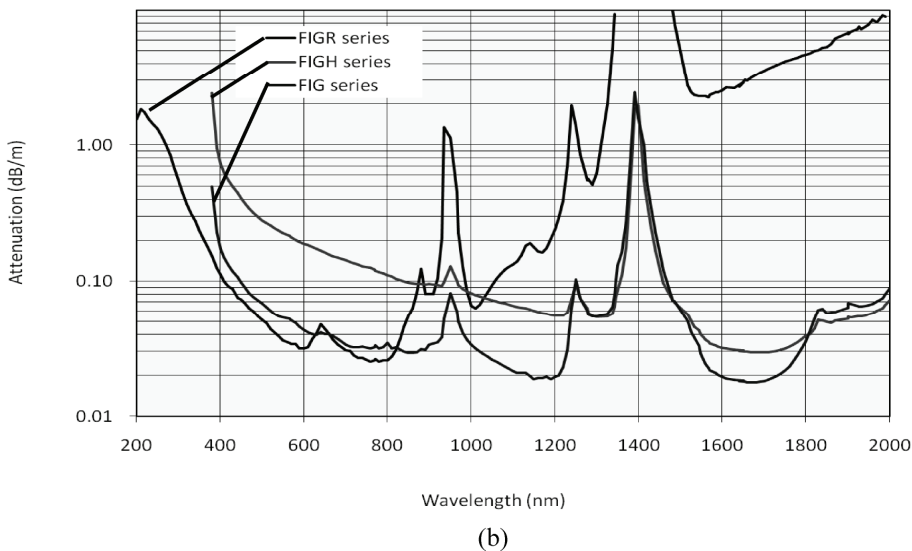


Figure 3.10 Typical transmission characteristics of fiber bundles: (a) fiber light guide with a fiber diameter of $50\ \mu\text{m}$ and an NA of 0.57, and (b) imaging fiber bundle. (Courtesy of Fujikura Ltd.)

close contact with an adjacent array. Also in the manufacturing process, some fibers may be displaced, which means that the order of the fibers at one end is not identical to the order at the other end. Fiber disorder is not critical for fiber light guides but should be avoided in coherent imaging fiber bundles because poor ordering will lower image quality.

3.3.6 Autofluorescence of fiber bundles

Because of impurities in fiber bundle materials and dopants used to increase the difference between the refractive indices of the fiber cores and cladding, fiber bundles may generate some autofluorescence, especially when illuminated with short wavelengths. In fluorescence imaging systems, autofluorescence from components other than the sample is background noise that reduces image contrast. Therefore, it is necessary to understand the fluorescence characteristics of fiber bundles.

Figure 3.11 shows the excitation-emission matrix (EEM) of three different fiber bundles from three different manufacturers. Data were collected from a 250- to 650-nm excitation light at 5-nm steps and with a 3-nm bandpass width, and the emission detection was started 5 nm above the center wavelength of the excitation band.¹³ Both the emission bandpass and spectral sampling were 3 nm. The fiber bundles from Sumitomo and Fujikura were Ge-doped fibers, while the fiber bundle from Schott was a quartz fiber. There were three autofluorescence peaks in the UV for the Ge-doped fibers [Figs. 3.11(a) and 3.11(b)], with maximum fluorescence excitations at 265 and 345 nm, and there is one autofluorescence peak for the quartz fiber [Fig. 3.11(c)], with a maximum excitation at 325 nm. The Fujikura fiber bundle had a fourth peak at a 380-nm excitation wavelength.

Therefore, it is necessary to measure the autofluorescence from the fiber bundle at the working wavelength and evaluate its impact on the fluorescence image before the fiber bundle is integrated into the system. The EEM measurements, as well as the background fluorescence images at working spectrum, provide useful information for optimizing the excitation/emission wavelengths, fluorophores, and the configuration of fluorescence filters for specific imaging applications. Filters and fluorophores can be chosen to minimize the impact of background fluorescence from a fiber bundle.

3.4 Illumination Using Fiber Bundles

Incoherent fiber bundles or fiber light guides are generally used for illumination in various medical instruments. There are several advantages to using a fiber bundle in an illumination system: 1) The fiber allows the source to be spatially separated from the illumination target. 2) By placing filters between the light source and the fiber surface, a fiber bundle prevents the unnecessary heat and unwanted content of radiation, such as UV, from reaching the target. 3) The flexibility of the fiber bundle makes it possible to deliver the light to places that would otherwise be impossible. 4) The nonuniform radiation from the light source is compensated for by randomizing the fibers in the bundle.

Typically, the individual fibers in a light guide have a core diameter of 15–50 μm and an outer diameter of 20–60 μm . The NA is usually larger than 0.5. Fibers made from glass are preferred to plastic when the light sources are high-power lasers or high-intensity arc lamps.

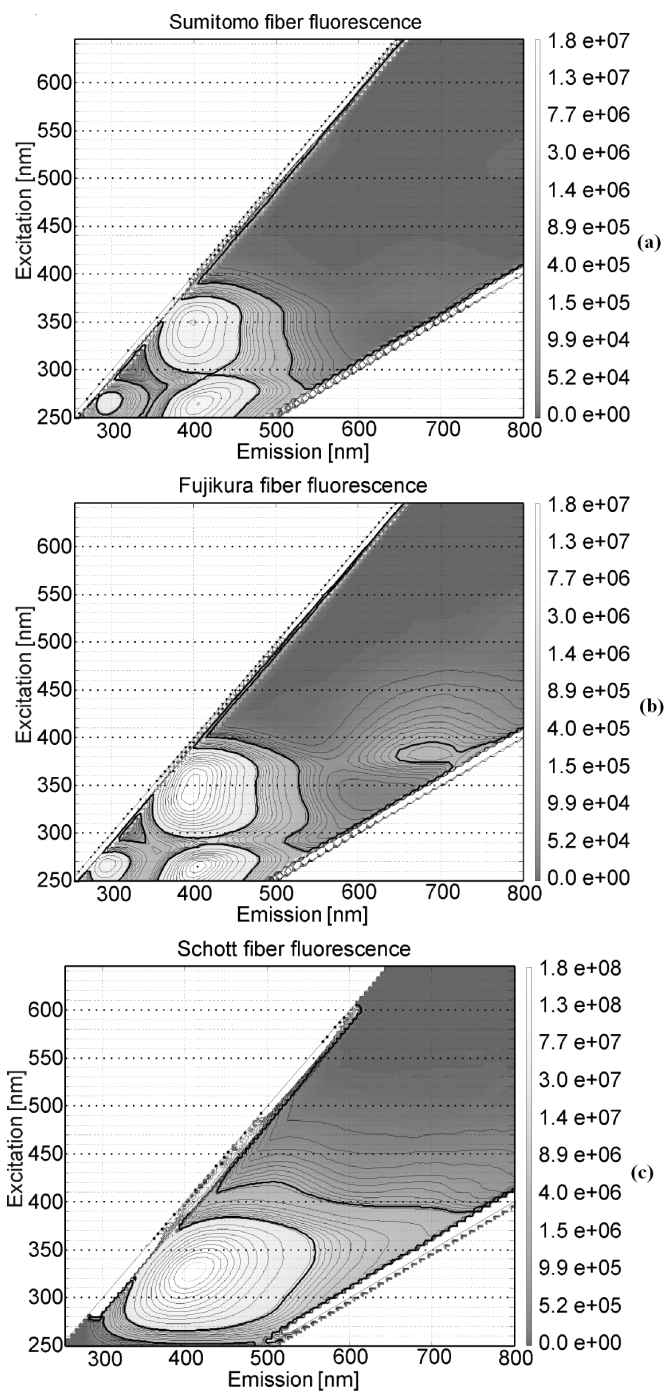


Figure 3.11 Excitation-emission matrices (EEMs) of imaging fiber bundles: (a) Sumitomo IGN-08/30, (b) Fujikura FIGH-30-850N, and (c) Schott fibers. The display scale for the Schott fiber has been increased by a factor of 10.¹³ (Reproduced with the permission from Optical Society of America.)

When light is focused on the input surface of a fiber bundle, the area occupied by the cores transmits light, whereas the area occupied by the cladding does not transmit light for a long distance. Therefore, the fiber bundle should have a large core diameter. The NA is another parameter that affects the efficiency in delivering the light from the light source to the illumination target. The collection efficiency is proportional to NA^2 . NA also determines the size of the illumination field if no additional optical components are used.

Figure 3.12 is a diagram of light being coupled from the light source to the fiber bundle. The light from the light source is usually focused on the input surface of the fiber bundle through the reflector and/or lens elements. Additional optical components, such as a hot mirror used to prevent IR light from reaching the input surface of the bundle, are sometimes placed between the light source and the fiber bundle.

Based on radiometry theory, the total flux that can be coupled into the fiber bundle is determined by the radiance of the light source, the area of the fiber cores, and the NA of the fiber bundle:¹⁴

$$\Phi = \pi AL(NA^2), \quad (3.10)$$

where Φ is the total flux, A is the total area of the fiber cores, L is the radiance, and NA is the numerical aperture of the fiber bundle.

Equation (3.10) determines the maximum useful light that can be taken from the light source. Once the light from the light source fills the input surface of the bundle and the cone of light is equal to the acceptance angle of the fiber, it is not possible to increase the amount of usable light from the same light source by using different coupling optics. Also, it is not always possible to increase the total usable light by using a light source with a higher power because its radiance can be smaller.

In the past, tungsten halogen and metal halide lamps were used in fiber-optic illumination systems for medical applications because these lamps provided reasonable illumination levels at a low cost. Halogen and metal halide lamps have some limitations, including short lifetime, low color temperature, low

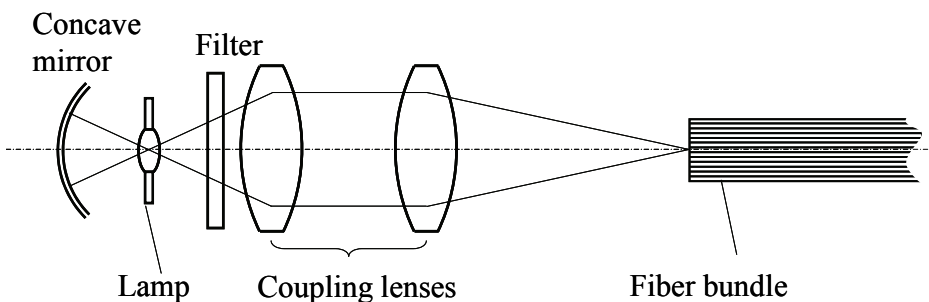


Figure 3.12 Diagram of light being coupled from a light source to a fiber bundle.

brightness, and low efficiency in coupling light into fiber bundles. In recent years, arc lamps, such as xenon and mercury lamps, have increasingly been used in applications in the medical field.

Figure 3.13 illustrates two configurations of fiber-optic illumination systems using arc lamps. A parabolic lamp is shown in Fig. 3.13(a), where the light from the arc is collimated by the parabolic reflector. Usually, a focus lens is used to focus the light onto the input surface of the fiber bundle and match the NA of the bundle. An aspherical lens is often used to achieve good coupling efficiency with the least complexity. A hot or cold mirror is usually used to block the infrared or UV light from the lamp. Figure 3.13(b) shows a configuration with an elliptical lamp. The light from the arc located at the first focal point is focused onto the second focal point, where the input surface of the fiber bundle is located. No additional optical elements are needed besides the hot or cold mirror. In both configurations, because of the cathode obstruction on the axis of the lamp, there is no ray on and near the optical axis.

One advantage of parabolic lamps is that the distance between the lamp and the fiber bundle can vary, as opposed to the fixed distance in elliptical lamps;

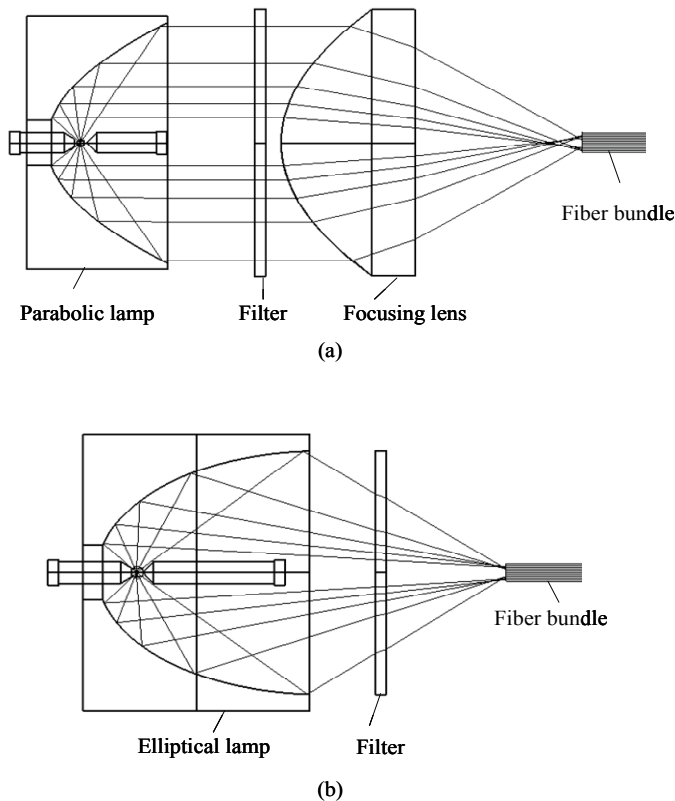


Figure 3.13 Basic configurations of light being coupled from (a) a parabolic lamp to a fiber bundle and (b) an elliptical lamp to a fiber bundle.

therefore, parabolic lamps can better accommodate additional elements, such as a hot mirror, in the optical path. Another advantage of parabolic lamps is that the ray angles are relatively small between the lamp and the focusing lens, which means that the hot or cold mirror can remove near-IR or UV light more effectively. In both configurations, the concave mirror can be a dichroic reflector, which is transparent to IR radiation but has high reflectivity for visible light.

Figure 3.14 shows the spatial and angular distribution of light at the focal planes of parabolic and elliptical lamps. Figures 3.14(a) and 3.14(c) are the irradiance plots for parabolic and elliptical lamps, respectively. Both have significant falloff from the center to the edge of the focal spot. Figures 3.14(b) and 3.14(d) are the raster charts of the intensity at the focal plane. Because there is no ray on and near the optical axis due to the cathode obstruction, there is a hole in the center of the intensity plot that corresponds to zero and the small ray angles. The size of the hole depends on the size of the cathode and the parameters of the reflective mirror.

The elliptical reflector can achieve a slightly higher number of lumens per watt than the parabolic reflector. This difference is mainly due to the lower light collection efficiency in parabolic lamps as compared to that of equivalent elliptical lamps.

Spatial and angular distributions are not uniform across the focal spot; therefore, a different amount of light with different angular distribution is coupled into each fiber in the fiber bundle. This nonuniformity on the input end of the bundle is partially corrected by randomizing the fibers in the fiber bundle.

In contrast to fiber bundles consisting of thousands of fibers, liquid light guides have only one component. Therefore, they are less sensitive to the spatial and angular nonuniformity at the input end and can provide more uniform light distribution at the output end. Liquid light guides are not regularly used in medical systems because of their rigidity and the difficulty in sterilizing them.

Even with some measures to prevent IR light from reaching the fiber bundle, the primary concern when using high-power arc lamps is the damage threshold of the fiber bundle because the center of the focused spot may have a high intensity peak. It is important to heat sink the fiber end to avoid thermal damage. To prevent the fiber bundle from overheating and to increase the coupling efficiency and uniformity, a light pipe can be placed at the focal point of the lamp to redistribute the light spatially and angularly, then the fiber bundle is attached to the other end of the light pipe, as shown in Fig. 3.15(a).

In this example, a square light pipe is placed at the focal point of the elliptical map, and a fiber bundle with a square input surface is attached to the other end of the light pipe. Figures 3.15(b) and 3.15(c) show the irradiance of the light on the focal plane of the elliptical plane and on the output surface of the light pipe, respectively. The square light pipe redistributes the light uniformly on its output surface. However, it does not change the intensity distribution, which means that there is still one hole in the intensity plot. In order to change the angular distribution to meet the NA of the fiber bundle, a tapered light pipe is often used.

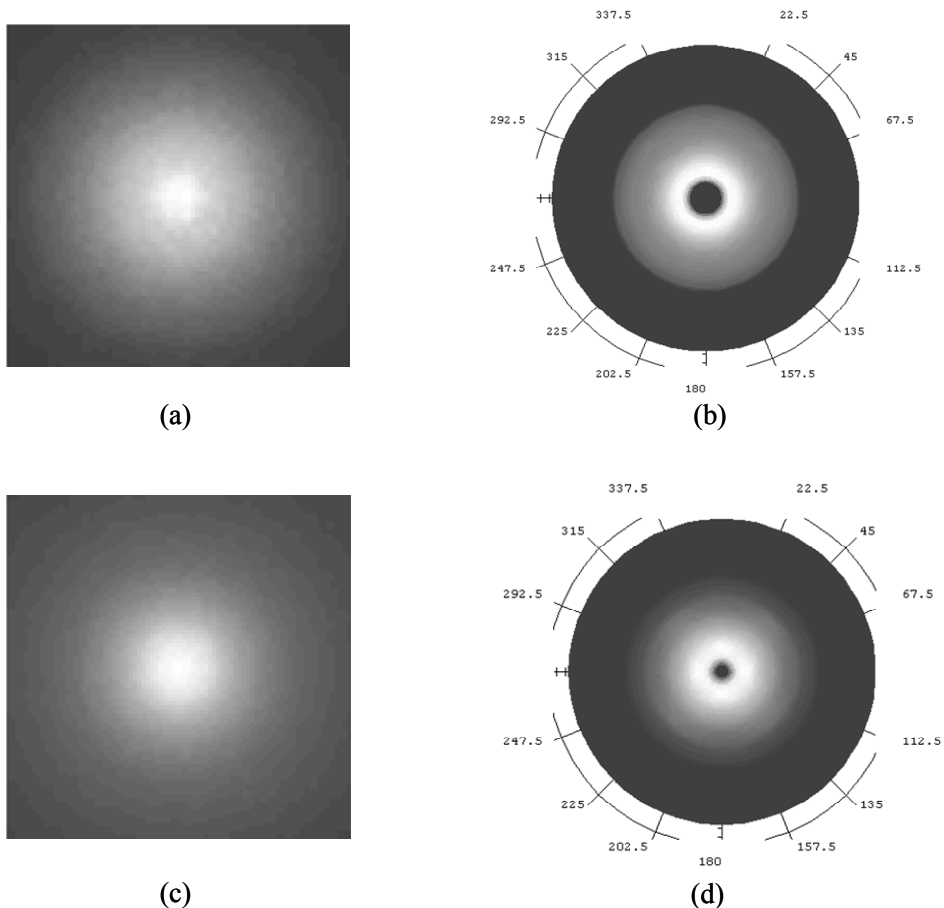


Figure 3.14 Spatial and angular distribution of light at a focal plane of a lamp. (a) and (c) are plots of irradiance at the focal plane of the parabolic and elliptical lamps, respectively, and (b) and (d) are the plots of intensity at the focal plane of the parabolic and elliptical lamps, respectively.

Given that a simple cylindrical light pipe is not an adequate light pipe homogenizer when the fiber bundle is circular (which is generally the case), a cylindrical light pipe with specific surface features can be used to make the spatial distribution of light from the arc lamp uniform.¹¹ Another solution is to use a special light pipe to convert the beam shape from square to circular after the square light pipe homogenizes the light distribution.

As a result of the recent advances in brightness and available color gamut, LEDs are emerging as important illumination light sources in medical imaging. LEDs have the benefit of significantly longer lifetime, low power consumption, small geometric dimension, and the availability of a wide range of discrete wavelengths. When direct LED illumination cannot be accommodated, LED fiber illumination is an option.

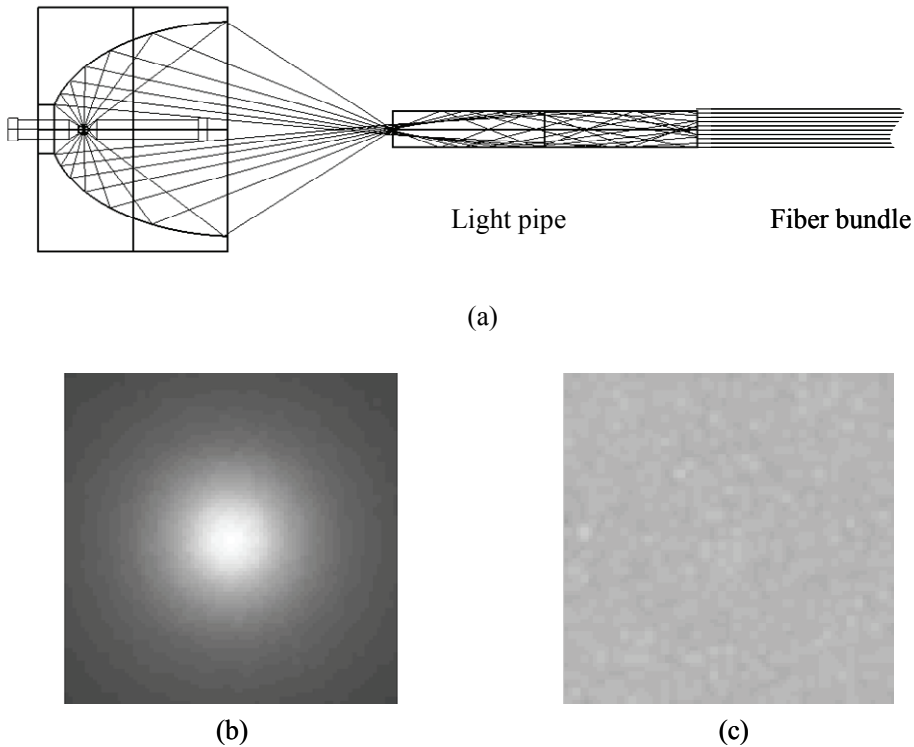


Figure 3.15 (a) Configuration of light being coupled from an arc lamp to a fiber bundle through a light pipe, (b) irradiance at the input end of the light pipe, and (c) irradiance at the output surface of the light pipe.

Several configurations can be used for coupling the light from LEDs to a large-core fiber or fiber bundle. The first and simplest method is through direct contact, as shown in Fig. 3.16(a). The disadvantage of this method is a low coupling efficiency because LED chips usually have a Lambertian angular distribution over a large angle. Any light with an angle larger than the NA of the fiber is wasted. The second method, shown in Fig. 3.16(b), is to image the LED emitting surface to the input surface of the fiber bundle. This configuration has the advantage that lenses can be optimized to maximize the coupling efficiency. Another advantage is that the light (from LEDs) with different spectra can be combined by a dichroic mirror and delivered to the input surface of the same fiber bundle. The third configuration is to use nonimaging optical elements, such as a compound parabolic concentrator (CPC), shown in Fig. 3.16(c). The LED is placed on the small end of the CPC, and the fiber bundle is on the other end of the CPC. The CPC reduces the ray angles so that the large-angle rays from the LED can be coupled into the fiber. None of these configurations provides uniform light distribution on the input surface of the fiber bundle; a light pipe can be placed between the LED and the fiber bundle to improve uniformity.

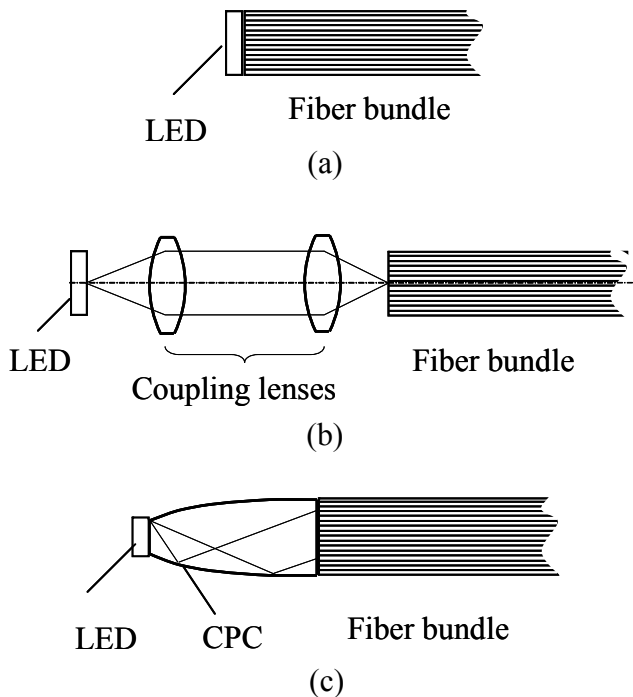


Figure 3.16 Configuration of light being coupled from an LED to a fiber bundle. (a) The fiber bundle is in direct contact with the LED, (b) the emitting surface of the LED is imaged onto the input surface of the fiber bundle, and (c) a nonimaging optical element (e.g., a CPC) reduces the angles of the rays from the LED and couples the light into the fiber.

Inside the fiber, the light is attenuated by bulk absorption and scattering losses that are wavelength dependent. Generally, the losses are higher for shorter wavelengths. When the fiber bundle is used for white-light illumination, it may render an erroneous appearance of the viewed object because the absorption of the light is higher in the blue end of the spectrum. When UV illumination is required, for example, to excite fluorescence, a bundle of fused silica fibers is generally used in place of ordinary glass and plastic fibers because of silica's better transmission properties.

The irradiance distribution on the sample can be accurately predicted by using the angular distribution of the light from the fiber bundle and the inverse-square law. The fibers in the bundles can be arranged in a ring or other forms to achieve a desired irradiance distribution. The output or input end of the optical fiber bundle can also be split into multibranch fiber bundles, such as bifurcated or trifurcated bundles. In multibranch fiber bundles, the bundles should be combed randomly to produce uniform outputs.

3.5 Fiber-Optic Imaging Systems

3.5.1 Introduction

Both single fibers and imaging fiber bundles have been used extensively in biomedical imaging systems. The applications and system configurations are quite different for systems using single fibers and fiber bundles.

Because of their small NA, small core diameter, and the ability to conserve coherence, single-mode fibers are typically used in imaging systems that have lasers as their light sources. The light from a single-mode fiber can be focused to a diffraction-limited spot, allowing high-resolution imaging. The same fiber core that delivers the illumination light can also act as a detection pinhole, enabling confocal detection. Single-mode fibers are particularly suitable for laser scanning imaging systems, such as confocal imaging, laser scanning endoscopes, and optical coherence tomography. In order to obtain 2D or 3D images, a miniaturized scanning mechanism is required inside the imaging head. The designs and configurations of these scanning methods will be discussed in Chapters 7 and 8.

To couple the laser beam into a single-mode fiber, the focusing lens should have a diffraction-limited performance in order to maximize the coupling efficiency. For a single-wavelength imaging system, a single aspherical lens is sufficient to couple the light into the fiber or focus the light from the fiber onto a diffraction-limited spot. When multiple wavelengths are used, the lens becomes more complicated in order to minimize chromatic aberrations.

Single multimode fibers are sometimes used in biomedical imaging systems. Because of their larger core diameter and higher NA, multimode fibers are more suitable for fluorescence collection. Single multimode fibers can be used to directly collect the light without the use of additional optical elements. The disadvantage of multimode fibers in imaging systems is that the light from the fiber cannot be focused to a diffraction-limited spot, limiting the achievable resolution.

An imaging fiber bundle usually contains thousands of high-refractive-index cores in a common lower-refractive-index cladding and is typically used as an image relay component in biomedical imaging systems. Fiber bundles have been used in endoscopy, fluorescence imaging, confocal scanning, two-photon imaging, and other biomedical imaging systems.

3.5.2 Optical properties of imaging fiber bundles

An imaging fiber bundle typically contains up to $\sim 100,000$ individual optical fibers packed into a hexagonal array. The spatial arrangements of the fibers at both ends are identical, resulting in a spatially coherent transmission of the image. The optical fibers are multimode, preserving the color of the light. The overall diameter of the imaging bundle may vary between 0.16 and 2 mm and may be over 2 m in length. The core diameters and separations are on the order of a few microns. The refractive index contrast in an imaging fiber bundle is

increased from that of a standard, single-mode fiber in order to more tightly confine the light and reduce crosstalk between the cores. Most imaging fiber bundles are made from quartz or doped silica. One of the two most common dopants, germanium (Ge), increases the refractive index of the fiber cores relative to the cladding. Another dopant, fluorine (F), is usually added to reduce the refractive index of the cladding.

The key advantages of using imaging fiber bundles are that (1) no scanning component is required for nonconfocal imaging modalities and (2) the scanning components can reside on the proximal end of the bundle for scanning imaging systems. Designs and configurations of scanning methods will be discussed in Chapters 7 and 8.

There are two basic requirements for an imaging fiber bundle:

- (1) High efficiency: The individual fiber must have an NA large enough to collect sufficient light at both ends. Also, the fill factor should be large to maximize the amount of light that can be coupled into and transmitted by the fiber bundle. This means that the core diameter must be large and the cladding layer should be thin. However, there is a limit to the thinness of the cladding. If the cladding layer is too thin, the light transmitted within one fiber may leak into its neighboring fibers, resulting in crosstalk. On the other hand, the core diameter should not be too large in order to maintain a reasonable flexibility.
- (2) High resolution: The resolution of imaging fiber bundles is determined by the core diameter and cladding thickness. High resolution requires a small core diameter and a thin cladding layer.

The above requirements are contradictory and cannot be satisfied simultaneously. The core diameter should be small to obtain a high-resolution image, but it should be large enough to achieve reasonable light-collection efficiency.

Ideally, the packing fraction should be as large as possible to transmit more light, and the fiber core should be as small as possible to achieve a high resolution. However, given that there is a limit to the practical thickness of the cladding layer as the diameter of the core is reduced to improve resolution, the packing fraction becomes unacceptably low, resulting in a dark image and a low signal-to-noise ratio (SNR).

The coupling between the cores is one of the limiting factors of using fiber bundles. The degree to which two cores will interact, or couple, depends on the wavelength, the cladding thickness, and the core size. Because of the small core size and separation of imaging fibers, light will begin to disperse among the cores after a propagation distance of much less than a meter.¹⁵ Crosstalk between fibers can blur the images transmitted through fiber bundles. However, nonuniformity in the core size, as well as variations in shape and separation, helps to reduce the efficiency of the coupling between adjacent cores, substantially suppressing the crosstalk. Therefore, the manufacture of functional image fibers with closer and smaller pixels requires the pixels to be nonuniform in order to enhance the mismatch between neighboring cores.

The pixelated nature of the fiber bundle is another limiting factor in imaging systems. The image from a fiber bundle is of a discrete nature. Pixelation reduces the lateral optical resolution in the object plane to about twice the average core-to-core distance divided by the optical magnification of the imaging lens. For the hexagonal configuration of the fiber bundle in Fig. 3.17(a), in order for the light beam to always be detectable regardless of position, the diameter of the image point D_s should be larger than $2(D/\sqrt{3} - r)$, where D is the distance between the centers of the adjacent cores and r is the radius of the core. To be distinguishable, the two spots should be separated by at least $2(D/2 + r + D_s/2)$. Also, as shown in Fig. 3.17(a), the line width T should be larger than $(\sqrt{3}/2)D - 2r$ in order for the line to be detected, regardless of its position and orientation.

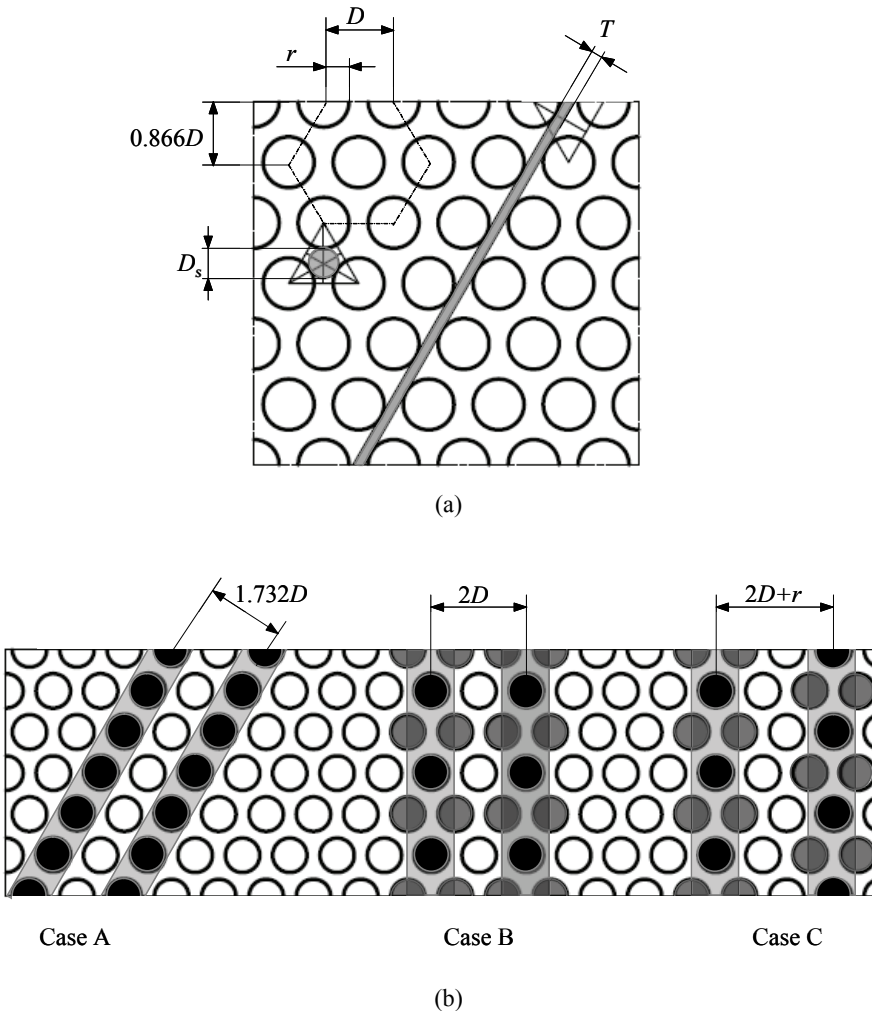


Figure 3.17 (a) Minimal resolvable point and line of a fiber bundle and (b) line objects with various azimuthal orientations on the fiber bundle surface.

As shown in Fig. 3.17(a), fiber bundles packed in hexagonal configurations have two effective spacings between the fibers in the orthogonal directions, D and $0.866D$, which means that the resolution changes from maximum to minimum for every 30-deg rotation of the fiber bundle.

The resolution of the imaging system depends on the orientation of the object. As shown in Fig. 3.17(b), the best attainable resolution using hexagonal fiber bundles is $1/(1.732D)$ in line pairs per millimeter when the two lines align perfectly with the fibers (case A). Generally, the resolution is $\leq 1/(2D + r)$ (case C), where there is always one fiber without light between two lines. In case B, the two lines are barely resolvable because not all of the fibers between the two bars are light free. Therefore, the realistic resolution is

$$d = \frac{1}{2D + r} . \quad (3.11)$$

The resolution can be increased through dynamic imaging in which the fiber bundle is vibrated randomly in the transverse direction by keeping both ends stationary. The resolution of a dynamic image, in general, is higher than that of an equivalent static image obtained without vibration.¹⁶

The image quality is also determined by the total number of fibers in the image area. The image quality is further affected by the following factors:

- A limited depth of focus/field compared to a lens relay imaging system,
- Imperfections in the ordering of fibers,
- Faulty or broken fibers,
- Lower transmittance of light in the blue end of the visible spectrum,
- Leaky modes excited in the fibers, and
- Crosstalk between the fibers.

Other limitations include the Moiré pattern produced by a superposition of the image from the fiber bundle and the digital imaging sensor, such as CCD and CMOS.

3.5.3 Depth of focus/field

Depth of focus is an important consideration in imaging systems using fiber bundles. When a lens relay system is used, the out-of-focus object point is transferred to the object space of the eyepiece or imaging lens as an out-of-focus image point, which can be seen in sharp focus by adjusting the eyepiece or imaging lens. However, this is not true when a fiber bundle is used as a relay system because the light pattern imaged on one end of the bundle will be reproduced at the other end. The out-of-focus image blur falls on several fibers and is transferred to the other end. It is not possible to make this blur from several fibers appear to be emerging from a single fiber by adjusting the eyepiece or imaging lens. Therefore, any loss of image quality at the input end of the bundle caused by a focus error is an unrecoverable loss.

The depth of field can be calculated from the permissible focus error ρ . In Fig. 3.18, an objective lens with a focal length F_0 images the object point O_0 perfectly onto the fiber surface and images the object point O_1 in front of the fiber surface, producing a blurred image ρ' on the fiber surface. D_0 , D_1 , and D_2 are the distances from the focal point to the object points O_0 , O_1 , and O_2 ; and D'_0 and D'_1 are the distances from the focal point to the image points O'_0 and O'_1 .

The distance from the defocused image point O'_1 to the fiber surface Z'_1 is

$$Z'_1 = D'_0 - D'_1. \quad (3.12)$$

Using Newton's formulas,

$$D_0 D'_0 = -F_0^2, \quad (3.13a)$$

$$D_1 D'_1 = -F_0^2, \quad (3.13b)$$

Eq. (3.12) becomes

$$Z'_1 = D'_0 - D'_1 = F_0^2 \left(\frac{1}{D_0} - \frac{1}{D_1} \right). \quad (3.14)$$

The radius of the blurred image ρ' is

$$\rho' = Z'_1 \tan(\alpha'_1), \quad (3.15)$$

where α'_1 is the ray angle of the marginal ray from D_1 . The magnification m_1 of the object point D_1 is given by

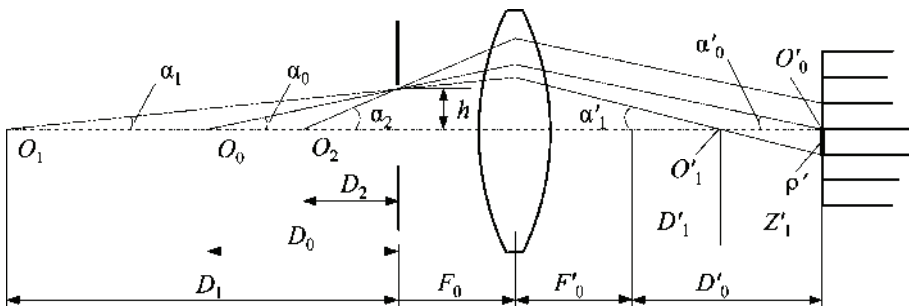


Figure 3.18 Depth of focus/field of an imaging system employing a fiber bundle.

$$m_1 = \frac{\sin(\alpha_1)}{\sin(\alpha'_1)} = \frac{F_0}{D_1}. \quad (3.16)$$

Therefore,

$$\sin(\alpha'_1) = \frac{D_1 \sin(\alpha_1)}{F_0}, \quad (3.17)$$

For a small angle,

$$\sin(\alpha'_1) \approx \tan(\alpha'_1). \quad (3.18)$$

Therefore,

$$\sin(\alpha'_1) = \frac{D_1 \sin(\alpha_1)}{F_0} = \frac{D_1 \tan(\alpha_1)}{F_0} = \frac{h}{F_0}. \quad (3.19)$$

where h is the radius of entrance pupil, as shown in Fig. 3.18. By combining Eqs. (3.14), (3.15), and (3.19), we have

$$\frac{1}{D_1} = \frac{1}{D_0} - \frac{\rho'}{hF_0}. \quad (3.20)$$

For the maximum permissible focus error ρ , the largest and nearest object distances can be estimated from the following equations:

$$\frac{1}{D_{\max}} = \frac{1}{D_0} - \frac{\rho}{hF_0}, \quad (3.21)$$

$$\frac{1}{D_{\min}} = \frac{1}{D_0} + \frac{\rho}{hF_0}. \quad (3.22)$$

From Eqs. (3.21) and (3.22), the larger the distance of the best focused object point D_0 from the focal point, the greater the depth of field ($D_{\max} - D_{\min}$).

When D_{\max} is infinity, the best focused distance is

$$D_0 = \frac{hF_0}{\rho}, \quad (3.23)$$

and the minimum object distance is

$$D_{\min} = \frac{hF_0}{2\rho}. \quad (3.24)$$

When the focal length of the imaging lens and the diameter of the fiber core are fixed, the only way to increase the depth of field is to reduce the radius of aperture h . However, the brightness of the image is reduced by a factor of $(h'/h)^2$, where h' and h are the new and original radii of the aperture, respectively.

3.5.4 Optical properties of imaging systems using fiber bundles

Figure 3.19 shows a typical optical configuration of an imaging system employing a fiber bundle. The objective lens forms an image, OA , of a line onto the fiber bundle surface, then the fiber bundle samples the image. The light from each sampled point travels along the fiber and emerges from the other end of the fiber bundle. The image, $O'A'$, acts as the object for either the eyepiece to produce a magnified virtual image or the imaging lens to form an image onto the image sensor.

In a conventional endoscope with a lens relay system, the eyepiece or imaging lens can be designed to compensate for the residual aberrations accumulated from the objective and relay lenses. The eyepiece or imaging lens in a fiber endoscope can only give a sharp image of what appears at the proximal end of the fiber bundle, and cannot correct the blur of the image formed on the distal end by the objective lens and transmitted by the fiber bundle. Therefore, the objective lens should provide a high-quality image.

The objective lens should be telecentric in the image space, which means that the chief ray of each image point is parallel to the optical axis of the objective lens. When the optical axis is normal to the fiber surface, all of the chief rays are perpendicular to the fiber surface. The NA of the objective lens should be equal to or slightly smaller than that of the fiber bundle to reduce the crosstalk between the cores.

If a nontelecentric objective lens is designed for a fiber-optic imaging system, some or all of the light from the off-axis object point will not be coupled into and transmitted through the fiber, even if the NA of the objective lens in fiber space is smaller than the NA of the fiber. This loss occurs because the incident angle of some off-axis light is larger than the acceptance angle of the fiber. As shown in Fig. 3.20(a), only the portion of light from the off-axis object point filled with hatched lines is within the acceptance angle of the fiber and can be transmitted through the fiber. Therefore, less light is coupled into the fiber for the off-axis point than for the on-axis point.

The central rays of the light emerging from the fibers are also perpendicular to the surface of the fiber bundle. The eyepiece or imaging lens is required to be telecentric on the fiber side in order to use as much light from the fiber bundle as possible. Otherwise, the light from the off-axis point may not be able to enter the eyepiece or imaging lens.

As discussed in Sec. 3.3, the straight multimode fiber redistributes light spatially at the output end of the fiber without changing the range of ray angles. Therefore, when the objective lens is not telecentric, the light emerging from the fiber surface is a cone, as shown in Fig. 3.20(a). This effect can be minimized by twisting the fibers in the fiber bundle.

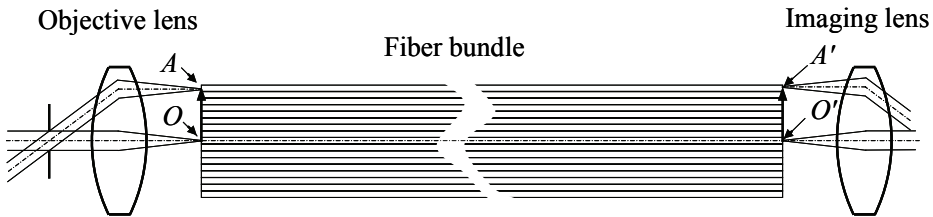


Figure 3.19 Optical layout of an optical system using a fiber bundle.

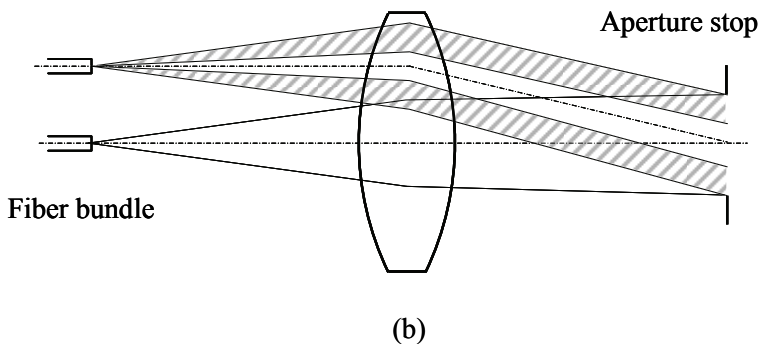
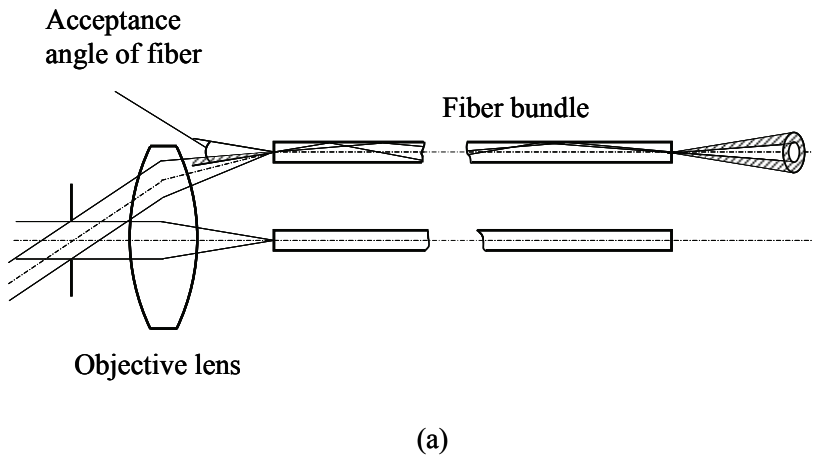


Figure 3.20 (a) Optical layout of a fiber-optic relay with a nontelecentric objective lens. (b) The ray path of the light emerging from the fiber when an ideal telecentric lens is used.

When a telecentric imaging lens is used, light passes only through the outer parts of the aperture stop where the aberrations are usually larger than the central portion of the pupil, as shown in Fig. 3.20(b). Even with a perfect imaging system, the contrast of the image is degraded when only the outer ring of the aperture stop is filled. Figure 3.21 shows the MTFs of an ideal imaging lens when the pupil is fully filled and annularly filled. MTF in the middle frequency is significantly reduced when only the outer annular portion of the aperture stop is filled with light.

For a digital imaging system, where the output surface of an imaging fiber bundle is imaged onto a digital imaging detector, such as a CCD or a CMOS, aliasing is possible because the fiber bundle may not adequately sample the image produced by the objective lens, and the detector may not adequately sample the image produced by the imaging fiber bundle. Therefore, when designing an imaging system to image the output surface of an imaging fiber bundle to a detector, one must ensure that the detector provides adequate sampling for the image detail transmitted through the fiber bundle.

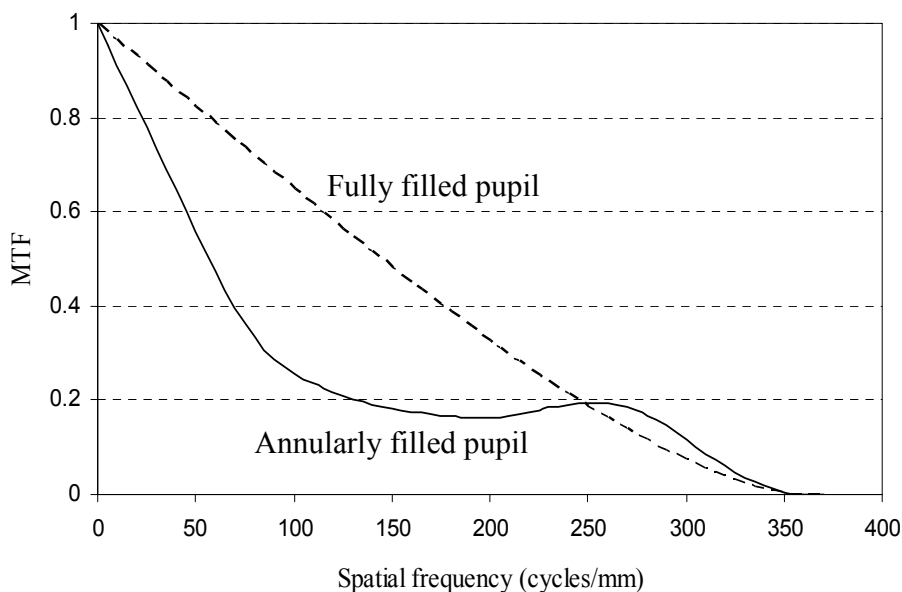


Figure 3.21 MTFs of an ideal telecentric lens when the pupil is fully filled and annularly filled.

References

1. T. Takahashi, I. Sugimoto, T. Takabayashi, and S. Yoshida, "Optical transmission loss of liquid core silica fibers in the infrared region," *Opt. Commun.* **53**, 164–168 (1985).

2. A. Inberg, M. Ben-David, M. Oksman, A. Katzir, and N. Croitoru, "Theoretical model and experimental studies of infrared radiation propagation in hollow plastic and glass waveguides," *Opt. Eng.* **39**(5), 1316–1320 (2000) [doi: 10.1117/1.602505].
3. M. Ben-David, A. Inberg, I. Gannot, and N. Croitoru, "The effect of scattering on the transmission of IR radiation through hollow waveguides," *J. Optoelectron. Adv. Mater.* **3**, 23–30 (1999).
4. M. Ben-David, I. Gannot, A. Inberg, and N. I. Croitoru, "Mid-IR optimized multilayer hollow waveguides," *Proc. SPIE* **4253**, 11–18 (2001) [doi: 10.1117/12.427923].
5. F. Zolla, G. Renversez, A. Nicolet, B. Kuhlmeiy, S. Guenneau, and D. Felbacq, *Foundations of Photonic Crystal Fibres*, Imperial College Press, London (2005).
6. F. Poli, A. Cucinotta, and S. Selleri, *Photonic Crystal Fibers: Properties and Applications*, Springer, Dordrecht, The Netherlands (2007).
7. C. Yeh, *Handbook of Fiber Optics: Theory and Applications*, Academic Press, San Diego, CA (1990).
8. J. S. Sanghera and I. D. Aggarwal, Eds., *Infrared Fiber Optics*, CRC Press, Boca Raton, FL (1998).
9. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, John Wiley and Sons, New York (1992).
10. R. A. Shotwell, *An Introduction to Fiber Optics*, Prentice Hall, Englewood Cliffs, NJ (1997).
11. W. J. Cassarly and T. L. R. Davenport, "Non-rotationally symmetric mixing rods," *Proc. SPIE* **6342**, 63420Q (2006) [doi: 10.1117/12.692256].
12. R. M. Verdaasdonk and C. Borst, "Optics of fibers and fiber probes," in *Optical-Thermal Response of Laser-Irradiated Tissue*, A. J. Welch and M. J. C. van Gemert, Eds., pp. 619–666, Plenum, New York (1995).
13. J. A. Udovich, N. D. Kirkpatrick, A. Kano, A. Tanbakuchi, U. Utzinger, and A. F. Gmitro, "Spectral background and transmission characteristics of fiber optic imaging bundles," *Appl. Opt.* **47**, 4560–4568 (2008).
14. H. H. Hopkins, "Physics of the fiberoptic endoscope," in *Endoscopy*, G. Berci, Ed., pp. 27–63, Appleton-Century-Croft, New York (1976).
15. K. L. Reichenbach and C. Xu, "Numerical analysis of light propagation in image fibers or coherent fiber bundles," *Opt. Express* **15**, 2151–2165 (2007).
16. N. S. Kapany, J. A. Eyer, and R. E. Keim, "Fiber optics. Part II. Image transfer on static and dynamic scanning with fiber bundles," *J. Opt. Soc. Am.* **47**(5), 423–425 (1957).

Chapter 4

Microscope Optics

Optical microscopes are one of the oldest optical instruments and are used to produce magnified visual or photographic images of small objects. The oldest microscope was the magnifying glass, which was successfully used to observe some large bacteria. In the early 1600s, the first compound microscope with an objective lens and an eyepiece was invented by the Janssen brothers in the Netherlands. All modern microscopes are developed from the compound configuration.

To meet various requirements, many types of microscopes have been developed, for example, polarization, phase contrast, fluorescence, confocal, digital, and nonlinear. Optical microscopes have become an essential tool for biomedical applications, from investigating biological processes and diagnosing diseases to quantitatively measuring biological processes *in vitro* and *in vivo*.

Microscope optics is fundamental to many other biomedical imaging systems, such as endoscopy, confocal imaging, fluorescence imaging, and polarization imaging, all of which will be discussed in the following chapters.

4.1 Introduction

Figure 4.1 shows two basic configurations of microscope optics. The exit pupil of the objective lens is set at the rear focal plane; however, this is not an essential requirement for a microscopic objective lens. The advantage of an exit pupil at the rear focal plane is that the objective lens is telecentric in object space, which means that the chief rays are parallel to the optical axis.

Figure 4.1(a) depicts a classic configuration with a finite-corrected objective lens whose image plane is located at a finite distance from the objective lens. The distance between the rear focal plane and the intermediate image plane is defined as the optical tube length. The eyepiece forms a virtual image of the intermediate image at a distance 250 mm from its aperture, where the iris of the observer is located for an optimal viewing condition. Sometimes the eyepiece is designed to image the intermediate image to infinity.

In order to insert additional components, such as a prism, beamsplitter, or filter, between the objective lens and the eyepiece without shifting the final image plane and without introducing aberrations, a Telian telescope system is

necessary to provide a collimated space where the flat optical components can be placed without degrading image quality and shifting the image plane. A Telan system typically consists of two doublets, with a negative doublet to collimate the light and a positive doublet to focus the light onto the image plane. The light between the two doublets is collimated. The Telan system does not shift the image plane, but it may change the magnification and, possibly, the pupil location.

Figure 4.1(b) illustrates the layout of a modern microscope with an infinity-corrected objective lens, which images objects to infinity. The light from a point object in the object plane is collimated by the objective lens. A tube lens is generally used to focus the collimated light onto the intermediate image plane located at its back focal plane.

Although image-forming characteristics are insensitive to changes in the optical path length between the objective lens and the tube lens, the location of the exit pupil of the system does change. The infinity-corrected objective lens gives a better modular structure because there is a collimated space behind the

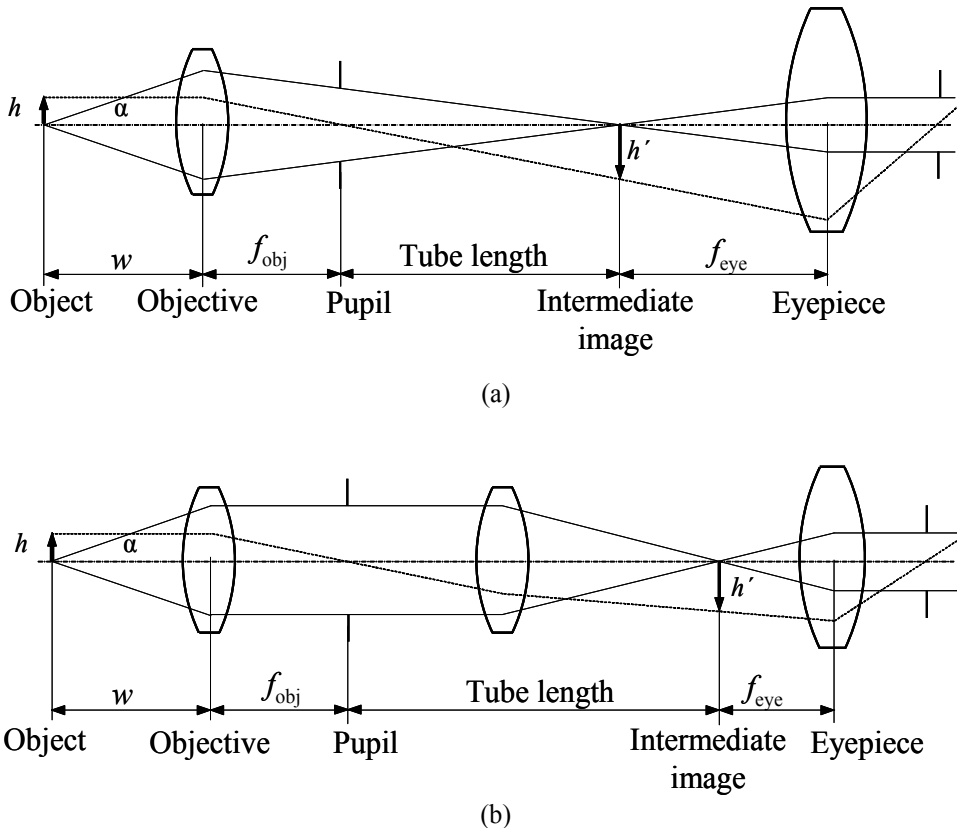


Figure 4.1 Basic configurations of (a) a classic microscope with a finite-corrected objective lens and (b) a modern microscope with an infinity-corrected objective lens.

objective lens. It is possible to insert subsystems, such as a fluorescence filter cube, between the objective and tube lenses without significant impact on the optical performance of the system. In addition, it is possible to focus the image by moving only the objective lens, instead of moving the stage or the entire optical system.

Although the infinity-corrected objective lenses are designed for an infinite conjugate, they are not, in general, interchangeable between microscope systems from different manufacturers, because, as will be discussed later, some objective lenses are optimized with specific tube lenses so that the residual aberrations of the objective lens are compensated by the tube lens.

Figure 4.2 shows the illumination and imaging ray paths of a microscope with an infinity-corrected objective lens and Köhler illumination. Consideration of field and pupil imaging is important in understanding microscopic image formation. When an eyepiece is used to observe an object, the retina is conjugate to the object plane, intermediate image plane, and field aperture. The pupil plane is conjugate to the light source, aperture stop of the condenser, and the pupil of the objective lens. The aperture stop of the condenser also defines the conditions of coherence.

A modern microscope typically has the following major components:

- Objective lens: images the object onto the intermediate image plane or to infinity;
- Tube lens: focuses the collimated light to the intermediate image plane;
- Eyepiece: magnifies the intermediate image for the human eye;
- Imaging detector: captures the live or still image;
- Light source: provides illumination light;
- Light collector: collects the light from the light source;
- Condenser: focuses the light onto the object; and
- Beamsplitter: separates the illumination light and the detection light in an epi-illumination mode.

4.1.1 Key specifications

For practical purposes, the image size, total conjugate distance (the distance from the object to the image), and other parameters of the objective lens can be optimized according to the application. However, for standard microscopes, each manufacturer usually standardizes those parameters so that the microscopes can be built in a modular way and the objective lenses and eyepieces can be interchanged.

Figure 4.3 shows the key parameters generally used to standardize objective lenses. The diagram is a simplified, unfolded optical path. In a real microscope, the optical axis is not a straight line, and there are prisms between the objective lens and eyepiece.

The parfocal distance is the distance between the object plane and the objective lens shoulder by which the objective lens is supported on a turret. For

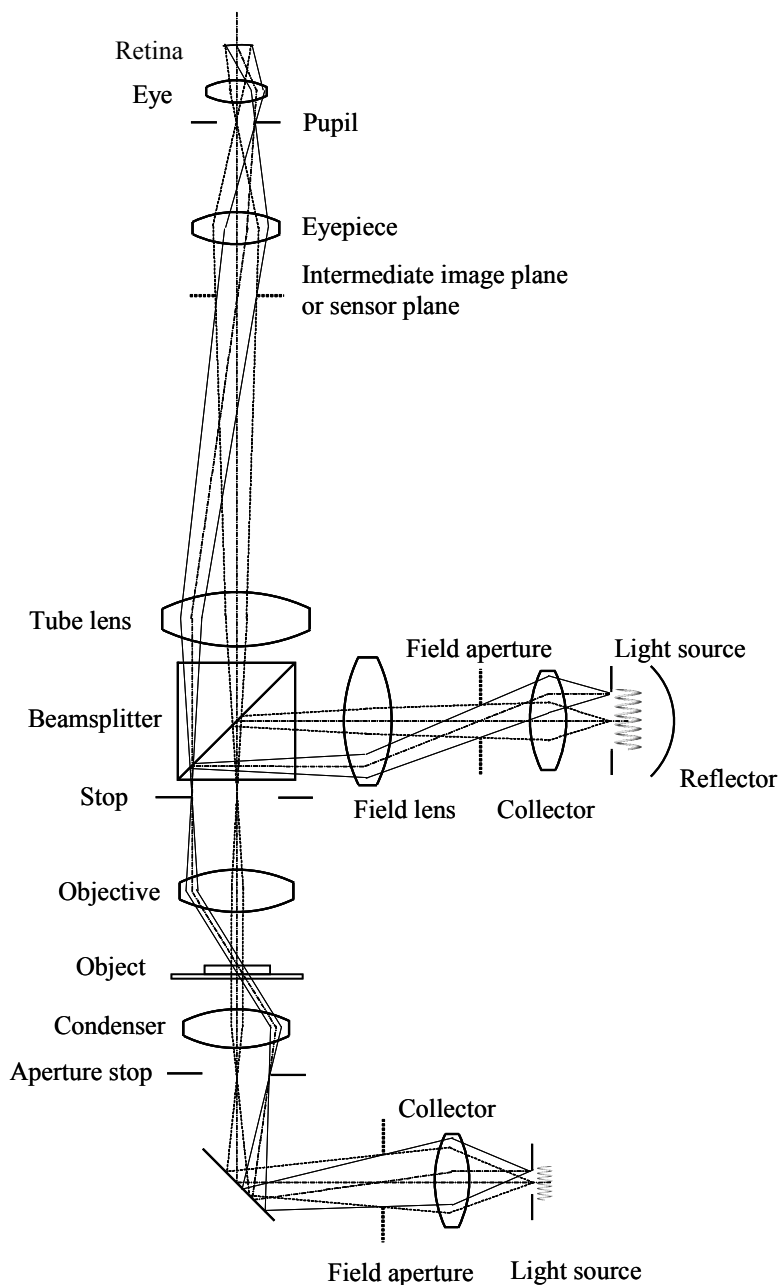
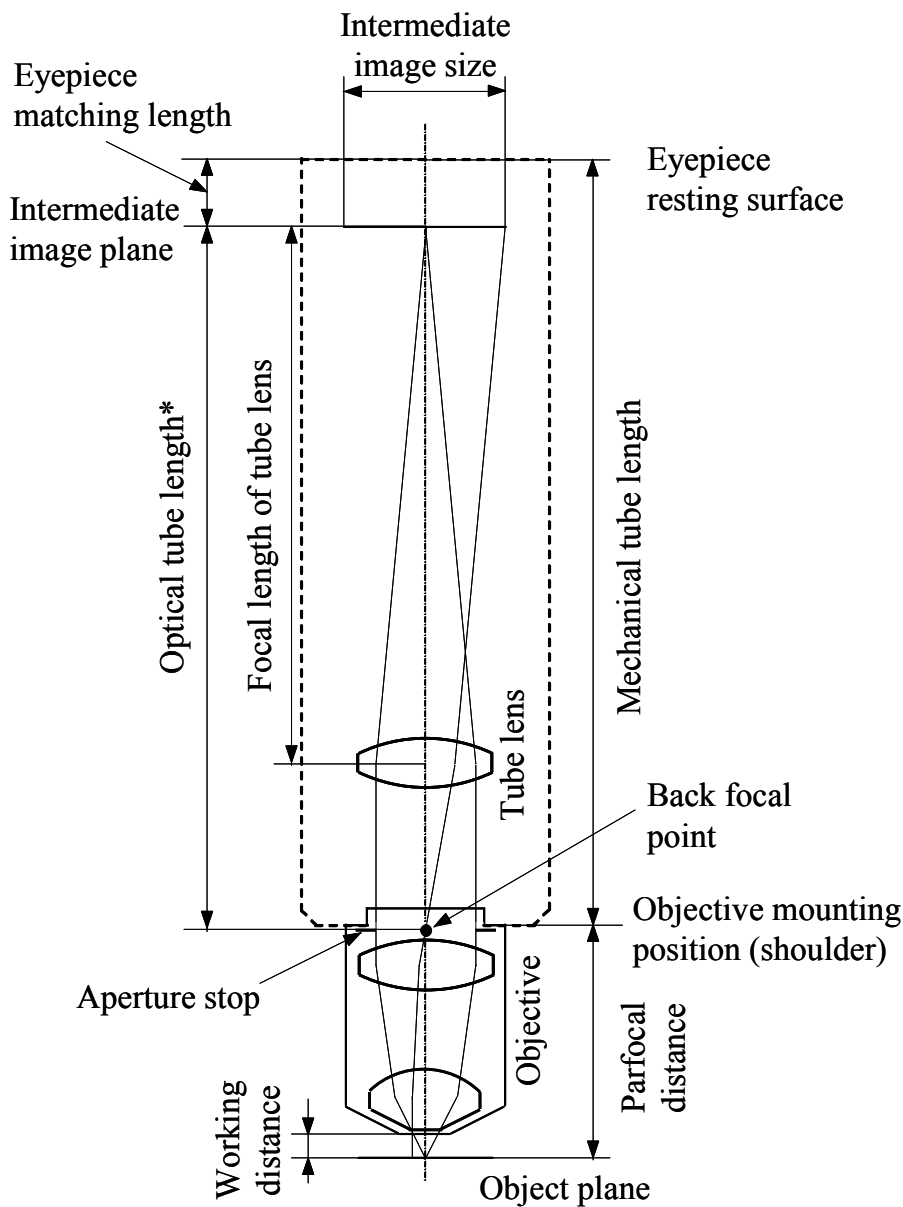


Figure 4.2 Layout of a compound microscope. The dashed ray path indicates the pupil imaging, and the solid ray path indicates the field imaging.

classic microscopes with finitely corrected objective lenses, the objective lenses from most manufacturers have a parfocal distance of 45.0 mm. Because of different design criteria for aberration correction, working distance, and field



* Typically specified for finitely corrected objective.

Figure 4.3 Specifications of an objective lens and a compound microscope.

size, infinity-corrected objective lenses from the same manufacturer usually have identical parfocal distances, but these distances may not be 45.0 mm.

The optical tube length is the distance between the rear focal plane of the objective lens and the intermediate image and is typically used in microscopes with finitely corrected objective lenses. The mechanical tube length is the distance from the turret opening, where the objective lens is mounted, to the top edge of the observation tubes, where the eyepiece rests. The tube length of 160 mm suggested by the Royal Microscopical Society is commonly used for finitely corrected objective lenses, but other tube lengths, for example, 170 and 210 mm, are also used by some manufacturers. For an infinity-corrected objective lens, the mechanical tube length is determined by the tube lens and its location from the objective lens. In a finite-tube-length microscope system, whenever an optical element, such as a beamsplitter, is placed in the light path between the objective lens and the eyepiece, the optical tube length becomes greater than the designed length. Thus, it is necessary to add a Telan system to bring the intermediate image to the original plane.

The focal length of the tube lens is usually between 160 and 250 mm, depending on the manufacturer. With a fixed intermediate image size, longer focal lengths produce smaller off-axis angles for diagonal light rays, producing correspondingly smaller shifts in off-axis rays passing through accessory components, such as dichroic mirrors or filters.

The intermediate image size is usually defined by each manufacturer at a fixed diameter. The typical sizes are 18, 20, 23, 25, and 28 mm. The location of the intermediate image plane inside the microscope tube is usually standardized to 10 mm from the eyepiece end. This distance is called the eyepiece matching distance.

In terms of optics, the working distance is the distance from the last surface of the objective lens to the closest surface of the cover glass or the actual surface of the object if no cover glass is used. In practice, the working distance is calculated from the last mechanical surface, instead of the last lens surface of the objective lens. The working distance of an objective lens generally decreases as its NA increases.

4.1.2 Key parameters

4.1.2.1 Numerical aperture

The most important design or selection criterion of an objective lens is NA, which is defined as

$$\text{NA} = n \sin(\alpha), \quad (4.1)$$

where n is the refractive index of the medium between the front lens of the objective lens and the object, with a value that ranges from approximately 1.00 for air to 1.51 for a specialized immersion medium and even to 1.78 for TIR fluorescence applications; α is the acceptance angle of the objective lens, as shown in Fig. 4.1. The NA determines the light gathering power, the resolving

power, and the depth of field of the objective lens. The light collection efficiency is proportional to NA^2 .

In order to obtain higher working NAs, one can increase the acceptance angle α of the objective lens by using a more complex design. The theoretical maximum acceptance angle is 90 deg; therefore, the maximum NA in air is 1.0. To further increase the NA, an immersion liquid, such as synthetic oil with very low autofluorescence, glycerin, or water, is usually used between the front lens and the object or cover glass to increase the NA by a factor of n . For example, when the acceptance angle of the objective lens is 65 deg and water is used as the immersion medium, the NA is $1.33\sin(65\text{deg}) = 1.21$.

In biomedical microscopic imaging, the sample is typically protected by a cover glass, and the sample touches the cover glass. Given that the cover glass is in the imaging path, its effect on image formation should be considered when designing and using the objective lens. Figure 4.4 shows how the immersion medium increases the light collection when a cover glass is used. As shown in Fig. 4.4(a), when there is no immersion medium, ray a from the object can pass the edge of the aperture stop in the objective lens. Ray b exits the cover glass but is blocked by the objective lens because the exit angle is larger than the acceptance angle of the objective lens. Ray c, with a larger angle, cannot exit the cover glass because of TIR. On the other hand, when an immersion liquid is used, as shown Fig. 4.4(b), all three rays, a, b, and c, exit the cover glass and enter and pass through the objective lens. This effect occurs because the ray exit angle becomes smaller when the refractive index of the medium between the objective lens and the cover glass increases, and the TIR condition is broken when the refractive index of the immersion medium matches that of the cover glass.

4.1.2.2 Magnification

The total magnification of a microscope system with an eyepiece is the product of the magnifications of the objective lens and the eyepiece:

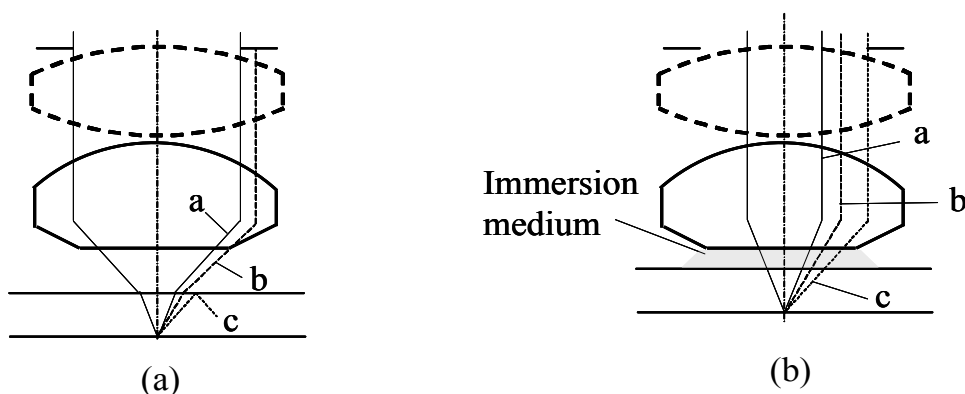


Figure 4.4 Effect of the immersion liquid on a high-NA objective lens. (a) Without immersion liquid and (b) with immersion liquid.

$$M = M_{\text{obj}} M_{\text{eye}}. \quad (4.2)$$

The magnification of the eyepiece is approximated as

$$M_{\text{eye}} \approx \frac{250}{f_{\text{eye}}}, \quad (4.3a)$$

where f_{eye} is the focal length of the eyepiece in millimeters.

For finitely corrected objective lenses, the magnification is usually standardized, roughly with a factor of $\sqrt[10]{10} = 1.259$ between two succeeding values when the magnification is larger than 10. This standardization gives the following typical magnifications: 10, 12.5, 16, 20, 25, 32, 40, 63, 80, and 100.

For an infinity-corrected objective lens, the magnification is determined by the focal length of the objective lens and that of the tube lens:

$$M_{\text{obj}} = \frac{f_{\text{tube}}}{f_{\text{obj}}}, \quad (4.3b)$$

where f_{tube} and f_{obj} are the focal lengths of the tube lens and the objective lens, respectively.

The objective lens is designed for diffraction-limited performance; therefore, the magnification is usually related to the NA, but not strictly. The NA typically increases with the magnification in order to achieve a usable resolution. For the same magnification, the NA generally increases as the aberration correction is improved. For example, the NAs are 0.25, 0.30, and 0.40 for 10× achromat, plan fluorite, and plan apochromat, respectively.

4.1.2.3 Resolution

The resolution of a microscope objective lens is defined as the smallest distance between two points that can still be distinguished. NA determines the resolving power of an objective lens, but the total resolution of a microscope system is also dependent on the NA of the illumination condenser.

Using the Rayleigh criterion, two points are resolvable if they are separated by a distance such that the peak of one Airy disk pattern falls onto the first zero of the second Airy disk pattern. The lateral resolution, which is the distance between the two points, is

$$r = \frac{0.61\lambda}{\text{NA}}, \quad (4.4)$$

where λ is the wavelength for imaging.

Equation (4.4) is valid when the object is self-luminous, such as a fluorescent molecule, or when the range of the illumination angle is equal to or larger than the acceptance angle of the objective lens. When the NA of the illumination is smaller than that of the objective lens, the pupil of the objective lens is partially filled, and the resolution is

$$r = \frac{1.22\lambda}{n \sin \theta_{\text{ill}} + n \sin \theta_{\text{obj}}}, \quad (4.5)$$

where θ_{ill} and θ_{obj} are the illumination angle of the condenser and the acceptance angle of the objective lens, respectively.

Equation (4.5) shows that the higher the NA, the better the resolution, as illustrated in Fig. 4.5(a). This equation also indicates that shorter wavelengths yield higher resolution. The greatest resolving power in optical microscopy is achieved with near-UV light and a high-NA liquid-immersion objective lens.

Axial resolution, like lateral resolution, is determined only by the NA of the objective lens and the working wavelength λ .¹

$$z_{\text{min}} = 2 \frac{n\lambda}{\text{NA}^2}, \quad (4.6)$$

where z_{min} reduces inversely proportionally with the square of NA instead of the first power of NA, meaning that the axial resolution is affected by NA much more than the lateral resolution is affected. This characteristic is clearly shown in Fig. 4.5(b).

The ratio of axial-to-lateral resolution is

$$R = \frac{z_{\text{min}}}{r} = \frac{3.28n}{\text{NA}}. \quad (4.7)$$

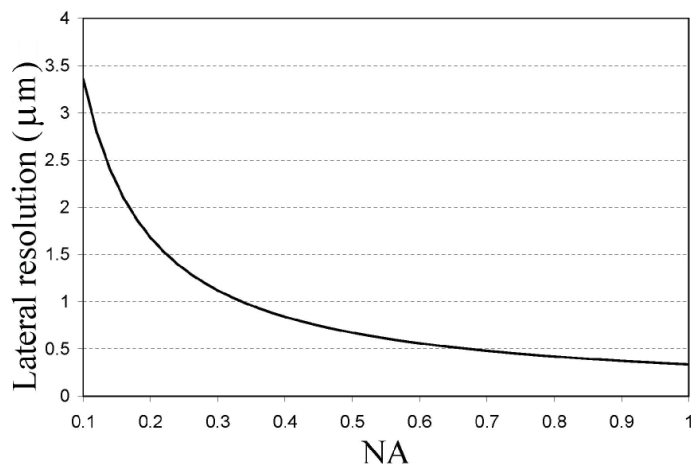
The ratio is larger than 1 and does not depend on the wavelength.

4.1.2.4 Depth of field

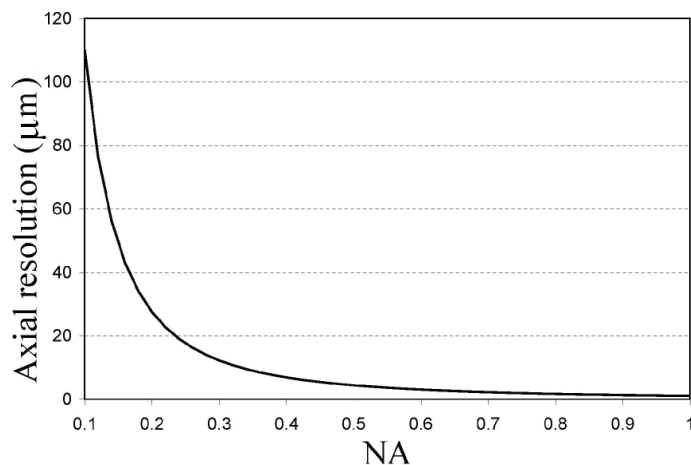
Depth of field is determined by the distance from the nearest object plane that is in focus to the farthest plane that is also in focus. Within this range, an objective lens can be focused without any appreciable change in image sharpness. Another commonly used parameter, DOF, is the range of the acceptable focus in the image plane.

The total depth of field is the sum of the wave and geometrical optical depth of field, as given by

$$d_{\text{tot}} = \frac{n\lambda}{\text{NA}^2} + \frac{ne}{M\text{NA}}, \quad (4.8)$$



(a)



(b)

Figure 4.5 Resolution of an objective lens as a function of NA: (a) lateral resolution and (b) axial resolution.

where e is the smallest distance that can be resolved by a detector placed in the image plane of the objective lens with a magnification of M . For CCD or CMOS, e is the pixel size. The first term of Eq. (4.8) is the diffraction-limited depth of field, which is a contribution from wave optics, and the second term is related to the geometrical optical circle of confusion.

Equation (4.8) is valid when the illumination is incoherent and the NA of the condenser is greater than or equal to that of the objective lens. The depth of field generally increases as the coherence of illumination increases, which means the condenser's NA decreases. The depth of field can be much shallower than that predicted from Eq. (4.8) for some phase-based microscopes.

Normally, the human eye can accommodate from infinity to about 250 mm, so that the depth of field is greater than that predicted by Eq. (4.8) when one observes a microscopic image through an eyepiece. On the other hand, in digital and video microscopes whose sensors are flat, the depth of field is given by Eq. (4.8).

Depth of focus is the counterpart of the depth of field on the image side and is determined by the NA of the objective lens on the image side and the pixel size when a detector is used. Depending on the magnification, the DOF can be larger for an objective lens with a high NA and higher magnification than that of an objective lens with a small NA and a low magnification.

4.1.3 Key components

4.1.3.1 Objective lens

The objective lens is the most important optical element in a microscope because it is responsible for the primary image formation and it determines the image quality. It is also the most difficult to design and manufacture.

The objective lens is usually specified by NA, FOV, magnification, working distance, and the degree of aberration correction. An objective lens is typically classified by its type of aberration correction (see Table 4.1):

- **Achromat:** The lens is corrected for axial chromatic aberration at only two wavelengths, usually at 486 and 656 nm, and for spherical aberration at 587 nm. The field curvature is not corrected. The NA is usually low, and the magnification is medium. Limited color correction can lead to substantial artifacts when a broadband light source is used. The achromat provides high contrast and sharpness within about one-third of the linear field, which is sufficient in many microscopic studies.
- **Plan achromat:** In addition to requiring the same chromatic and spherical aberration corrections as the classic achromat objective lens, the field is flat with a low Petzval curvature and a small astigmatism.
- **Fluorite:** The lens corrects both chromatic and spherical aberrations for two or three wavelengths. Calcium fluoride, which has a low refractive index equal to 1.43 and an unusual dispersion of 95.0, is typically used to minimize chromatic aberrations.
- **Plan fluorite:** The lens is fluorite with a flat field.
- **Apochromat:** The chromatic aberrations are corrected for three colors (red, green, and blue), and the spherical aberration is corrected for either two or three wavelengths. Calcium fluoride is often used as a positive element that is cemented to two negative glasses.
- **Plan apochromat:** In addition to requiring the correction for chromatic and spherical aberrations for three wavelengths, the lens has a flat field with excellent corrections of Petzval curvature and astigmatism.

Table 4.1 Types of objective lenses and aberration corrections.

Objective type	Spherical aberration	Chromatic aberration	Field curvature
Achromat	1 color	2 colors	No
Plan achromat	1 color	2 colors	Yes
Fluorite	2-3 colors	2-3 colors	No
Plan fluorite	2-3 colors	2-3 colors	Yes
Apochromat	3-4 color	3-4 colors	No
Plan apochromat	3-4 color	3-4 colors	Yes

In addition to some special objective lenses developed for phase contrast, polarized light, confocal, fluorescence, and other imaging modalities, the following are some examples of special objective lenses:

- Long working distance objective lenses have a longer working distance for some applications, such as cell perfusion experiments. The lens size is usually larger than the regular microscope objective lens with similar specifications.
- Objective lenses with a correction collar have a movable lens group to compensate for the spherical aberrations introduced by the small variation in the refractive indices of immersion and mounting media, as well as small thickness variation in the cover glass.
- Water-dipping objective lenses do not require that the specimen be observed with a cover glass and can be dipped directly into the cell culture dish, thus making them highly desirable for the study of live cells and physiology. The NAs are slightly lower than those of water-immersion lenses, but the working distances are much longer. These lenses are ideal for any application requiring micromanipulation, such as physiology, imaging of brain tissue, and *in vitro* fertilization applications.

4.1.3.2 Tube lens

The tube lens is used together with the infinity-corrected objective lens to form an image at the intermediate image plane. In some cases, in addition to focusing the collimated light from the objective lens to the intermediate image plane, the tube lens also compensates for the lateral chromatic aberration from the objective lens. For commercial microscopes, the focal lengths of the tube lenses range from 160 to 250 mm, depending on the manufacturer. For custom digital microscopes, the focal length is generally determined by the size of the detector.

As shown in Fig. 4.6, the light from the on-axis point on the object (shorter-dashed lines) is collimated by the objective lens and is parallel to the optical axis. The light from the off-axis point on the object (solid lines) is also collimated by the objective lens but is not parallel to the optical axis. The angle β between the optical axis and the collimated ray increases as the point moves away from the optical axis.

Given that the light rays for any individual point on the object are parallel in the space between the objective lens and the tube lens, optical components with flat surfaces can be placed in this space without introducing aberration and changing the distance between the object and the intermediate image.

The location of the tube lens with respect to the objective lens is a major consideration when designing tube lenses. If the tube lens is too close to the objective lens, the space for auxiliary optical elements is limited. However, if a tube lens is too far from the objective lens, the rays from the off-axis points may be blocked by the tube lens, resulting in an image with darkened edges. It is necessary to optimize the distance between the objective lens and the tube length in order to maximize the flexibility of a microscope and maintain system performance.

4.1.3.3 Eyepiece

As shown in Fig. 4.7, an eyepiece is used to further magnify the intermediate image formed by the objective lens. It relays the intermediate image to the human eye, compensates for some remaining aberrations from the objective lens, and provides measurement capabilities with a measuring graticule.

The characteristic parameters of an eyepiece include focal length, field angle, and eye relief. The focal length determines the magnification of the eyepiece, as indicated in Eq. (4.3a). Typically, eyepieces have a semifield angle ω between 20 and 30 deg. Larger FOVs are possible but require more complicated designs.

Eye relief is the distance between the last surface of the eyepiece and its exit pupil where the iris of the human eye is located. Typically, the eye relief varies between 0.5 and 0.8 of the focal length. As a general guideline, the eye relief should be 15 mm or longer for a person wearing spectacles.

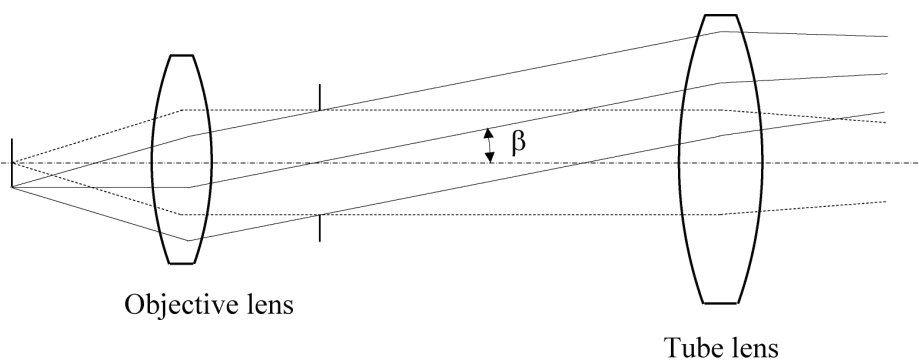


Figure 4.6 Tube lens for the infinity-corrected objective lens.

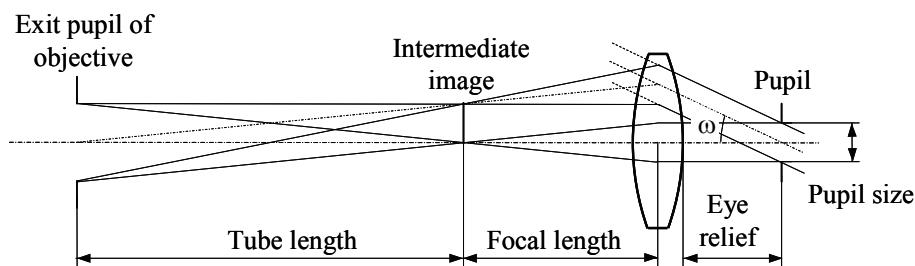


Figure 4.7 Principle of an eyepiece.

There is no hard aperture stop in the eyepiece. The pupil of a human eye, which varies in size from 1–6 mm, forms the physical stop for the eyepiece. The variations in human pupils should be considered when designing an eyepiece or evaluating the performance of an eyepiece. A 5-mm exit pupil diameter is usually used as the design parameter. If the pupil diameter is too small, part of the image will disappear when the human eye moves.

The eyepiece is used in conjunction with an objective lens; therefore, the entrance pupil of the eyepiece should match the exit pupil of the objective lens, in both size and location. This means that the eyepiece images the exit pupil of the objective lens onto the iris of the human eye. The ray diagram in Fig. 4.7 shows those relationships.

4.1.3.4 Light source

The basic requirements for light sources are optimal spectral coverage and sufficient radiance. The selection criteria of light sources for microscope illumination include source structure, spatial light distribution, spectrum, spatial and temporal stability, brightness, coherence, lifetime, and the degree to which these various parameters can be controlled.

Tungsten-halogen lamps are one of the most commonly used light sources because of their low cost and long lifetime. They are relatively stable over long time periods and are subject to only minor degrees of output fluctuation (under normal operating conditions). They are relatively bright with a color spectrum centered at 3200 K. Tungsten-halogen lamps may need color filters to convert the color temperature to a daylight equivalent.

Arc lamps, such as mercury, xenon, and metal halide, used in optical microscopy are generally several orders of magnitude more radiant than tungsten-halogen lamps of comparable wattage, primarily because of the small size of the arc compared to the incandescent lamp filament. Xenon lamps are commonly used in microscopes. They have a relatively even output across the visible spectrum and a color temperature close to daylight. They are also very bright and have long lifetimes. Arc lamps are significantly less stable than tungsten-halogen lamps because their gas plasma is inherently unstable and can be affected by magnetic fields and the erosion of electrode tips. Both tungsten-

halogen and arc lamps require considerable time after ignition to reach thermal equilibrium.

In recent years, LEDs have been increasingly accepted in microscope illumination. LEDs are compact semiconducting devices that emit incoherent narrow-spectrum light. The spectrum of the emitted light depends on the composition and condition of the semiconducting material used. LEDs can emit UV, visible, and near-IR light. White light can also be emitted by using phosphor conversion. LEDs produce an output with a spectrum bandwidth wider than that of the laser but much narrower than the arc lamp. LEDs possess some desirable qualities: small size, low power consumption, instant switching, and long lifetime. The diverse spectra afforded by LEDs make it possible to select an individual LED to supply the optimum illumination wavelength for fluorescence imaging and other purposes. The color image can be obtained by using either white LEDs, based on phosphor conversion, or RGB LEDs. For RGB LED light sources, the independent and accurate control of red, green, and blue LEDs can create any desired color. With RGB LEDs as the light source, color images can also be produced from monochrome cameras, giving full noninterpolated color at every pixel.

All light sources produce a significant level of heat, and their emission output depends on the source temperature. LEDs have the lowest operating temperature of all lamps and are the most stable in temporal, spatial, and spectral light distributions. For long-term stability, tungsten-halogen lamps and LEDs exhibit better performance than arc lamps.

4.1.3.5 Illumination system

In order to observe a nonluminous object under a microscope, the object must be illuminated. Illumination is very important in achieving high-quality images in microscopy. There are two illumination modes in microscope imaging: epi-illumination and transillumination. Each of these modes has bright-field and dark-field methods to illuminate an object; thus there are a total of four illumination configurations. For bright-field illumination, critical and Köhler illumination systems are typically used.

Critical illumination images the entire light source onto the object plane. A larger solid angle can be used to deliver more light to the object plane. Unfortunately, any nonuniformity in the light source will appear as fluctuations in brightness. Coherence of the light source is preserved in critical illumination.

The Köhler illumination system was shown in Fig. 4.2. In transillumination mode, a large-aperture collector gathers the light from the light source and focuses the light onto the aperture stop of the condenser. The condenser images the field aperture to the object plane. The size of the field aperture can be adjusted to match the illumination and observation fields.

4.1.3.6 Imaging detector

The imaging detector, typically a solid-state detector, is placed at the intermediate imaging plane to capture and digitize the image formed by the microscope objective lens and the tube lens. Digitization of an image captured through an optical microscope results in a dramatic increase in ability to enhance features, extract information, and modify the image.

Solid-state detectors consist of a dense matrix of photodiodes incorporating charge-storage regions. CCD and CMOS detectors are the two most commonly used imaging detectors in optical microscopes. In each of these detectors, a silicon-diode photosensor is coupled to a charge-storage region that is connected to an amplifier that reads out the quantity of the accumulated charge. In a CCD, there is only one amplifier at the corner of the entire array. The stored charge is sequentially transferred through the parallel registers to a linear serial register and then to an output node adjacent to the readout amplifier. CMOS devices have an amplifier and digitizer associated with each photodiode in an integrated on-chip format. The individual amplifier associated with each pixel helps reduce noise and distortion levels. It allows gain manipulation of the individual photodiode, region-of-interest readout, high-speed sampling, electronic shuttering, and exposure control.

The efficiency of the image capture through the imaging detector depends on the properties of the microscopic imaging system and the imaging detector. When selecting an imaging detector for microscopic imaging, resolution, image transfer speed, quantum efficiency (QE), dynamic range, and color fidelity are all key parameters to consider.

The final resolution of a microscopic imaging system is a function of the spatial density of pixels in the detector and the optical resolution of the objective lens. Adequate resolution of an object can only be achieved if there are at least two samplings for each resolvable unit. When the optical resolution of the microscope objective lens is better than the resolution of the imaging detector, the final resolution of the digital image is limited only by the image detector. When the resolution of the imaging detector is better than the optical resolution of the microscope objective lens, the final resolution of the digital image is limited only by the objective lens.

4.2 Objective Lenses

4.2.1 Introduction

The objective lens is the most important component in the microscopic instrument; it determines the overall resolution and light collection efficiency. The objective lens is an aperture-dominant optical system in that the FOV is small and the NA is large.

A microscope objective lens is designed to be used either with or without a cover glass over the object. A standard cover glass is made of glass K5 with a thickness of 0.17 mm. In a strongly divergent or convergent cone of light, this

cover glass will introduce a significant amount of aberration, mainly spherical aberration.

To design a high-performance objective lens, the aberration correction should be diffraction limited, at least near the optical axis. The objective lens is usually optimized to minimize the wavefront error. Wavefront aberration or the Strehl ratio is typically used to analyze lens performance in the final stage. According to the Maréchal criterion, the RMS wavefront error of a diffraction-limited objective lens should be less than $\lambda/14$.

The lens elements in a high-performance objective lens can be divided into three groups, as shown in Fig. 4.8(a):²

- (1) Front group: These are the lenses closest to the object. Since the light must be bent at huge angles, the front group usually consists of aplanatic-concentric surfaces to prevent large spherical aberration and coma.
- (2) Middle group: These are generally a combination of cemented elements and singlets located next to the front group. The lenses in the middle group are used to correct spherical aberration, coma, and chromatic aberration. The spherical aberration may be overcorrected to compensate for the remaining spherical aberration from the front group. To meet these goals, materials with abnormal partial dispersion are commonly used in this group.
- (3) Rear group: This group typically consists of a meniscus lens or its variation. It is located close to the tube lens or the intermediate image and generally after the aperture stop. The rear group is used to correct field curvature, astigmatism, and chromatic aberrations. This group may not be necessary, depending on the requirements for controlling field curvature and chromatic aberration.

In real microscope objective lenses, the separation between each group may not be clear. Each group will introduce all five monochromatic and two chromatic aberrations and contribute to the correction of all aberrations.

Microscope objective lenses are usually designed as telecentric in object space, which means that the aperture stop is located at the rear focal point and the entrance pupil is at infinity. In a telecentric objective lens, the chief rays are parallel to the optical axis. For example, the objective lens in Fig. 4.8(a) is made telecentric in object space by placing the aperture stop at the real focal plane. Figure 4.8(b) shows the ray path in the front group. The off-axis chief ray is parallel to the optical axis.

Telecentricity is necessary to achieve uniform resolution across the entire FOV. For a reflectance microscope with epi-illumination whose illumination and imaging systems share the same light path, telecentricity prevents vignetting and provides uniform conditions for resolution. Also, telecentricity minimizes the change in magnification of the system and the change in FOV during defocusing. In transillumination mode, the condenser should be telecentric in object space to match the telecentricity of the objective lens. If the condenser is not telecentric,

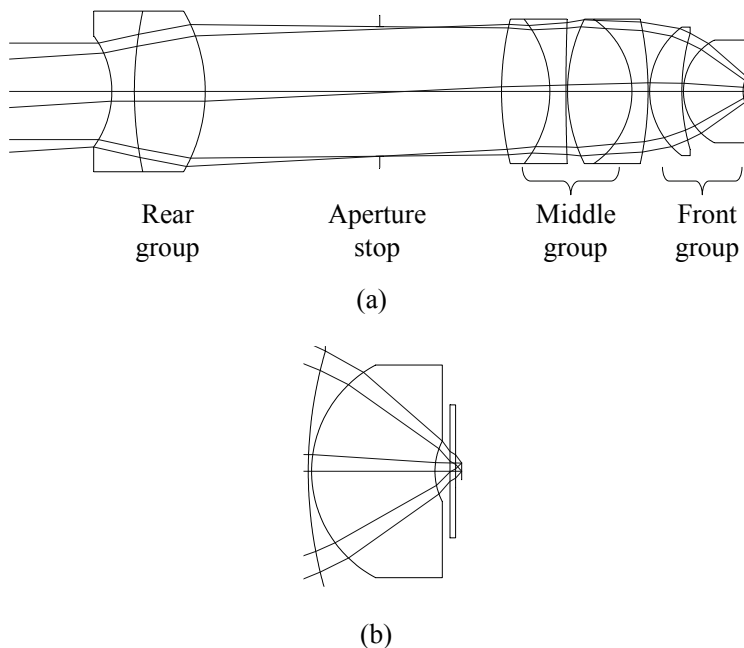


Figure 4.8 (a) Basic configuration of a high-performance microscope objective lens. (b) Ray path in the front group.

its NA should be large enough that the pupil of the objective lens is completely filled.

The exit pupil is usually located inside the objective lens. Its position inside the lens depends solely on the individual design of the microscope objective lens. The size of the pupil decreases as the magnification increases.

4.2.2 Aberration characteristics of objective lenses

In order to obtain diffraction-limited performance, it is important to achieve a high level of aberration correction. As discussed in Sec. 4.1.3.1, according to the correction degree of spherical aberration, field curvature, and chromatic aberration, objective lenses are classified into several categories: achromat, plan achromat, fluorite, plan fluorite, apochromat, and plan apochromat.

4.2.2.1 Spherical aberration

Spherical aberration is due to the different refractions of the light path between the center of the lens and its periphery. Light passing near the center of the lens is refracted only slightly, whereas light passing near the periphery is refracted to a greater degree. As a result, the peripheral light rays come into a shorter focus closer to the back focal plane of the objective lens than do the rays traveling through the central region of the lens.

Spherical aberration is the most serious aberration in microscope objective lenses because it causes the object to be hazy or blurred and out of focus. It has a direct impact on the resolution of the objective lens, which seriously affects the object's sharpness and clarity.

Besides that which is generated by the lens, spherical aberration can be introduced by the improper use of the tube lens, cover glass, and immersion medium. As the NA increases, the spherical aberration generated by those components becomes severe and must be minimized.

4.2.2.2 Coma

Coma is a variation of image location and size with the zonal radius in the pupil and produces a comet-like blur. Coma is similar to spherical aberration, but it only affects off-axis objects. The magnitude of coma is proportional to the field angle.

Coma is most severe when a lens is not properly aligned; coma is considered one of the most problematic aberrations because each image point formed by the lens with coma is asymmetric. This aberration must be corrected for high-performance objective lenses.

4.2.2.3 Field curvature and astigmatism

Field curvature is the natural result of using lenses that have curved surfaces. A lens with field curvature will image points from an extended flat object onto a curved surface.

An objective lens with field curvature produces an image that is out of focus in the circumferential section and in focus in the center, or vice versa. The objective lens must be continuously refocused to obtain detailed information between the center and the edge.

Field curvature may not be an issue when the user observes the sample directly by focusing the object at a different region, but it can be a serious problem for digital imaging, given that the detector is flat. For some applications in biomedical imaging, especially in tissue imaging, field curvature correction is not as critical because the object is not flat.

Generally, correction for field curvature adds a considerable number of lens elements to the objective lens. One direct impact of field correction is the sacrifice of a considerable amount of free working distance.

Astigmatism occurs because light rays in the tangential and sagittal planes of the lens are refracted differently. A lens with astigmatism will image an off-axis point to a series of elongated images ranging from linear to elliptical, depending on the location of the image plane. Corrections of astigmatism and field curvature are generally accomplished together.

4.2.2.4 Distortion

Distortion occurs because the real ray bends more than the paraxial ray. Distortion is a change in magnification as a function of the FOV and does not

blur the image. Most objective lenses designed for use in biological microscopes can have a distortion of up to 1–2%. For visual observation, a distortion of 2–3% is acceptable. Compared with other lenses, for example, a photographic lens, this acceptable distortion is fairly large against the FOV.

4.2.2.5 Chromatic aberrations

Chromatic aberrations come from the dispersion of optical materials. Axial chromatic aberration is the longitudinal variation of focus, and lateral chromatic aberration is the variation of image size with wavelength. An objective lens with chromatic aberrations cannot bring all of the light within the working spectrum onto a common focus, resulting in colored fringes surrounding the image.

In microscope objective lenses with large magnifications, the lateral chromatic aberration is difficult to correct, although the remaining aberration is not very large. Several approaches have been used to overcome this problem. A straightforward approach is to correct the lateral chromatic aberration independently for the objective lens and eyepiece; this correction for either can be fairly complicated.

For finitely corrected objective lenses, the remaining lateral chromatic aberration is often compensated for by the eyepiece. To compensate for the lateral chromatic aberration, the eyepiece should produce the same amount as but opposite sign of the lateral chromatic aberration from the objective lens. Typically, a lateral chromatic aberration of $\sim 1.5\%$ is acceptable for an eyepiece to compensate. For infinity-corrected objective lenses, the lateral chromatic aberration is generally compensated by the tube lens. This approach has been used by some manufacturers, such as Nikon and Zeiss.

4.2.3 Design techniques

Almost all lens design techniques for controlling aberrations can be applied in microscope objective lens design. These techniques include but are not limited to lens bending, power splitting, power combination, cemented surfaces, aplanatic surfaces, aspherical surfaces, and field lenses. This section will discuss three techniques commonly used in designing high-performance microscope objective lenses.

4.2.3.1 Aplanatic surfaces

As discussed in Sec. 2.7.4, a surface is free from spherical aberration, coma, and astigmatism when one of the following conditions is satisfied.³

- (1) $h = 0$: The object and image are both located at the surface.
- (2) $A = 0$: Both the object and image are at the center of curvature.
- (3) $\delta(u/n) = 0$: The object is at the aplanatic point.

From the paraxial ray trace of an aplanatic surface in Fig. 4.9, we have the object and image distances:

$$l = \frac{n' + n}{n} R, \tag{4.9a}$$

$$l' = \frac{n' + n}{n'} R. \tag{4.9b}$$

Parameters l and l' have the same sign, which means that an aplanatic surface cannot form a real image of a real object—either the object or the image must be virtual.

Meniscus lenses meeting aplanatic conditions are used in microscope objective lenses to enlarge the NA; each additional meniscus surface increases the NA by a factor of the refractive index. The principle of this method is shown in Fig. 4.10(a). The parameters of the meniscus lens are calculated in the following procedure.

The object distance l_1 of surface 1 is determined by the image distance l_0 of the previous surface and the air space t_0 between them:

$$l_1 = l_0 - t_0. \tag{4.10a}$$

To meet aplanatic condition 3, radius r_1 should be

$$r_1 = \left(\frac{1}{n + 1} \right) l_1, \tag{4.10b}$$

and the image distance is

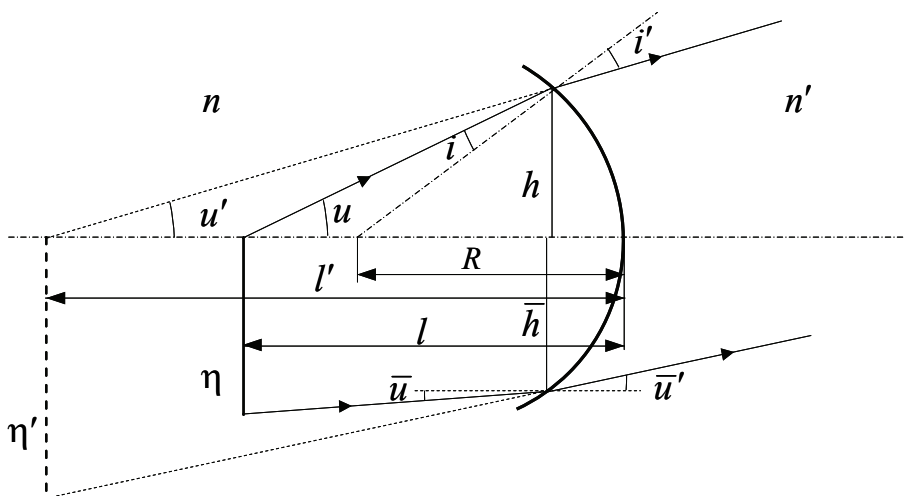
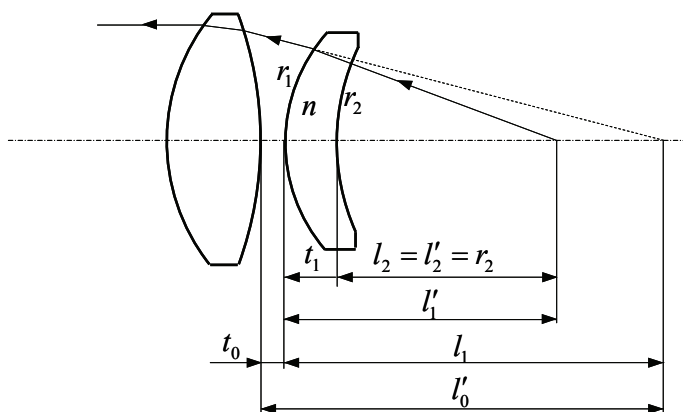
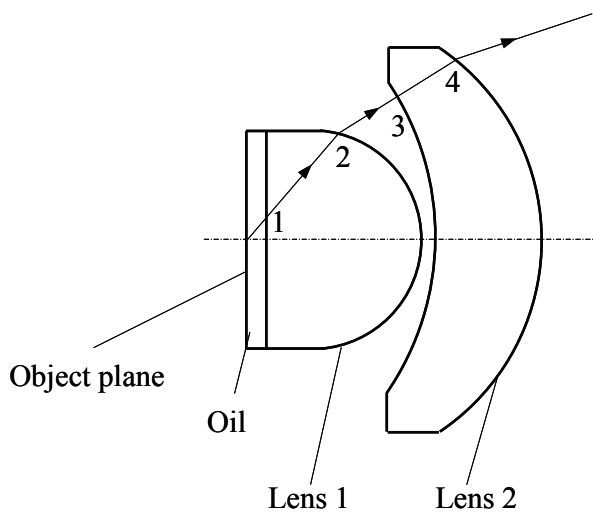


Figure 4.9 Principle of the aplanatic condition.



(a)



(b)

Figure 4.10 Principle of adding an aplanatic lens: (a) dry objective lens and (b) oil-immersion objective lens.

$$l'_1 = \left(\frac{n+1}{n} \right) r_1 = \frac{1}{n} l_1. \quad (4.10c)$$

The object distance l_2 of surface 2 is

$$l_2 = l'_1 - t_1. \quad (4.10d)$$

If surface 2 meets aplanatic condition 2,

$$r_2 = l_2. \quad (4.10e)$$

Surface 2 can also be designed to meet aplanatic condition 3. An additional aplanatic concentric meniscus lens can be added, following the same procedure, until the desired NA is met.

4.2.3.2 Liquid immersion

The primary purpose of an immersion medium is to increase the NA and, therefore, improve the resolution. For the objective lens designed with a cover glass, the immersion medium also helps to extract light out of the cover glass because the TIR condition is no longer satisfied when the immersion medium touches the cover glass.

The immersion liquid has two more advantages when a microscope is used to image an object through the cover glass or through tissue: it reduces the spherical aberration caused by the thickness variation of the cover glass, and it removes the specular reflection when the microscope works in reflectance mode. By using an immersion medium with a refractive index that is the same as that of the cover glass, image degradation due to the thickness variation of the cover glass is eliminated because there is no refraction event for all of the rays from the object. If the refractive index of the immersion liquid does not exactly match that of the cover glass, the change in cover glass thickness will still introduce spherical aberration. The induced spherical aberration is linearly proportional to the change in thickness and inversely proportional to the wavelength.³

The aplanatic condition is used in a classical oil-immersion microscope objective lens. As shown in Fig. 4.10(b), the second and fourth surfaces are aplanatic, meeting aplanatic condition 3. Surface 3 is also aplanatic with an object and an image at the center of the curvature. The rays pass through this surface without deviation.

By optimizing the lenses in the front group to meet the aplanatic conditions, they are free of spherical aberration and coma. In a real objective lens, there are some deviations from the perfect aplanatic condition, resulting in some spherical aberration.

4.2.3.3 Correction of field curvature

If there is no special effort to flatten the field, the objective lens has a Petzval sum of approximately $1/f$. To reduce the field curvature, we can either minimize the contribution from the lens surface or compensate for the field curvature using the lens in the rear group. The front group of the objective lens contributes the majority of the positive refractive power. Therefore, it is preferable to use optical materials with high refractive indices that enlarge the radii of the lens surfaces and therefore reduce the Petzval contribution.

In 1938, Boegehold introduced microscopic objective lenses with thick meniscus lenses for field curvature correction.^{3,4} As discussed in Sec. 2.4.4, according to the Petzval theorem, every surface between two media with different refractive indices contributes to the Petzval curvature:

$$\frac{1}{r_p} = -n'_m \sum_i \frac{n'_i - n_i}{n'_i n_i r_i}, \quad (4.11)$$

where n_i and n'_i are the refractive indices of the media at the object side and image side of surface i with a radius of r_i , and m is the number of surfaces. For a thin lens in the air, the Petzval curvature is

$$\frac{1}{r_p} = -\sum_i \frac{1}{n_i f_i}, \quad (4.12)$$

where f_i is the focal length of i th lens. The refractive power of a thick lens is

$$\Phi = \frac{1}{f} = (n-1) \left(\frac{1}{r_1} - \frac{1}{r_2} \right) + \frac{(n-1)^2 d}{nr_1 r_2}. \quad (4.13)$$

If the lens has the same radius on both surfaces, the power of each surface is the negative of the other, so the Petzval sum is zero. The power is always positive if the radius is not zero and is a function of the lens thickness:

$$\Phi = \frac{(n-1)^2 d}{nr_1^2}. \quad (4.14)$$

Changing the lens thickness will change the refractive power, but it will not change the Petzval sum of the lenses if the lens curvatures are fixed. Therefore, we can either design a lens with positive power and a zero Petzval sum or design a lens with zero power and a negative Petzval sum to reduce the Petzval sum of a normal lens with a positive power. The disadvantage is that the lens must be thick, making it difficult to manufacture.⁵

Figure 4.11(a) shows one example of using a thick meniscus to flatten the field. Generally, a thick meniscus is placed after the aperture stop. Thick meniscus lenses in the rear lens group of an objective lens can reduce the Petzval sum and, therefore, reduce the field curvature. The greater the distance between the meniscus lens and the positive refractive group of the objective lens, the more effective this method will be in reducing the field curvature. Thick meniscus lenses also introduce astigmatism; thus, a combined correction is necessary. The effect of this method depends on the surface radius, the thickness, and the refractive index of the meniscus lens.

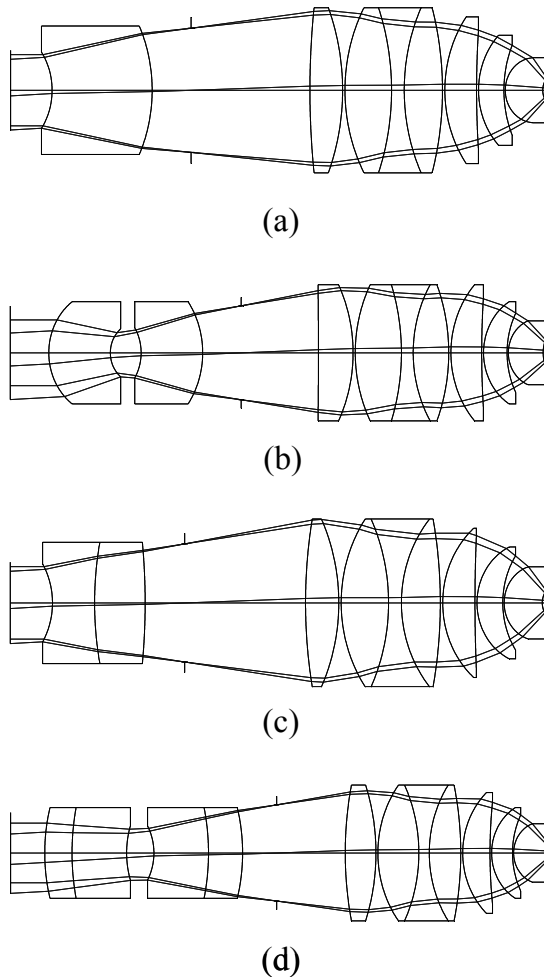


Figure 4.11 Configurations of a field-flattening lens group in a microscope objective lens. (a) Single meniscus, (b) double meniscus, (c) achromatized meniscus, and (d) double achromatized meniscus.

To reduce the field curvature without disturbing the infinite conjugate of the objective lens too much and to achieve better correction of the field curvature, an afocal flattening lens group with two thick meniscus lenses can be added to the lens, as shown in Fig. 4.11(b). To improve chromatic aberration correction, the meniscus is replaced with a cemented achromatic doublet, as shown in Figs. 4.11(c) and 4.11(d).

Generally, to correct the field curvature more effectively, the positive lenses should have high refractive indices and the negative lens should have a low refractive index.

According to the degree of field flattening, objective lenses can be classified as standard, semi-plan, or plan. Standard objective lenses without field curvature

correction have diffraction-limited performance only near the optical axis. Semi-plan objective lenses should be diffraction limited over 60% of the field, and plan objective lenses should be diffraction limited over 95% of the FOV. Figure 4.12 illustrates the difference in the Strehl ratio as a function of the field for three types of objective lenses.

Uncorrected field curvature in the semi-apochromat and apochromat objective lenses is one of the most severe aberrations. The region of interest must be continuously refocused for detailed information. Correction for field curvature adds a considerable number of lens elements to the objective lens and sacrifices a considerable amount of free working distance. Some of the high-magnification objective lenses have a concave front lens to minimize the field curvature, making it more difficult to clean and maintain the front lens.

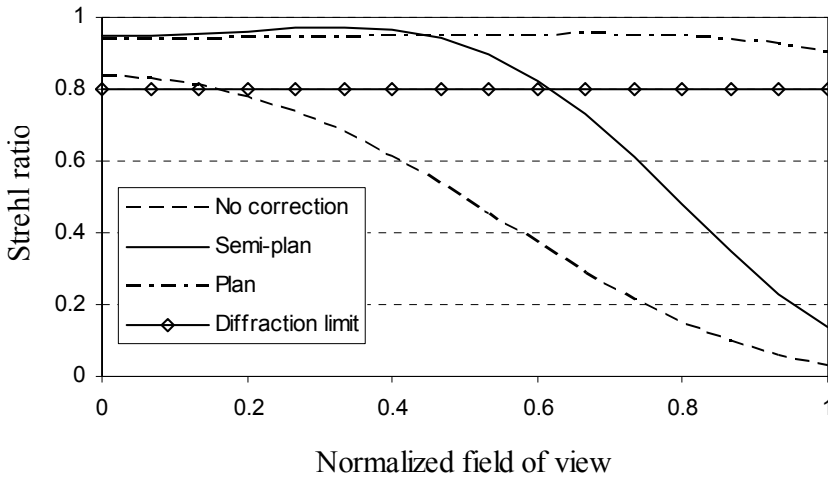


Figure 4.12 Strehl ratios of standard, semi-plan, and plan objective lenses.

4.2.4 Objective lens design

When designing a microscope objective lens, one usually starts at the image plane of the objective lens because it is generally much more convenient to design an optical system with the object at the long conjugate. The performance is calculated in object space. Theoretically, there is no difference if ray tracing starts from object space. However, ray tracing is more stable when starting from the image side of the objective lens because small changes in the front group may cause ray failure in the rear group.

There are several possible approaches in designing an objective lens. The fundamental method is to start with a simple doublet and then add doublets and aplanatic lenses to increase NA and FOV. Additional lenses, such as singlets, doublets, triplets, and thick meniscus lenses, are then added to correct

aberrations. Another approach is to start with a prior design that is similar in form and performance and then optimize it to meet the specifications.

For a microscope objective lens with low NA and low magnification, a simple achromat lens is generally used. For low and medium NA and magnification, a Lister objective lens, which consists of two achromatic doublets, is sufficient. To further increase the NA and magnification, one or more aplanatic elements are added in the front lens group, creating an Amici objective lens.

4.2.4.1 Low-power achromats

Achromats are microscope objective lenses with a limited spectral correction and a moderate FOV. They only correct chromatic aberration in two wavelengths and spherical aberration in one wavelength.

A doublet is ideal for an objective lens with a low NA (<0.1) and low magnification (<5). The Abbe numbers of the two materials used in a doublet should be different enough that the ratio is not close to 1. The larger the difference between the two Abbe numbers, the smaller the refractive power of the two elements; therefore, the lenses will have large radii and will be easy to manufacture. Typically, the performance of an achromat is limited by the uncorrected Petzval curvature and astigmatism.

Figure 4.13 shows an example of a $5\times/0.1$ microscope objective lens. The first element is BK7 glass, and the second is SF2 glass. The ratio of Abbe numbers is 1.9. As seen from the plot of RMS spot radius as a function of the field, the spot radius is smaller than the Airy disk for up to 70% of the field. The lateral chromatic aberration is well corrected, but the axial chromatic aberration is close to the depth of field and needs some improvement. The field curvature and coma limit the off-axis performance.

Sometimes, as shown in Fig. 4.14, low-power objective lenses have an additional meniscus element to reduce the Petzval curvature. A meniscus lens close to the object can significantly reduce field curvature. The Petzval radius is 1210 mm compared to 35.85 mm of the objective lens in Fig. 4.13. This objective lens has a diffraction-limited performance over the entire FOV for the d line (587.6 nm). Compared to the objective lens in Fig. 4.13, except for the astigmatism and axial chromatic aberration, all other aberrations, especially field curvature, are reduced. The drawback of this design is that the working distance is shorter.

4.2.4.2 Medium-power achromats

For a medium-power objective lens, a Lister lens is usually used. A Lister lens consists of a pair of widely separated doublets. Figure 4.15 shows an example of a $10\times/0.25$ microscope lens with a focal length of 16.5 mm and an FOV of 2 mm.

Typically, the separation between the two doublets is equal to the focal length of the doublet closer to the object, and the stop is at the plane of the first doublet so that the objective lens is telecentric in object space. The negative lens

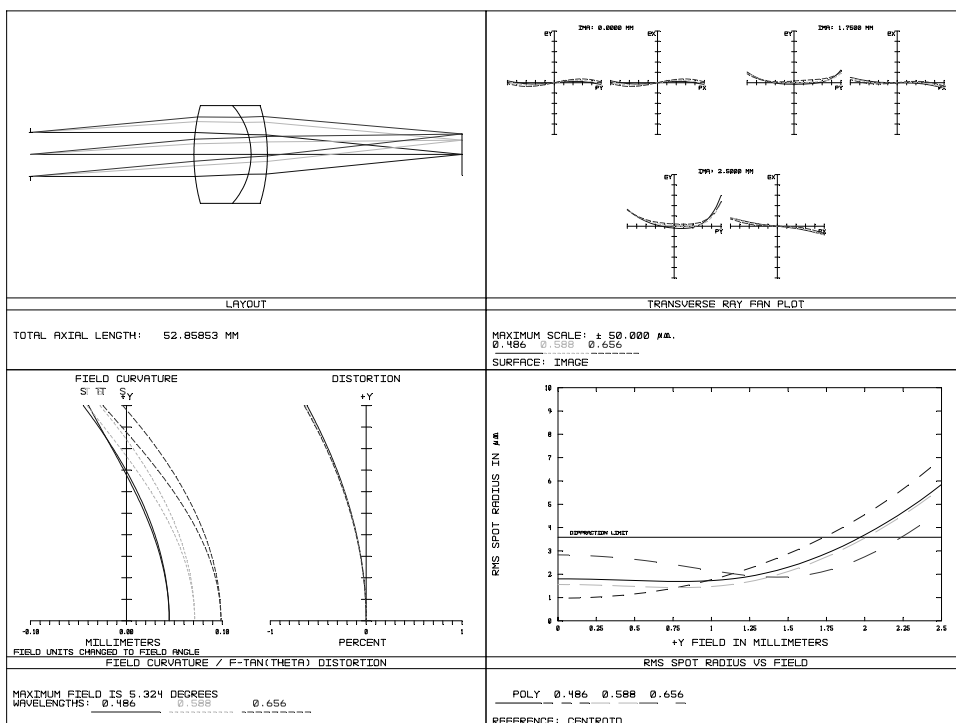


Figure 4.13 Layout and performance of an achromat microscope objective lens 5x/0.1.

of each doublet is on the object side, and the surface is nearly flat. Both doublets are corrected for spherical aberration and coma. The two doublets have nearly equal bending power so that the marginal ray has the same deflection angle for each doublet. The separation between the two doublets allows for partial correction of the field curvature with the proper amount of astigmatism in the cemented interface.

The plot of the Strehl ratio as a function of the FOV shows that this example is mainly corrected for the short wavelength. Performance drops quickly as the wavelength becomes longer than 600 nm, and it also decreases significantly from the optical axis to the edge of the field. Spherical aberration and field curvature are the two limiting aberrations.

4.2.4.3 High-power achromats

To meet the requirements of high-power objective lenses, a nearly aplanatic hyperhemispheric thick lens is added to the object side of a Lister objective lens. This type of lens is called an Amici objective lens. Objective lenses with an NA up to 0.65 can be achieved with an Amici configuration. One of the drawbacks of the Amici objective lens is its short working distance.

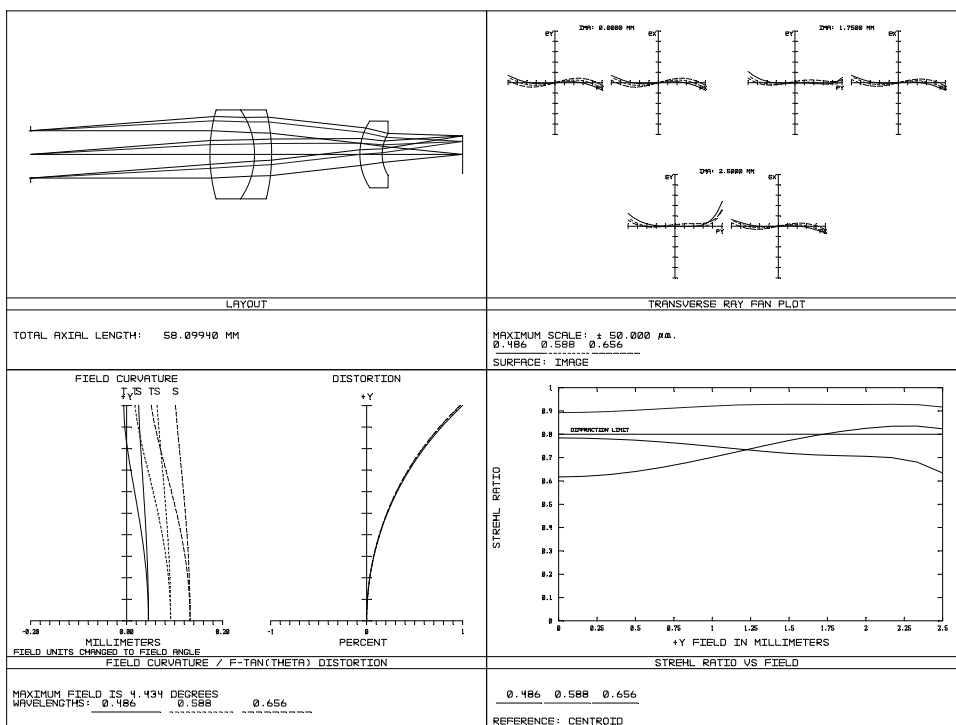


Figure 4.14 Layout and performance of an achromat microscope objective lens 5x/0.1 with an additional meniscus.

Typically, the front surface of the thick singlet is concave to reduce the Petzval sum; this surface introduces some spherical aberration because it is not in contact with the object. To compensate for the aberrations introduced by the first surface, the radius of the second spherical surface is slightly larger than the aplanatic solution. The distance between the thick singlet and the first doublet is small; otherwise, large lateral chromatic aberration would occur. The undercorrected spherical aberration introduced by the thick singlet is compensated by the other two doublets. The stop is usually placed at the plane of the last surface of the first doublet to achieve telecentricity in object space.

Figure 4.16 is the layout and performance of an Amici objective lens with an NA of 0.65 and magnification of 40. In this design, only one color (green) is diffraction limited on the axis, and the field curvature is not corrected, as can be seen in the plots of field curvature and Strehl ratio. There is also some lateral chromatic aberration, which is difficult to correct in such a simple objective lens. The lateral chromatic aberration should be compensated by the eyepiece or tube lens.

To further increase the NA, a second singlet, which has a nearly aplanatic-concentric meniscus shape, can be added between the first singlet and the first doublet. As will be seen later, this method, namely, adding one or more singlets,

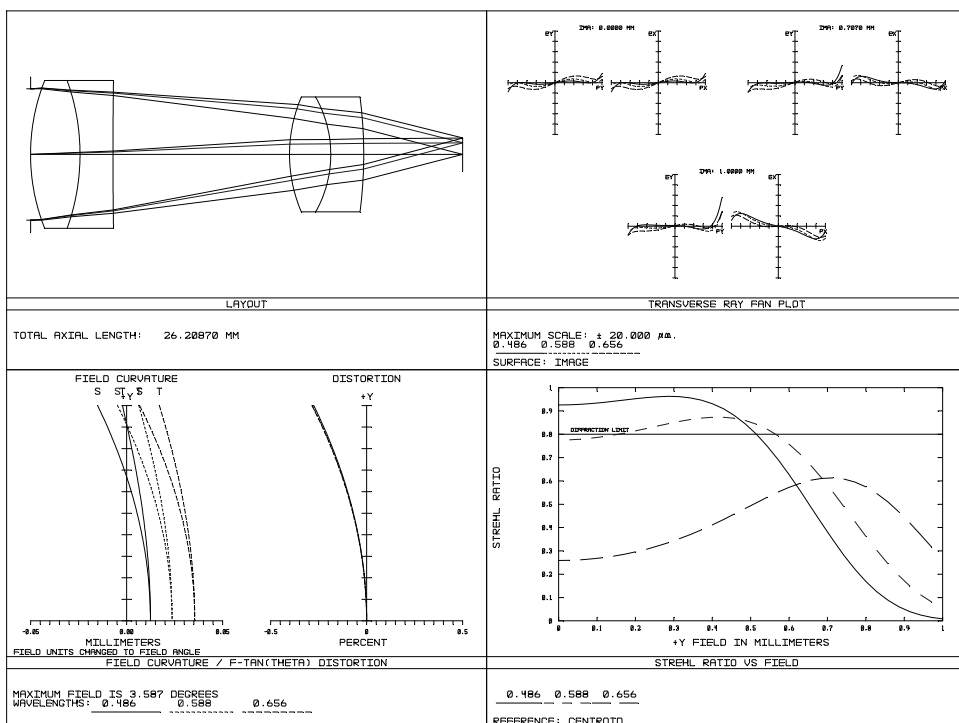


Figure 4.15 Layout and performance of a Lister microscope objective lens 10×/0.25.

is widely and effectively used in the design of high-performance objective lenses, such as oil-immersion objective lenses and plan objective lenses.

4.2.4.4 Oil-immersion objective lenses

To further increase NA, it is necessary to apply an immersion liquid between the first surface of the hyperhemispheric lens and the cover glass or the sample. The NA is increased by a factor of the refractive index of the immersion liquid. As a side benefit of using an immersion liquid, the reflection from the cover glass is minimized.

Figure 4.17 shows an oil-immersion objective lens 100×/1.0. A layer of oil is placed between the cover glass and the first lens surface of the objective lens. The refractive index of the immersion oil is 1.515, a value that nearly matches the refractive index of the cover glass. There are two cemented doublets with CaF_2 as the positive elements to correct the axial chromatic aberrations and the secondary spectrum. The first singlet is a plano-convex lens, which is easier to maintain and less expensive to manufacture. The second singlet is a nearly aplanatic-concentric meniscus lens that reduces the spherical aberration. The singlets convert a divergent cone of light with a large NA from the object point into a cone with a lower NA. This lens has a diffraction-limited

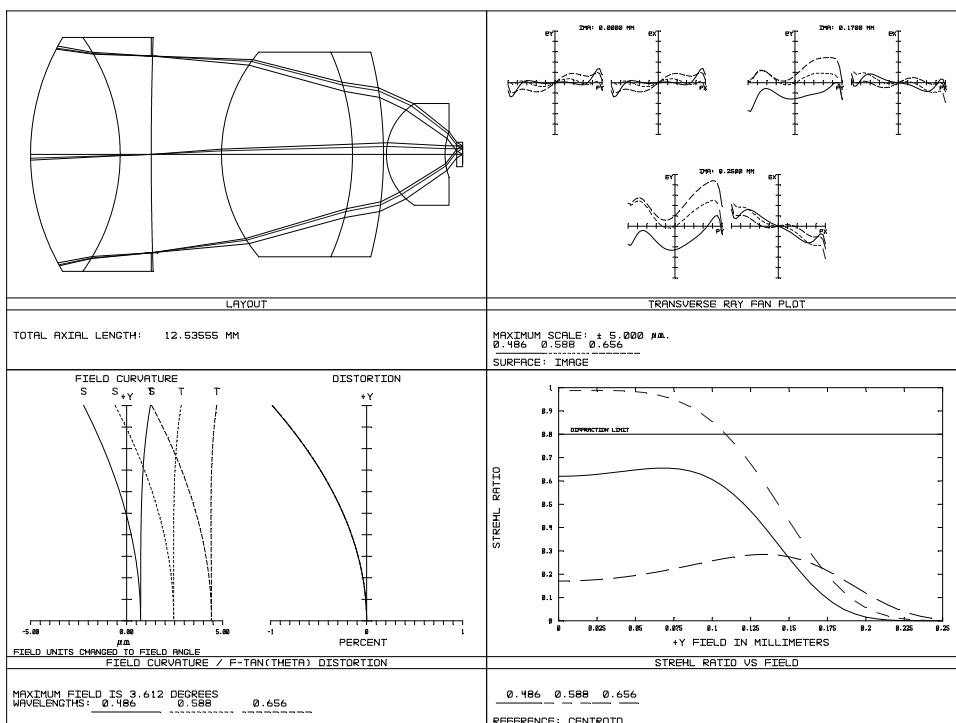


Figure 4.16 Layout and performance of an Amici microscope objective lens 40x/0.65.

performance within 40% of the FOV for a visible wavelength shorter than 620 nm. Astigmatism is well corrected, but the field curvature correction is not satisfactory.

4.2.4.5 Plan achromats

Small field curvature and astigmatism are very useful for a wide range of biological applications. Any application using a detector generally requires a flat-field objective lens.

As discussed earlier, the main positive power of a high-power microscope’s objective lens comes from the front group, which consists of singlets. To reduce the Petzval contribution, a thick meniscus lens with a high refractive index is generally used. However, a thick meniscus lens also introduces astigmatism.

Figure 4.18 illustrates an objective lens with enough correction for field curvature and chromatic aberrations. This lens has a similar format to that shown in Fig. 4.16, with a thick meniscus lens in front; however, it has an additional positive lens as the last element to compensate for the spherical aberration, field curvature, and lateral chromatic aberration generated in other elements. This objective lens is a plan fluorite, with a magnification of 10 and an NA of 0.32. It

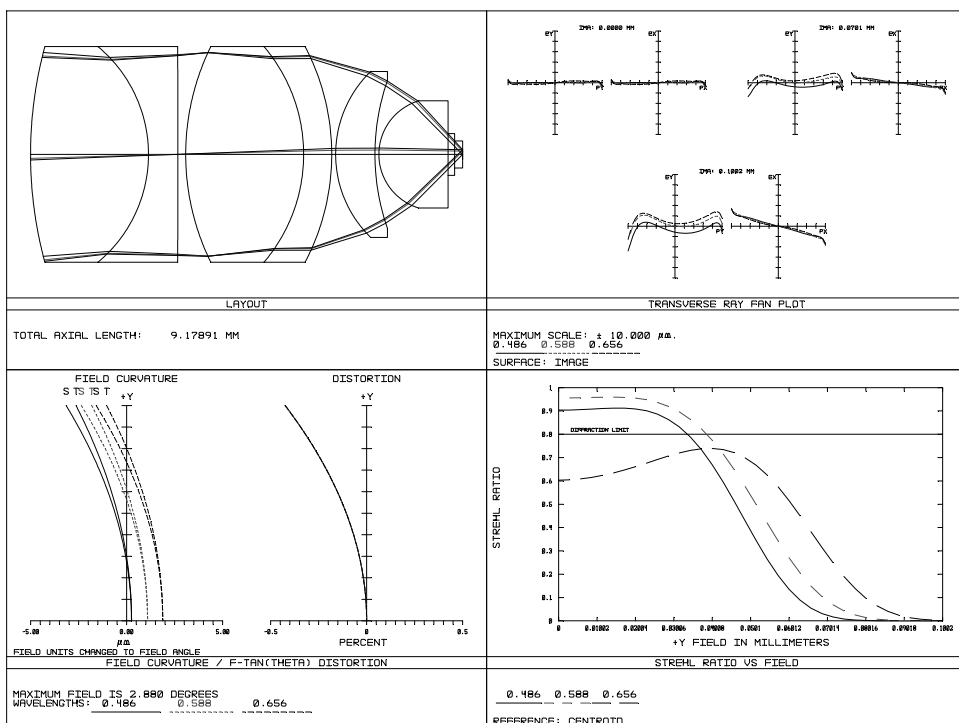


Figure 4.17 Layout and performance of an oil-immersion objective lens 100 \times /1.0.

has a diffraction-limited performance over the entire FOV for the visible spectrum. The lateral chromatic aberration is within the Airy disk, and the Petzval radius is 124 mm.

In order to improve the correction of the field curvature for a high-performance objective lens, it would be more effective to add a rear lens group consisting of a single thick meniscus lens, two meniscus lenses, a triplet, or their achromatic variations, which would better correct for chromatic aberration. In addition to correcting the field curvature, this lens group reduces the astigmatism.

Figure 4.19 shows an objective lens with a single thick meniscus lens in the rear group. It has an NA of 0.65 and a focal length of 4.0 mm. The objective lens is designed with a cover glass that introduces significant spherical aberration that must be corrected. Two singlets in the front group operate close to the aplanatic condition. Compared to the objective lenses in Figs. 4.16 and 4.17, the field curvature is reduced, and 70% of the field is flat for the short visible wavelength. However, the lateral chromatic aberration needs further improvement.

Figure 4.20 shows an objective lens with a doublet in the rear group to minimize the lateral chromatic aberration and to reduce the field curvature. It has the same NA and focal length as the objective lens in Fig. 4.19. The field curvature is well controlled across the entire FOV, and the lateral chromatic aberration is reduced by a factor of 4.

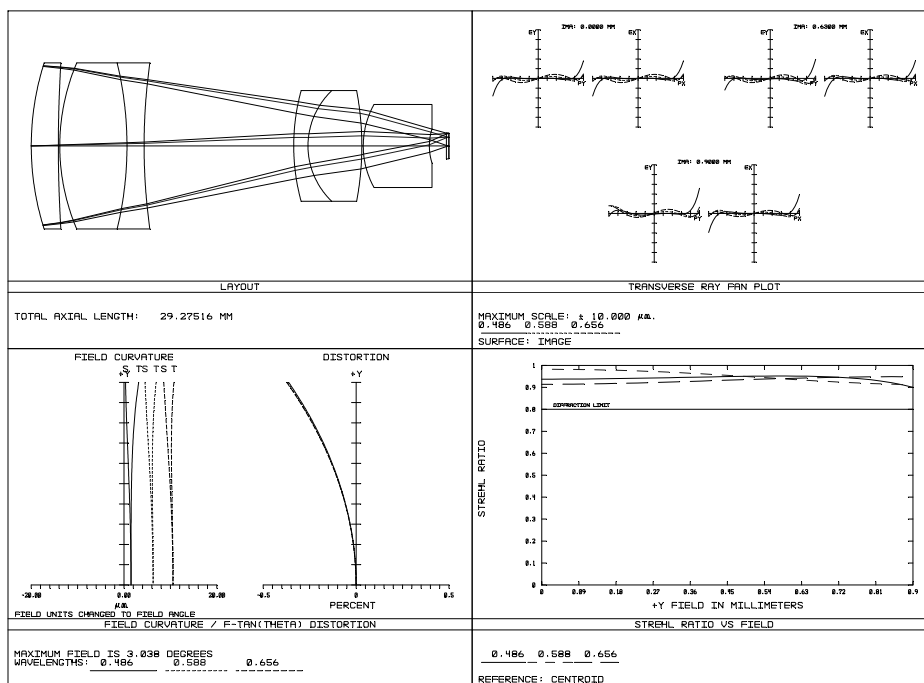


Figure 4.18 Layout and performance of a microscope objective lens 10 \times /0.32 with field curvature correction.

4.2.4.6 Plan apochromats

The basic design form for plan apochromat objective lenses is a plan achromat with one or two doublets replaced by a triplet comprising CaF₂ and low-index flint glasses. In addition, there are a few thick meniscus elements in the rear lens group to provide improved correction of axial secondary spectrum and Petzval curvature.

Figure 4.21 shows a plan apochromat lens with a thick meniscus achromat pair to correct the field curvature. The focal length is 2.0 mm, and the NA is 1.2. The front group consists of three aplanatic-concentric meniscus lenses to increase the NA. The middle group has three cemented lenses to correct the spherical aberration, coma, and chromatic aberration. The two inner face-to-face surfaces in the rear lens group introduce a significant negative field curvature to compensate for the positive field curvature from the front and middle groups.

4.2.4.7 Long-working-distance objective lenses

As can be seen from the examples discussed above, the working distance is quite small for a high-NA objective lens. The short working distance is favorable for correcting spherical aberration, field curvature, and chromatic aberration. Generally, the working distance decreases as the magnification and NA increase.

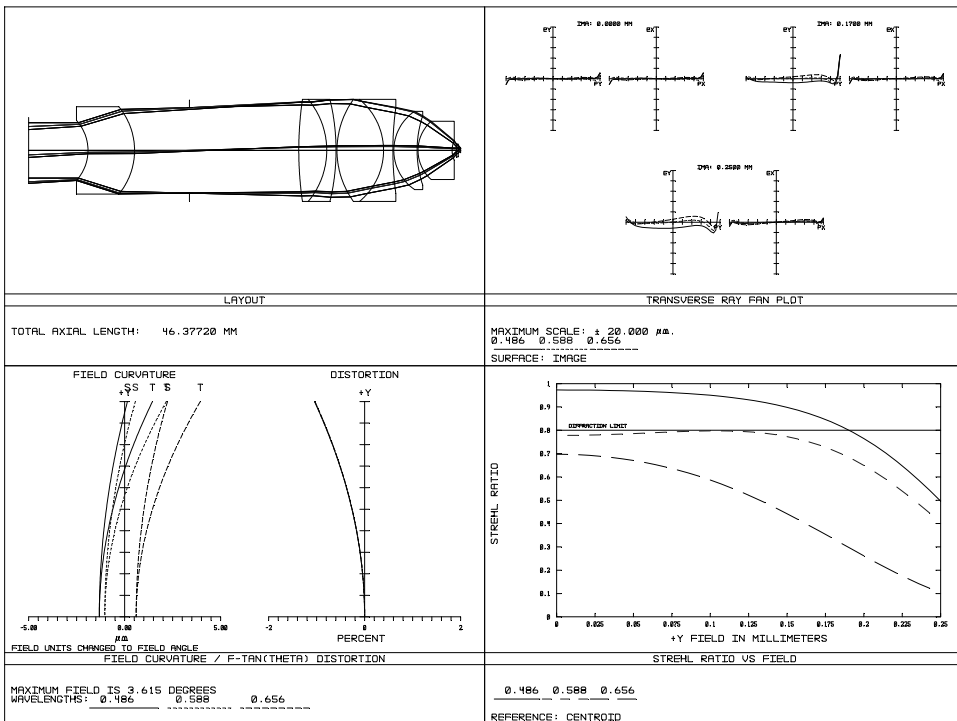


Figure 4.19 Layout and performance of a microscope objective lens with field curvature correction using a thick meniscus lens.

To maintain the NA and achieve a long working distance, the marginal ray bundle must have a large diameter. Thus, this type of lens has a relatively large front lens, and the correction of the spherical aberration becomes more challenging. In order to reduce the spherical aberration, the front group typically consists of several thin meniscus lenses.

Figure 4.22 is an objective lens 40×/0.55 with a working distance (3.8 mm) slightly shorter than the effective focal length of 4.5 mm, but much larger than those of the objective lenses discussed above.⁶ A double-Gaussian lens group with thick meniscuses is used in the rear group to flatten the field. It has diffraction-limited performance over the entire field for visible wavelengths shorter than 620 nm. This objective lens uses FK51, FK54, and KZFSN4 to reduce the secondary spectrum. As can be seen from the ray fan, the correction of the lateral chromatic aberration is not satisfactory. Compared to the objective lens in Fig. 4.15, which has the same NA and magnification, the lens diameter of this long-working-distance objective lens is almost double.

In order to meet the requirements of a small pupil diameter in the rear part of the system and a long working distance, a retrofocus lens is needed. Figure 4.23 shows a diagram of a retrofocus lens, which is a combination of a front negative lens group and a rear positive lens group, with the intention of shifting the focal

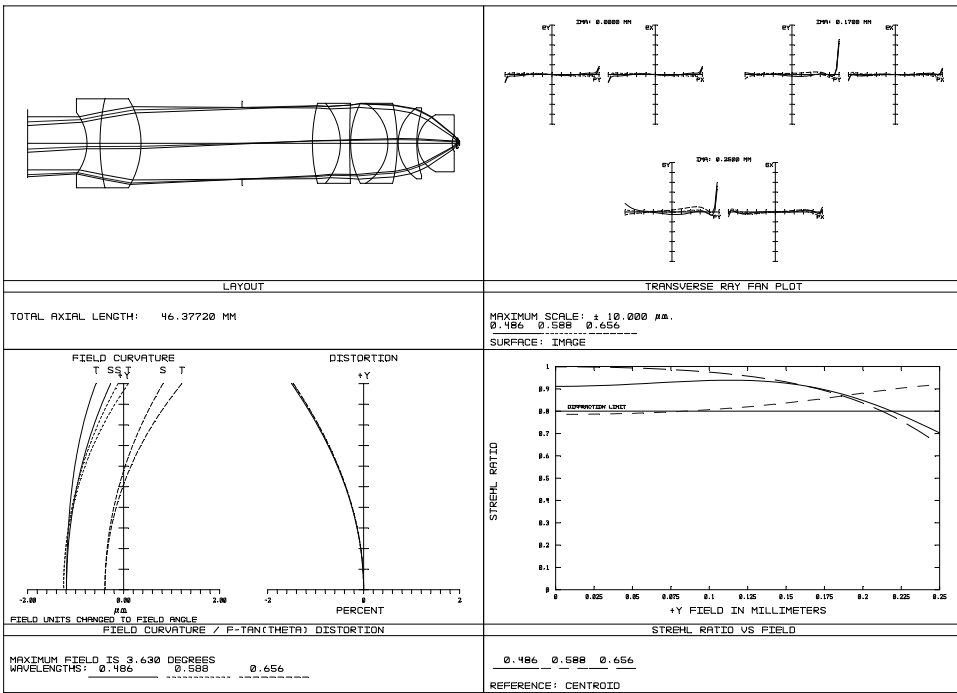


Figure 4.20 Layout and performance of a microscope objective lens with field curvature correction using a cemented doublet.

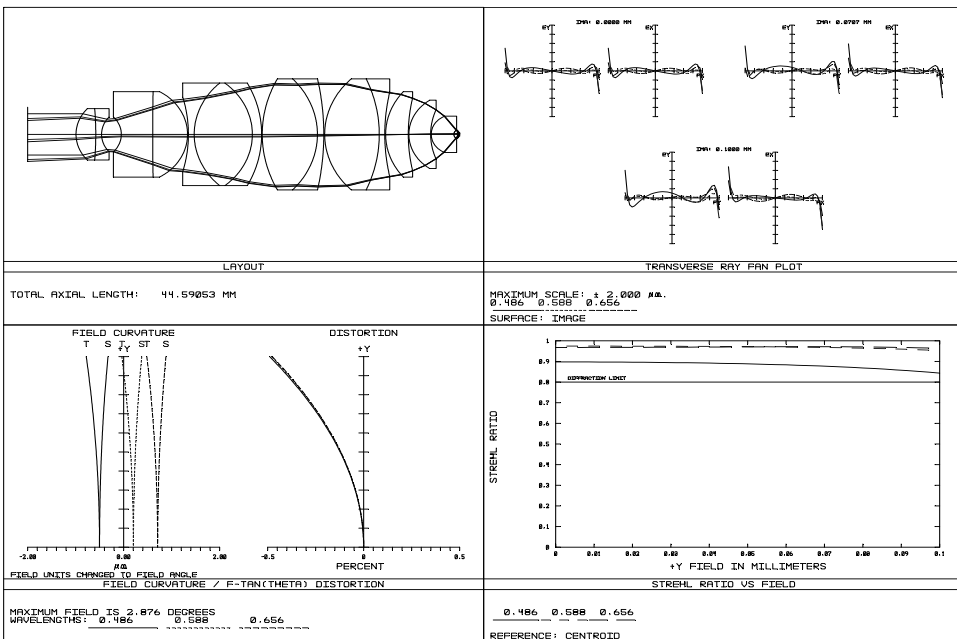


Figure 4.21 Layout and performance of a plan apochromat objective lens 100x/1.2.

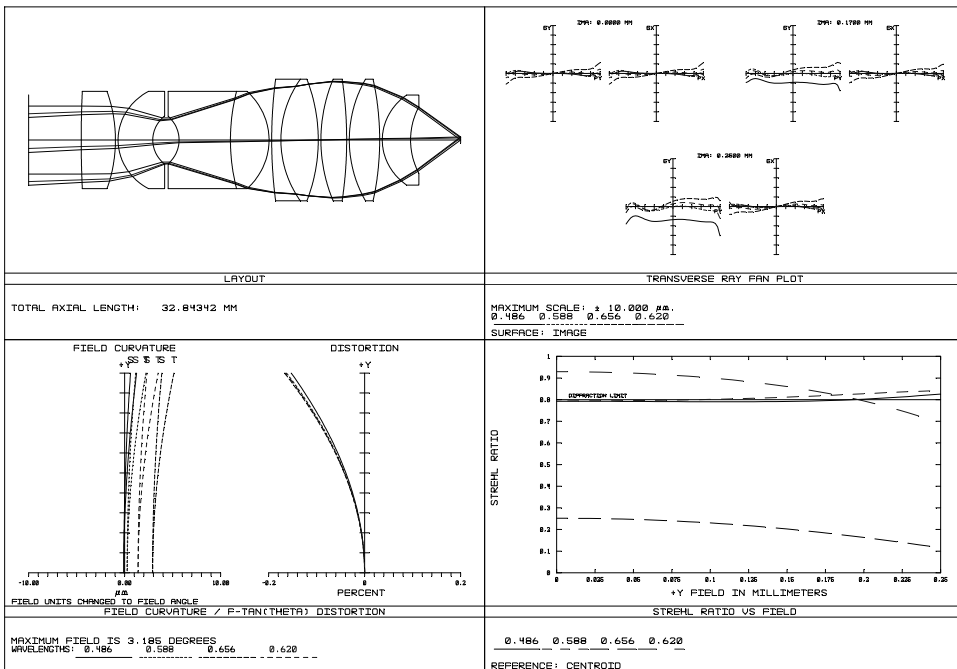


Figure 4.22 Layout and performance of a double-Gaussian, flat-field, long working distance microscope objective lens.

plane out of the lenses. The back focal length (BFL) is longer than the effective focal length (EFL). An extremely long BFL is possible; however, the Petzval field has a tendency to become strongly overcorrected. Using a retrofocus configuration, we can obtain an objective lens whose working distance is much larger than the focal length.

Figure 4.24 illustrates an objective lens $60\times/0.55$ whose working distance of 5.25 mm is longer than the focal length of 3.0 mm.⁶ This is a retrofocus lens with a negative lens separated from the positive lens group; this negative doublet also helps to flatten the field. CaF_2 and FK5 are used to correct the secondary spectrum, and the cemented triplet is used to reduce chromatic aberrations. The long separation between the negative and positive lens groups in this form also allows better correction of the lateral chromatic aberration.

As can be seen from the performance plots, this objective lens has diffraction-limited performance over the entire FOV across the visible spectrum. The lateral chromatic aberration is well under control, leaving spherical aberration as the main remaining aberration. The front lens group has two thin meniscus lenses that introduce a very small amount of spherical aberration and some axial chromatic aberration.

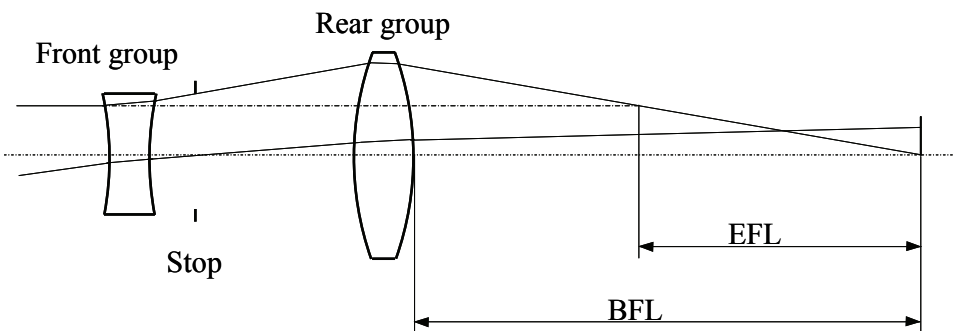


Figure 4.23 A retrofocus lens is characterized by a long back focal length.

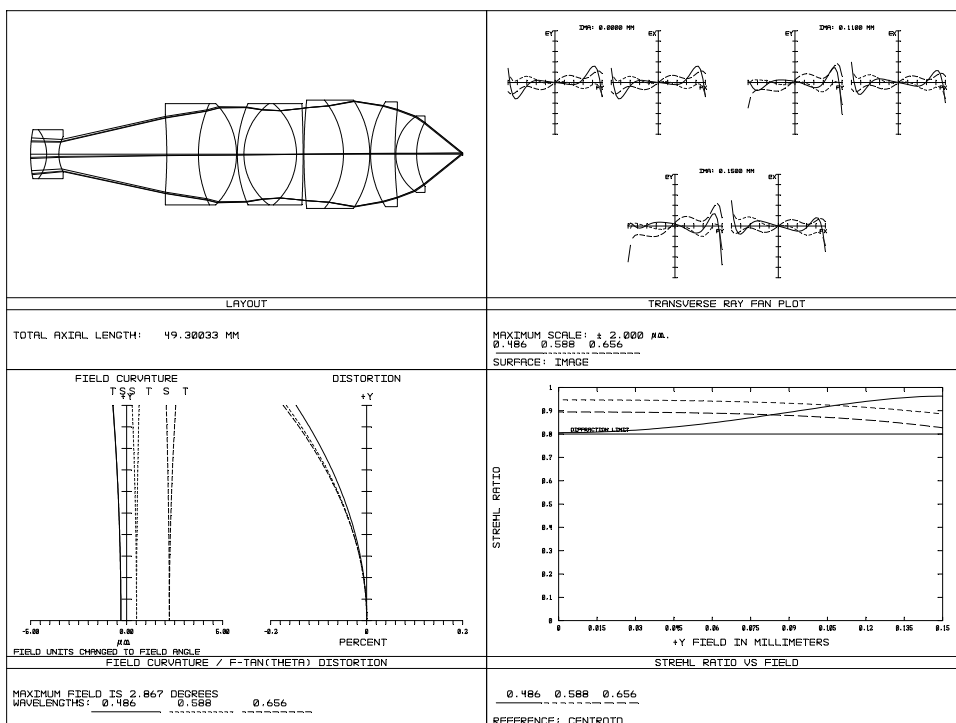


Figure 4.24 Layout and performance of a flat-field, long working distance microscope objective lens.

4.2.4.8 Objectives with correction lenses

Microscope objective lenses are generally designed with predefined parameters, such as the working distance, thickness of the cover glass, the refractive index of the immersion liquid, etc. In practice, these parameters may not be guaranteed. For example, the thickness of the cover glass can deviate from the nominal value,

and the immersion liquid may have a different refractive index. All of these deviations will induce spherical aberration.

A small thickness change of the cover glass becomes noticeable when the NA is above 0.35. When the NA is larger than 0.7, even extremely small deviations (± 0.01 mm) from the cover glass thickness have a significant effect on system performance.

The difference in the refractive indices of the immersion and mounting media will also induce spherical aberration. The more the refractive index of the immersion medium deviates from that of the mounting medium, the more spherical aberration is induced. For example, an objective lens that is designed for oil immersion ($n = 1.52$) will have considerable spherical aberration if the specimen is in a watery solution ($n = 1.33$).

In a practical microscope, the object detail to be observed may not contact the cover glass directly; it can be slightly below the cover glass. If the refractive index of the sample is different from that of the immersion liquid, this small distance will also introduce spherical aberration. The magnitude of the spherical aberration depends on the difference between the refractive indices as well as the distance between the cover glass and the observation plane.

In order to compensate for the spherical aberration induced by deviations in the refractive index and thickness of the cover glass as well as deviations in the refractive indices of the immersion and mounting media, a special objective lens with a movable lens group inside has been developed. The induced spherical aberration is minimized by moving the lens group with the help of a collar. However, this method is usually not applicable to objective lenses with very high NAs because the higher-order spherical aberration is difficult to compensate for.

Figure 4.25 illustrates an example of a correcting objective lens with an NA of 0.65 and a magnification of 60. The spherical aberration generated by a change in the thickness of the cover glass is minimized by moving an internal lens group along the optical axis. The movement of this group has a very small effect on coma and astigmatism.

For comparison, the Strehl ratio as a function of the field at 587.6 nm for a cover glass with thicknesses of 0.10–0.24 mm is plotted in Fig. 4.26(a). The objective lens maintains its diffraction-limited performance for a cover glass with a 0.17 ± 0.07 -mm thickness variation. The movement of the internal lens group as a function of the cover glass thickness is plotted in Fig. 4.26(b), which shows that it has an approximately linear dependence on the cover glass thickness.

4.2.4.9 Reflective objective lenses

As discussed above, chromatic aberration, especially lateral chromatic aberration, is difficult to correct in microscope objective lenses. For biomedical imaging systems working in a wide spectrum, correcting chromatic aberration with refractive elements makes the objective lens fairly complicated. An alternative approach is to use reflective elements because they are free of chromatic aberration.

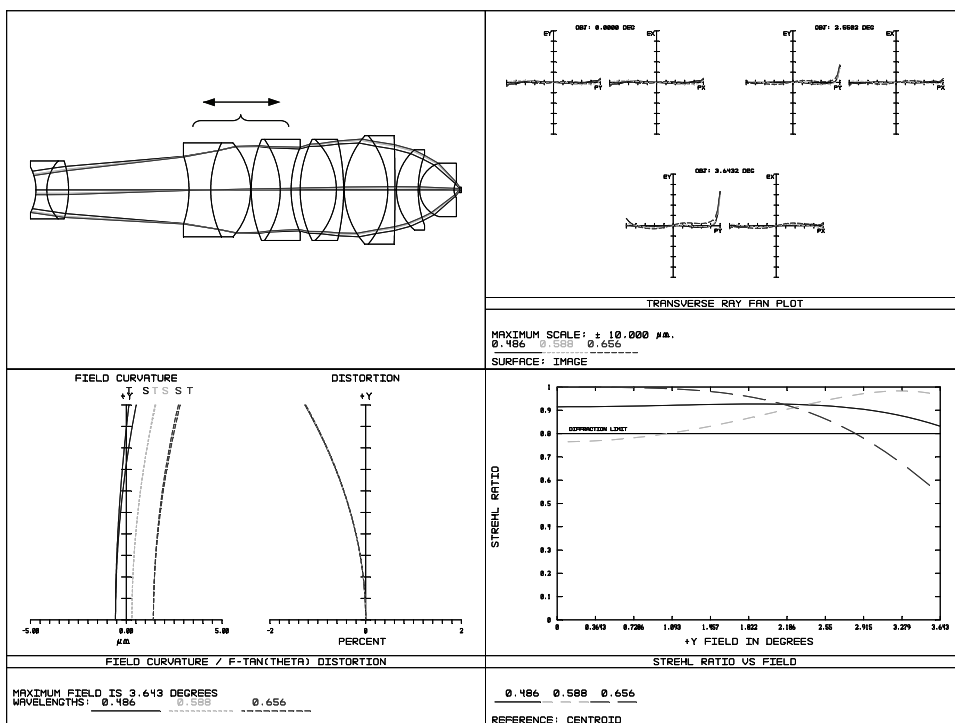


Figure 4.25 Layout and performance of a microscope objective lens with an internal movable lens group to compensate for the spherical aberration induced by the thickness variance of the cover glass.

Around 1904, Karl Schwarzschild invented a two-mirror objective lens that was free of spherical aberration, coma, and astigmatism. The system is formed by two concentric spherical mirrors, as shown in Fig. 4.27. The advantages of this objective lens include its inherent color correction, a long working distance, and the capability of completely correcting third-order spherical aberration, coma, and astigmatism. In practice, the design slightly deviates from the monocentric condition to balance third- and fifth-order aberrations. However, the Schwarzschild objective lens also has some limitations. The first limitation is that the system is obscured in the center, which increases the side lobes in the diffraction pattern and reduces lens performance; although the resolution is not decreased by a ring-shaped pupil, the energy throughput is reduced along with image contrast. The second limitation is a severely curved focal plane that limits the design to a small FOV.⁷

The Schwarzschild objective lens has been studied analytically based on third-order aberration theory in several publications.⁸⁻¹² Figure 4.27 shows the first-order parameters of a two-mirror objective lens. The two spherical mirrors are concentric with a common center of curvature. There are four independent parameters: two radii of curvature r_1 and r_2 , distance d between the two mirrors, and the working distance l or the imaging distance l' .

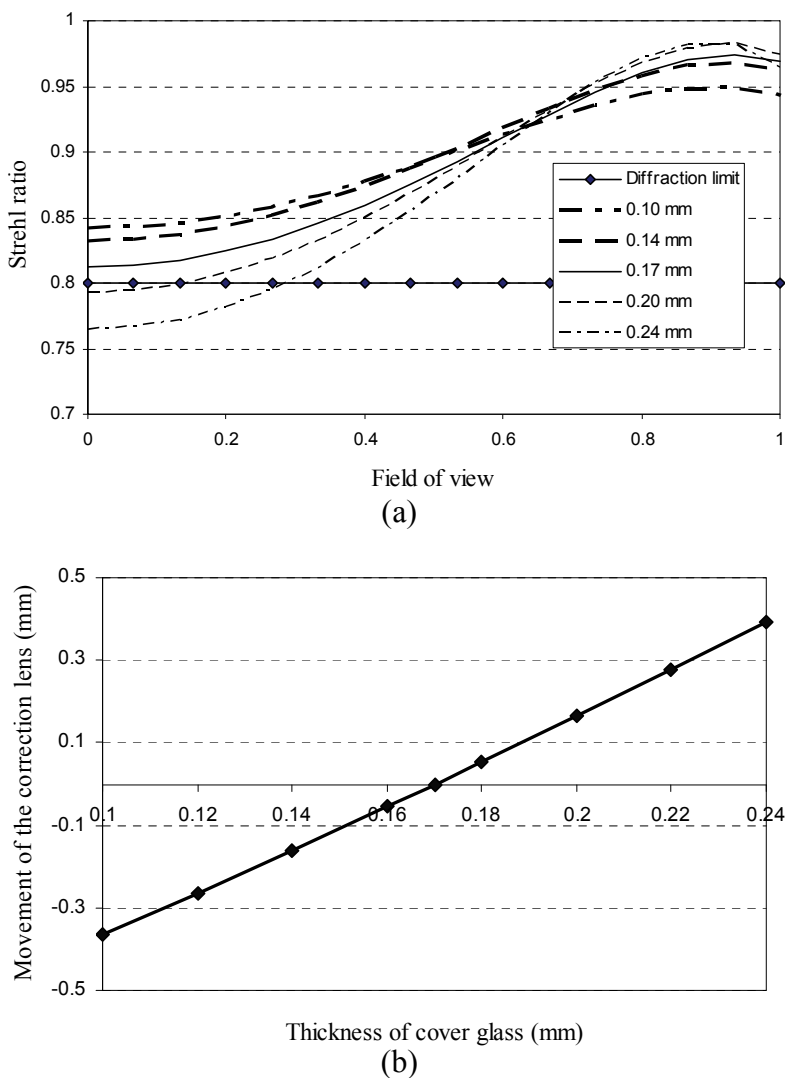


Figure 4.26 (a) Strehl ratios of the objective lens in Fig. 4.25 for different cover glass thicknesses after spherical aberration is compensated for by moving the internal elements. (b) The movement of the internal lens as a function of the cover glass thickness.

For an infinite imaging configuration with concentric mirrors, the third-order spherical aberration, coma, and astigmatism are given by the following equations:²

$$S_1 = \frac{1}{32M^4 f_1^3} \left[1 - \frac{(f_1 - d)(f_1 + f)(f_1 - f)^2}{f^3 f_1} \right], \quad (4.15a)$$

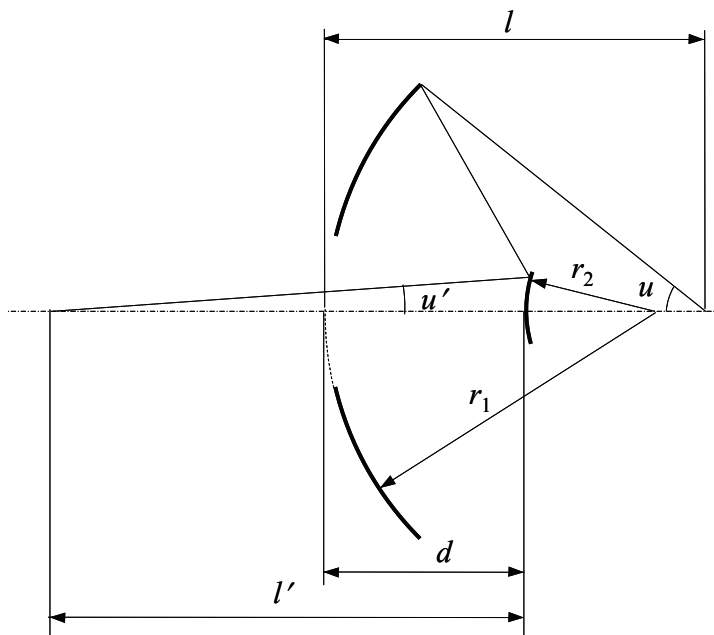


Figure 4.27 A Schwarzschild microscope lens with only two mirrors.

$$S_{II} = \frac{1}{4M^3 f^3} \left[1 + \frac{d(f_1 + f)(f_1 - f)^2}{2ff_1^3} \right], \quad (4.15b)$$

$$S_{III} = \frac{1}{2M^2 f^4 (f_1 - d)} \left[-f_1(f + d) - \frac{d^2(f_1 + f)^2(f_1 - f)^2}{4ff_1^2} \right], \quad (4.15c)$$

where

$$M = \frac{f(f_1 - d)}{f_1(f + d)} \quad (4.16)$$

and f_1 is the focal length of the first mirror.

The relationships between the focal lengths and the distance between the two concentric mirrors are

$$f_2 = \frac{f(f_1 - d)}{(f + f_1)}, \quad (4.17a)$$

$$d = \frac{2f_1^2}{f_1 - f}. \quad (4.17b)$$

Therefore,

$$S_I = \frac{1}{32M^4 f_1^3} \left[1 - \frac{(f - f_1)(f_1 + f)^2}{f^3} \right]. \quad (4.18)$$

To correct the spherical aberration, $S_I = 0$, we have

$$f = \frac{1 + \sqrt{5}}{2} f_1, \quad (4.19a)$$

$$d = -2f, \quad (4.19b)$$

$$f_2 = f + f_1, \quad (4.19c)$$

$$\frac{r_2}{r_1} = \frac{3 + \sqrt{5}}{2} = 2.618. \quad (4.19d)$$

From Eqs. (4.15b) and (4.15c), we have $S_{II} = S_{III} = 0$. Therefore, the system is free from spherical aberration, coma, and astigmatism.

For a finite image distance configuration with two concentric mirrors, the system is free from spherical aberration, coma, and astigmatism when the radius ratio of the two mirrors meets the following condition:⁷

$$\frac{r_2}{r_1} = \frac{3}{2} + \frac{r_2}{r_2 - l'} + \sqrt{\frac{5}{4} - \left(\frac{r_2}{r_2 - l'} \right)}. \quad (4.20)$$

The focal length is

$$f = \frac{1}{2} \left(\frac{r_1 r_2}{r_1 + r_2} \right), \quad (4.21)$$

and the Petzval sum is equal to

$$P = 2 \left(\frac{r_1 + r_2}{r_1 r_2} \right). \quad (4.22)$$

The radius of the curved image field is equal to the focal length.

Figure 4.28 is a Schwarzschild objective lens with the image at infinity. The parameters are calculated from Eq. (4.19a)–(4.19d). The focal length is $f = 2$ mm,

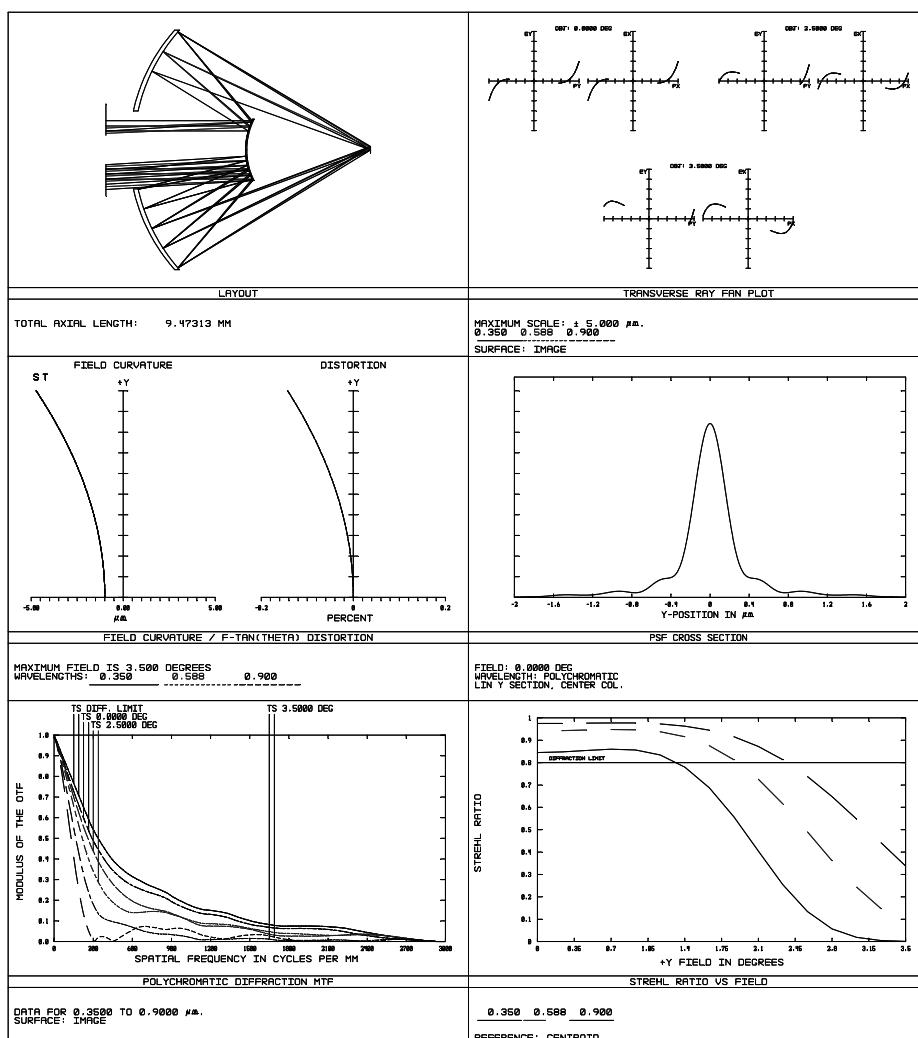


Figure 4.28 Layout of a Schwarzschild objective lens 10x/0.45 and its performance plots.

the NA is 0.45, and the FOV is ± 3.5 deg. The spectral coverage extends from 0.35 to 0.90 μm . The third-order spherical aberration, coma, and astigmatism are zero. The first-order parameters calculated from Eq. (4.19a)–(4.19d) correct only the third-order spherical aberration, coma, and astigmatism, so at least one of the mirrors has to be aspherical if higher-order aberrations need to be corrected. The residual aberration is field curvature, which is the radius of the focal length of the objective lens.

The ray tracing in the lens layout, ray fan plot, and MTF plot all show the effect of the central obscuration. Given that this is a mirror system, the performances are the same for all wavelengths at 0.35–0.90 μm . The Strehl ratio plot shows that the performance of the system degrades as the FOV increases

because of the uncorrected field curvature. The system is diffraction limited within 2.5 deg. The side lobes are clear in the point PSF plot.

Another limitation of the standard Schwarzschild objective lens is that its NA is limited to less than 0.6.⁷ In order to increase the FOV and the NA, many modified Schwarzschild objective lenses with additional refractive lenses have been developed.^{13–15} These lenses are usually called catadioptric lenses. A catadioptric objective lens uses the strengths of both reflective and refractive elements but avoids their disadvantages. Typically, in a catadioptric design, reflective elements are used to generate power, and refractive elements are used to correct third- and higher-order aberrations.

4.3 Tube Lens

A tube lens is necessary to form an intermediate image for an infinity-corrected objective lens. In some cases, the tube lens is also designed to compensate for residual aberrations, mainly the lateral chromatic aberration from the objective lens. The focal length of the tube lens is usually standardized between 160 and 200 mm, depending on the manufacturer. For a custom digital microscope, there is no restriction on the focal length; it is generally determined by the size of the detector.

The location of the tube lens with respect to the objective lens is one of the key considerations in designing a tube lens. If the tube lens is too close to the objective lens, space for the auxiliary optical elements is limited. However, if the tube lens is too far from the objective lens, light collected by the tube lens is reduced, resulting in an image with darkened edges. The distance between the objective lens and the tube length must be optimized in order to maximize the flexibility of the microscope and maintain system performance.

Typically, the tube lens has one or two doublets that focus the collimated light onto the desired intermediate image plane and compensate for the aberrations from the objective lens. The four major considerations in optimizing the tube lens are image quality, aberration compensation, exit pupil size, and exit pupil location.

Figure 4.29 illustrates a tube lens with a focal length of 200 mm and an entrance pupil of 10 mm. The distance between the objective lens and the tube lens is 120 mm, assuming 100 mm is the total unfolded length of the prisms. This tube lens is designed without compensating for the lateral chromatic aberration from the objective lens. It has diffraction-limited performance over the entire FOV across the visible spectrum. The lateral chromatic aberration is within the Airy radius. The exit pupil is optimized to be 320 mm from the intermediate image plane—close to the exit pupil of the objective lens.

With two doublets, it is fairly easy to design a tube lens to compensate for a reasonable amount of lateral chromatic aberration from the objective lens while maintaining the diffraction-limited performance and exit pupil position. Figure 4.30 shows the lateral chromatic aberrations of three different tube lenses with diffraction-limited performance. Curve 1 is the lateral chromatic aberration of the tube lens in Fig. 4.29. This lens is optimized without compensating for the lateral

chromatic aberration from the objective lens. Curves 2 and 3 are the lateral chromatic aberrations of the tube lenses designed to compensate for different amounts of lateral chromatic aberrations from the objective lens.

4.4 Eyepiece

There are many eyepiece designs in the literature and patents. Except for using aspherical and diffractive surfaces, almost all fundamental eyepiece forms had been developed before 1950.²

When designing an eyepiece, one can optimize the eyepiece as an independent optical unit with standard predefined parameters, such as the size of the intermediate image and the tube length. Most eyepieces used in commercial microscopes are designed independently with a flat object, which is the intermediate image, and a fixed tube length.

Another approach is to design the eyepiece together with the objective lens to compensate for certain aberrations of the objective lens, such as field curvature and lateral chromatic aberration. This approach may have better overall image quality, but neither the eyepiece nor the objective lens is interchangeable between different designs, and they have to be used together for better image quality.

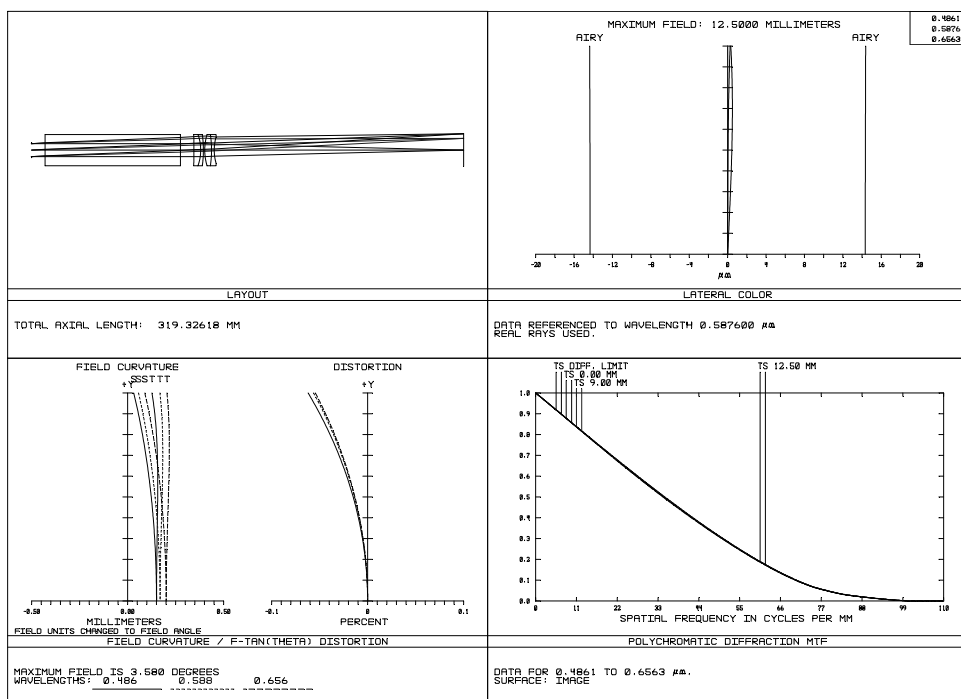


Figure 4.29 Layout and performance of a tube lens with a 200-mm focal length and a 10-mm entrance pupil.

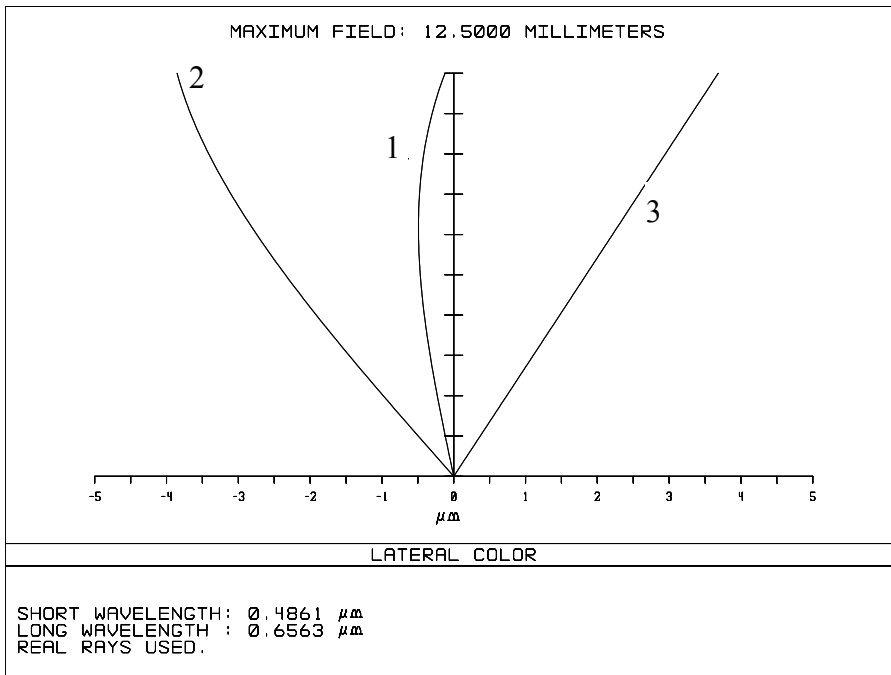


Figure 4.30 Lateral chromatic aberrations of three different tube lenses.

Given that the aperture is outside of the eyepiece, odd aberrations, such as coma, distortion, and lateral chromatic aberration, are more difficult to correct compared to other imaging systems, such as camera lenses, because there is absolutely no symmetry about the stop aperture. The aberrations of the eyepieces do not need to be corrected beyond what the normal human eye can detect, which means that angular aberrations do not need to be smaller than about 1 arcmin.

Correction of field curvature and astigmatism are important in eyepieces. It is sometimes preferred to overcorrect the tangential image surface so that the tangential and sagittal image surfaces are symmetrical with respect to a flat focal plane. The advantage to overcorrecting in this way is that the eyepiece design does not require any accommodation while viewing on axis. The drawback is that the astigmatism is large, and the image is blurred at the edges of the field. For a plan apochromat objective lens with high performance, a plan eyepiece with corrected field curvature is necessary. Most eyepieces have some pincushion distortion. The distortion can be as large 15–20% for wide-angle eyepieces. The human eye is sensitive to lateral chromatic aberration; therefore, this aberration must be reasonably controlled.

Another unique aberration to consider when designing an eyepiece is pupil spherical aberration, a phenomenon in which the position of the exit pupil is not fixed and it moves with field angles, as shown in Fig. 4.31. The larger the field, the closer the pupil location is to the eyepiece. Depending on the location of the iris of a human eye, the rays from some field points may not enter the eye or are strongly vignetted.

When the eye is closer to the eyepiece so as to view the edge of the field, zonal vignetting may appear in the zonal area of the field; this phenomenon is called the kidney bean effect. A displacement of 10% of the pupil diameter is typically acceptable for the pupil shift across the field. Aspherical surfaces can be used to correct pupil spherical aberration and are typically applied to the concave surfaces facing the pupil.

The design of the eyepiece usually begins with the selection of a basic eyepiece type according to the FOV. Then the lens is scaled to provide the correct magnification and is optimized to meet other requirements, such as image quality, eye relief, and compatibility with the objective lens.^{2,16}

Compared to other imaging systems, the design of the eyepiece requires more consideration because of the need to consider both image formation and pupil imaging.

Two approaches are generally used to trace the rays when designing an eyepiece. The first approach is to trace the rays in the reverse direction, from the pupil to the intermediate image plane, by setting the object at infinity, which corresponds to observation with a relaxed human eye. To optimize pupil imaging, a virtual stop is required to represent the exit pupil of the objective lens. The second approach is to trace the rays from the intermediate image back to the exit pupil of the objective lens and then to the eyepiece. A perfect lens is usually placed at the pupil plane to simulate the eye.

Figure 4.32 shows an Erfle eyepiece with two achromatic doublets and one positive singlet between them. The basic parameters of this eyepiece are 25-mm focal length, ± 30 -deg FOV, and 25-mm diameter of the intermediate image. The distance from the exit pupil of the objective lens to the intermediate image plane is 300 mm. This eyepiece is designed for an eye relief of 20 mm and a 5-mm pupil size.

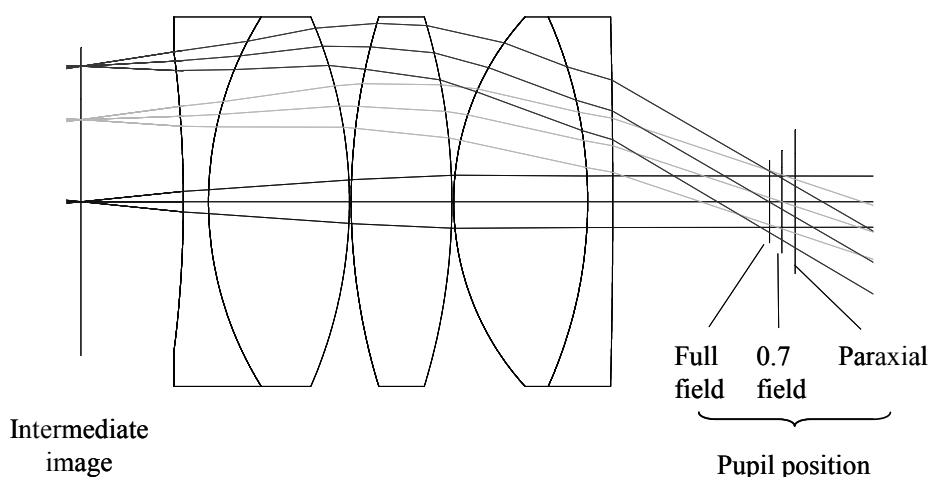
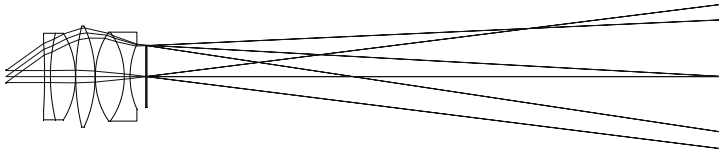
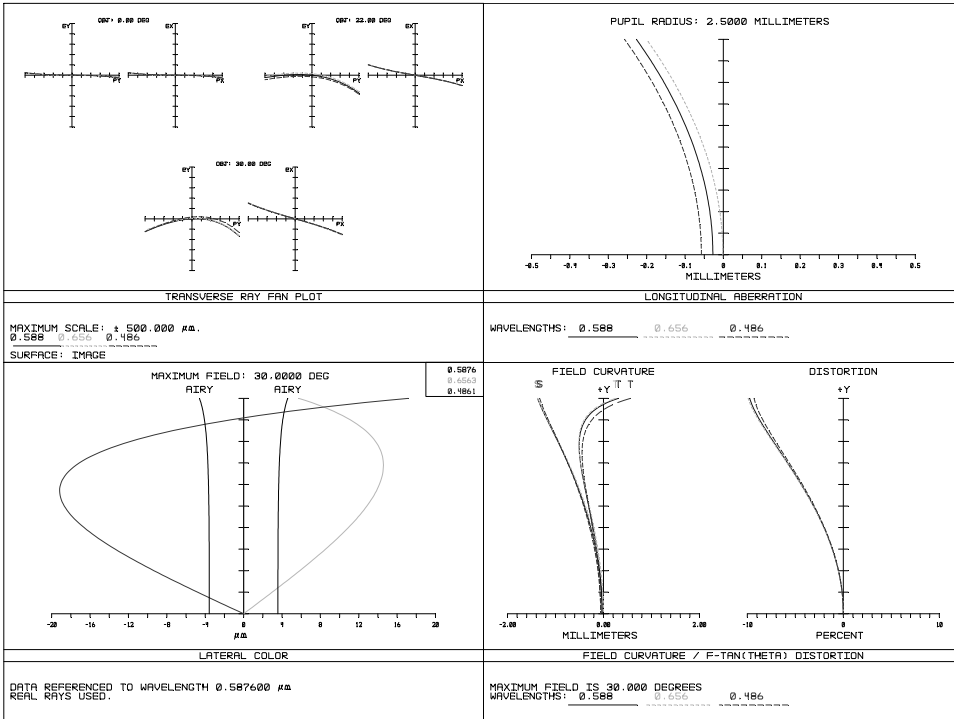


Figure 4.31 Pupil spherical aberration in the eyepiece.



(a)

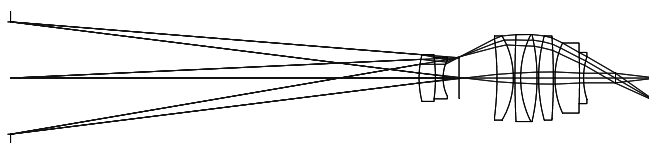


(b)

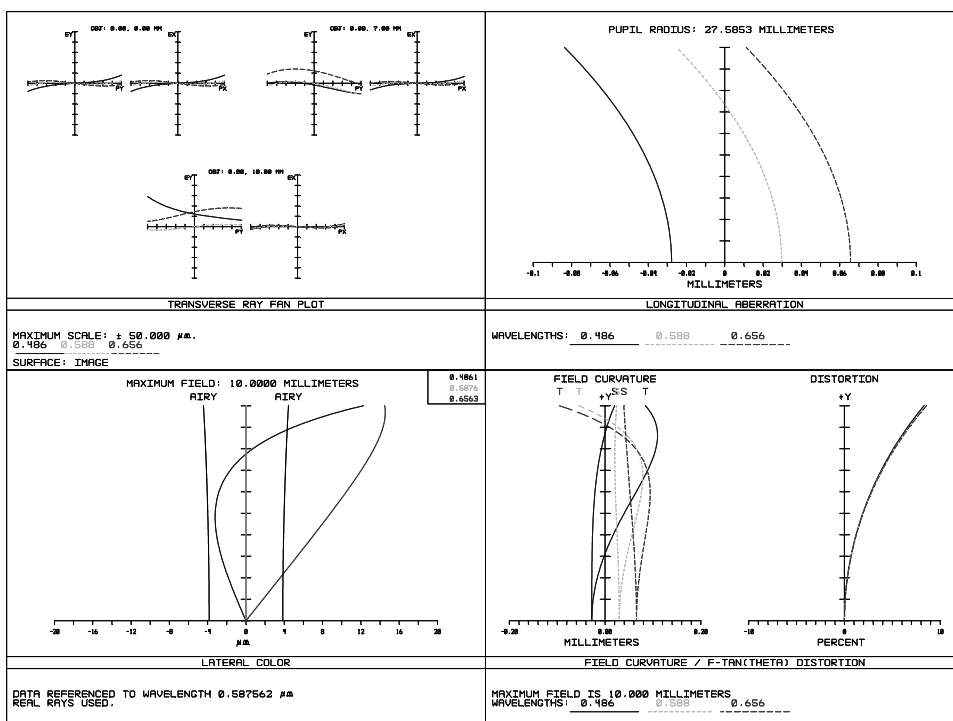
Figure 4.32 (a) Erfle eyepiece with two achromatic doublets and one positive singlet and (b) its performance plots.

The rays are traced from the pupil to the intermediate imaging plane to optimize image quality. During the optimization process, the chief ray positions are also controlled so that they intersect the optical axis at the exit pupil plane of the objective lens. As can be seen from Fig. 4.32(a), the pupil locations match reasonably well, but the edge of the pupil is blurred. The tangential surface is corrected, but the astigmatism is relatively large at the edge of the field. Spherical and axial chromatic aberrations are corrected reasonably well, but coma and lateral chromatic aberration are not satisfactory. Distortion is less than 10%.

Figure 4.33 illustrates a Nagler eyepiece with better astigmatism and field curvature correction than the Erfle eyepiece.¹⁷ The Nagler eyepiece has an FOV of ± 25 deg, a focal length of 25 mm, an intermediate image of 20-mm diameter, an eye relief of 19.5 mm, and a pupil size of 5 mm. As shown in Fig. 4.33(a), the rays are traced from the intermediate image to the exit pupil of the objective lens and then to the exit pupil of the eyepiece. A negative lens close to the intermediate image plane is used for controlling the astigmatism and field curvature and has very good correction of the pupil spherical aberration. The performance is calculated by placing a perfect lens with a 25-mm focal length at the pupil plane.



(a)



(b)

Figure 4.33 (a) Nagler eyepiece and (b) its performance plots.

4.5 Illumination System

Illumination is a critical determinant of the optical performance in a light microscope. The basic requirements of an illumination system for a microscope are sufficient light and good (angular, spatial, and spectral) illumination uniformity. Other requirements include the following: (1) the illumination field should be at least as large as the imaging field, and (2) the illumination angle should be at least as large as the acceptance angle of the objective.

There are two illumination modes in microscope imaging: epi-illumination and transillumination. Each mode has two methods to illuminate an object: bright field and dark field. Therefore, there are four types of illumination systems in total for microscope imaging, as shown in Fig. 4.34.

In the epi-illumination mode, the object is illuminated from the same side as the imaging system; therefore, the objective lens collects the reflected and backscattered light. In the transillumination mode, the illumination and imaging systems are on opposite sides of the object. The objective lens collects the transmitted light.

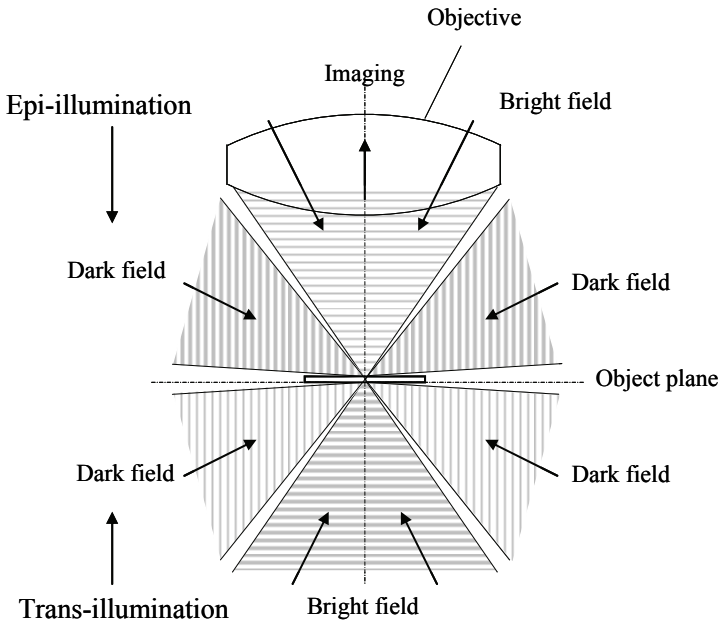


Figure 4.34 Illumination methods in a microscope.

For dark-field illumination, the smallest illumination angle should be larger than the acceptance angle of the objective lens so that the illumination light does not enter the objective lens directly by specular reflection or transmission. In contrast, the NA of a bright-field illumination system should be as large as or slightly larger than that of the imaging system to maximize the resolution and contrast.

Bright-field epi-illumination provides uniform illumination across the FOV and highlights surface information; it is not suitable for highly reflective surfaces. Dark-field epi-illumination is suitable for highly reflective shiny surfaces and emphasizes scatters. Transillumination improves the contrast of transparent objects. The bright-field mode emphasizes the object shape, while the dark-field mode emphasizes the edges.

4.5.1 Bright-field illumination

The two most commonly used configurations for bright-field illumination are critical illumination and Köhler illumination. The resolution obtained by either of these illumination methods should be the same. Köhler illumination provides a more uniform illumination over the entire FOV. Both configurations can be used in epi-illumination and transillumination modes. The aberration correction of the illumination system is of relatively minor importance because it only affects the illumination uniformity and light efficiency.

4.5.1.1 Critical illumination

The critical illumination method simply images the light source directly onto the sample, as shown in Fig. 4.35. The light source is placed in front of the field stop and is imaged by the condenser onto the object plane. A collector is sometimes used to increase light efficiency. Typically, there is an aperture stop between the field stop and the condenser. The amount of light entering the microscope can be controlled by the aperture stop. The illumination field is adjusted by the field aperture near the light source.

One requirement of the condenser is to image the light source or the field aperture onto the object plane with enough fidelity such that the illumination field has a sharp boundary. Another desired feature is telecentricity. Because most objective lenses are designed as telecentric in object space, the system will be more efficient if the condenser is telecentric or semitelecentric on the object side so that the chief ray angle is small.

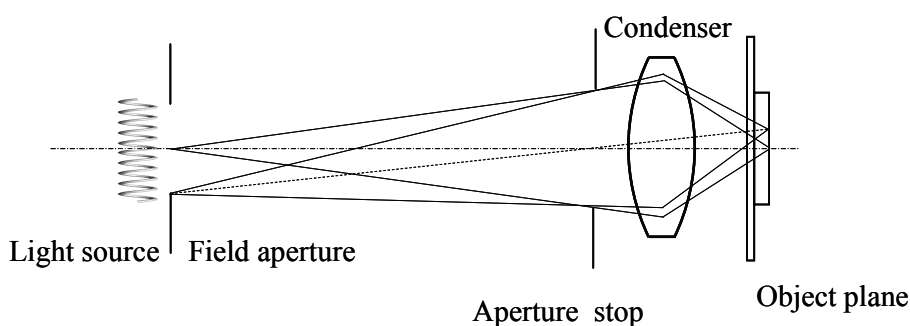


Figure 4.35 Critical illumination system.

Because the light source is imaged directly onto the object, any structural information and nonuniformity in the light source will appear at the object plane and will cause the image brightness to fluctuate. For example, the brightness fluctuation produced by convection within the plasma of an arc discharge can generate temporal variations in the object plane. Instead of imaging the light source directly onto the object plane, an effective solution is to first couple the light from the light source to a light homogenizer, such as a light pipe, fiber or liquid light guide, or a lenslet array, to reduce the angular, spatial, and temporal light variation and then to image the output surface of the light homogenizer onto the object plane. An alternative approach is to place a diffuser between the light source and the condenser. Defocus can also help to eliminate the structured information of the light source but is not very effective.

Figure 4.36 illustrates a condenser for critical illumination and its ray fan plot. The condenser has an illumination field of 2 mm and an NA of 0.9. The lens surface close to the aperture stop is aspherical to reduce the spherical aberration and coma. The chief ray angle is less than 1.5 deg. As the ray fan plot shows, spherical aberration is the major remaining aberration. Other aberrations include field curvature and axial chromatic aberration. The magnitude of the aberration is much larger than the imaging system but is acceptable to the illumination system. The condenser uses only one kind of crown glass, K5, and there is no effort to control the chromatic aberration. To minimize chromatic aberration and further reduce spherical aberration, more lenses, including cemented doublets, are commonly used.

Figure 4.37 shows the simulated irradiances at the field stop, aperture stop, and the object plane. The light distribution at the field aperture in Fig. 4.37(a) largely depends on the spatial light distribution of the light source if there is not a light homogenizer. As shown in Fig. 4.37(b), the light distribution at the aperture stop is relatively uniform; it is not affected by the spatial distribution of the light

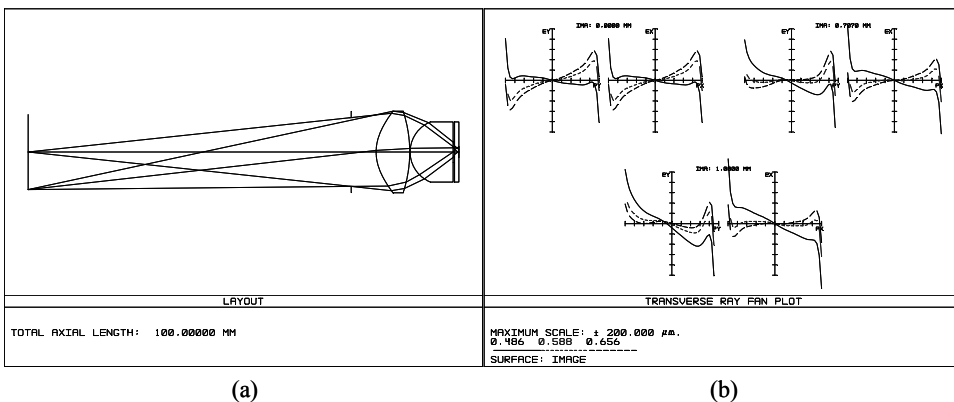


Figure 4.36 Condenser for critical illumination: (a) ray-tracing diagram and (b) ray fan plot.

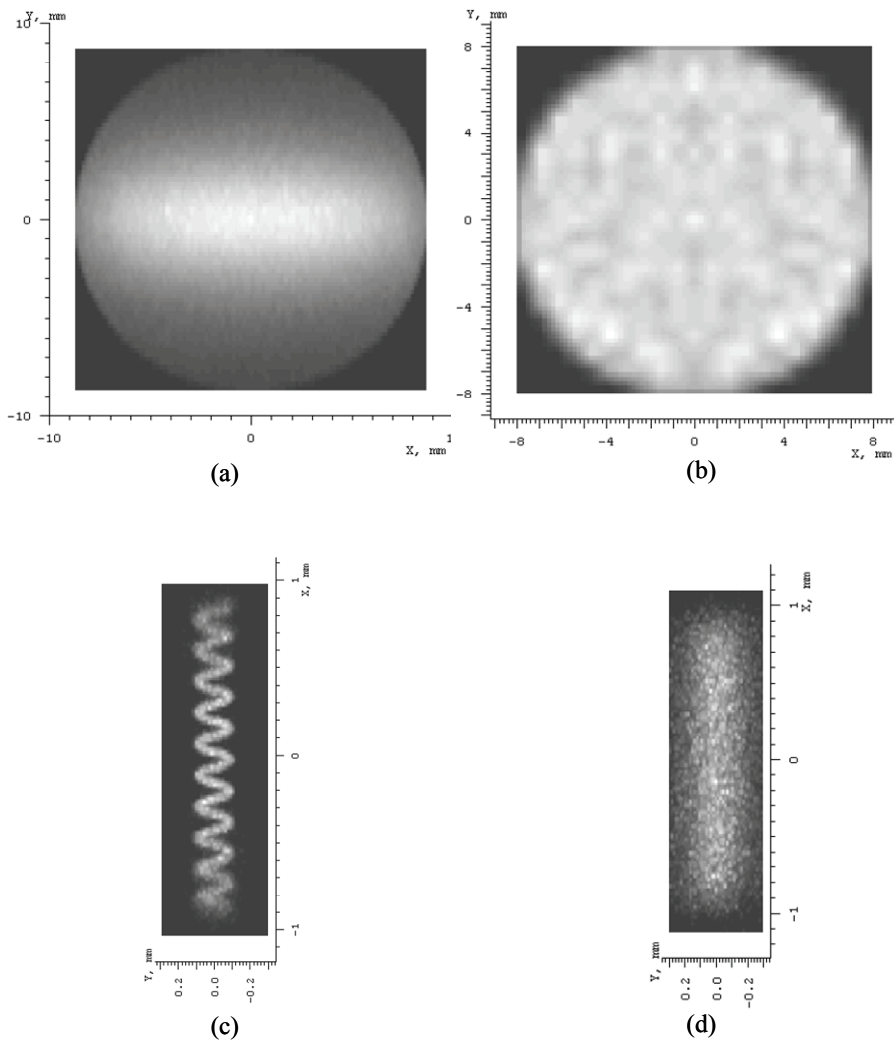


Figure 4.37 Light distributions at different planes in a critical illumination system. (a), (b), and (c) show the irradiances in the planes of the field aperture, the aperture stop, and the object plane, respectively, whereas (d) shows the irradiance in the object plane when a diffuser is placed between the light source and the field aperture.

source but depends on its angular distribution. On the other hand, the light distribution on the object plane is largely affected by the spatial distribution of the light source. The lamp filament is clearly reproduced in Fig. 4.37(c). Figure 4.37(d) shows the light distribution at the object plane when a diffuser is inserted between the light source and the field aperture. The illumination uniformity at the object plane is improved considerably.

4.5.1.2 Köhler illumination

Köhler illumination was first introduced in 1893 by August Köhler of the Carl Zeiss Corporation. As shown in Fig. 4.38, to overcome the issues related to critical illumination, the light source is imaged onto the aperture stop in front of the condenser by the collector, and the condenser images the field aperture behind the collector onto the object plane. The field aperture is adjustable to accommodate the size of the imaging field. The aperture stop is adjustable to control image brightness as well as the degree of partial coherence of the illumination.

In Köhler illumination, light rays from each point of the light source fill the field aperture, and some of the light rays then reach the object plane. Therefore, each point on the object sees a large portion of the light source, but each point sees only a portion of the intensity distribution. Therefore, the illumination uniformity is determined by the angular distribution of the light source. The size of the field aperture affects only the region of the illuminated field and not its brightness. Since the light source is also imaged onto the pupil plane of the observation system by the condenser and part of the objective lens, the illumination uniformity at the aperture stop of the objective lens is governed by the emission characteristics and geometry of the light source. Therefore, the angular characteristic of the objective lens may not be uniform, and the diffraction pattern produced by each point in the object plane may depart from the ideal Airy disk. To overcome this issue, a diffuser can be placed behind the collector.

As one of the key elements in Köhler illumination, the collector collects the light from the light source, and then delivers it to the aperture stop. The collector should have a large NA to capture as much light from the source as possible. A short focal length also helps to reduce the size of the light source that is projected onto the aperture stop. Generally, it is preferable to have a small chief ray angle on the light source side to keep the collection efficiency uniform across the entire field. Because the collector is close to the light source, the temperature in that area is generally too high for the cementing material. Therefore, only singlets are used in the collector, and chromatic aberration is not corrected.

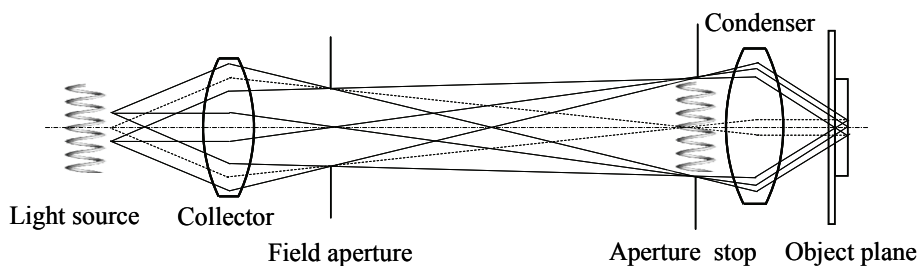


Figure 4.38 Köhler illumination system.

When the condenser stop aperture is completely filled with the image of the light source, the irradiance on the object plane is determined primarily by the intrinsic brightness of the light source as well as the NAs of the collector and condenser. The size of the light source affects the field irradiance only if the light source does not completely fill the aperture stop.

Ideally, the illumination field will be slightly larger than the size of the imaging field so that the object is fully illuminated. Also, the NA of the condenser should be as large as, or slightly larger than, that of the objective lens. The field aperture should be imaged onto the object plane with a sharp edge to reduce stray light.

It should be noted that the light source is generally not planar, except when a light guide is used to couple light from the light source to homogenize the light distribution or scramble the light to decrease the coherence. As a result, even though the illumination system perfectly images the central plane of the light source at the aperture stop of the condenser, a significant portion of the light away from the central plane will not be perfectly focused at the aperture stop. Therefore, some level of structure in the light source may still present at the object plane.

Köhler illumination is not the most efficient system because it does not use the full surface of the source or the full angular distribution of the emitted light. It should be noted that depending on the uniformity of the light source and the optical quality of the illumination system, the light from the source may not illuminate the object at all possible angles with equal intensity.

In summary, the irradiance at the object plane is determined by the radiance and the angular light distribution of the light source, the NA and the focal length of the collector, the condenser NA, and the overall transmittance of the illumination system.

Figure 4.39(a) shows a three-lens collector with an NA of 0.75 and a focal length of 11.2 mm for a light source with an emitting size smaller than 3 mm. Figure 4.39(b) is the ray fan plot showing that the collector has significant spherical aberration. In this design, the aperture is located at the front focal point; therefore, the chief rays in object space are parallel to the optical axis.

Figure 4.40 is a Köhler illumination system using the collector in Fig. 4.39 and the condenser in Fig. 4.36. The NA at the object plane is 0.85. The rays are traced from the object plane to the light source. Only singlets with crown glass are used; therefore, chromatic aberration is not corrected.

Figure 4.41 illustrates the simulated irradiance plots in the planes of the field aperture, the aperture stop, and the object plane. The irradiance plot in the aperture stop clearly shows the structural information of the light source. The irradiance plot in the field aperture reflects the angular light distribution of the light source. Since the uniformity at the object plane is determined by the angular characteristics of the light source, the light distribution is not perfectly uniform in the object plane unless the light source has uniform angular distribution, but it is much better than that achieved with critical illumination. To further improve the uniformity of critical illumination, beam homogenizers, such as diffusers and

light pipes, can be used to angularly and spatially redistribute light from the light source.

The critical and Köhler illumination systems discussed above are used in transillumination mode. In epi-illumination mode, the objective lens is used as a condenser. For critical illumination used in epi-illumination mode, an additional lens is generally required, together with the objective lens, to image the light source onto the object plane. Because the aperture stop of the objective lens is not accessible, the brightness of the illumination is adjusted by applying a different voltage to the lamp.

As discussed above, illumination uniformity is either determined by the spatial (critical illumination) or angular (Köhler illumination) light distribution of the light source. Fiber illumination is one of the most effective methods to improve illumination uniformity in microscopic imaging systems. The basic

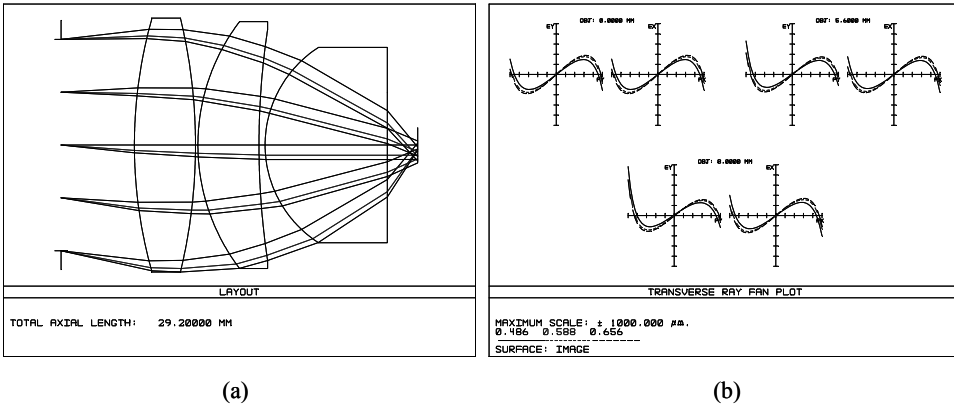


Figure 4.39 A 0.75 NA collector with three singlets: (a) ray-tracing diagram and (b) ray fan plot.

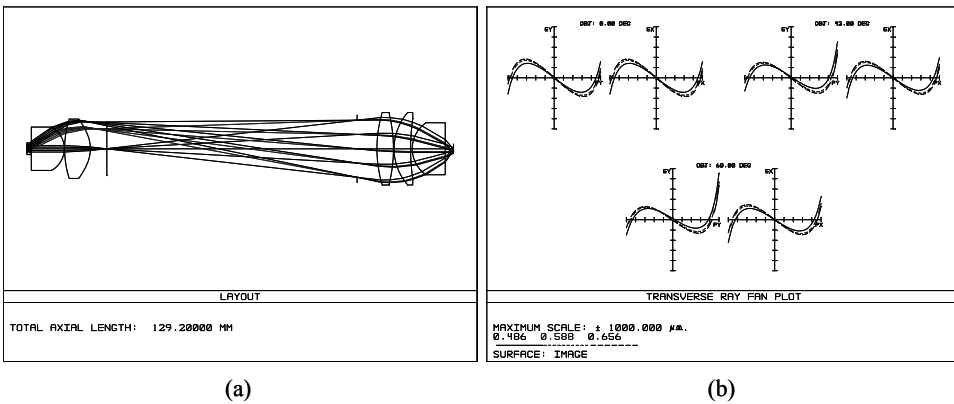


Figure 4.40 Köhler illumination with the collector in Fig. 4.39 and the condenser in Fig. 4.36: (a) ray-tracing diagram and (b) ray fan plot.

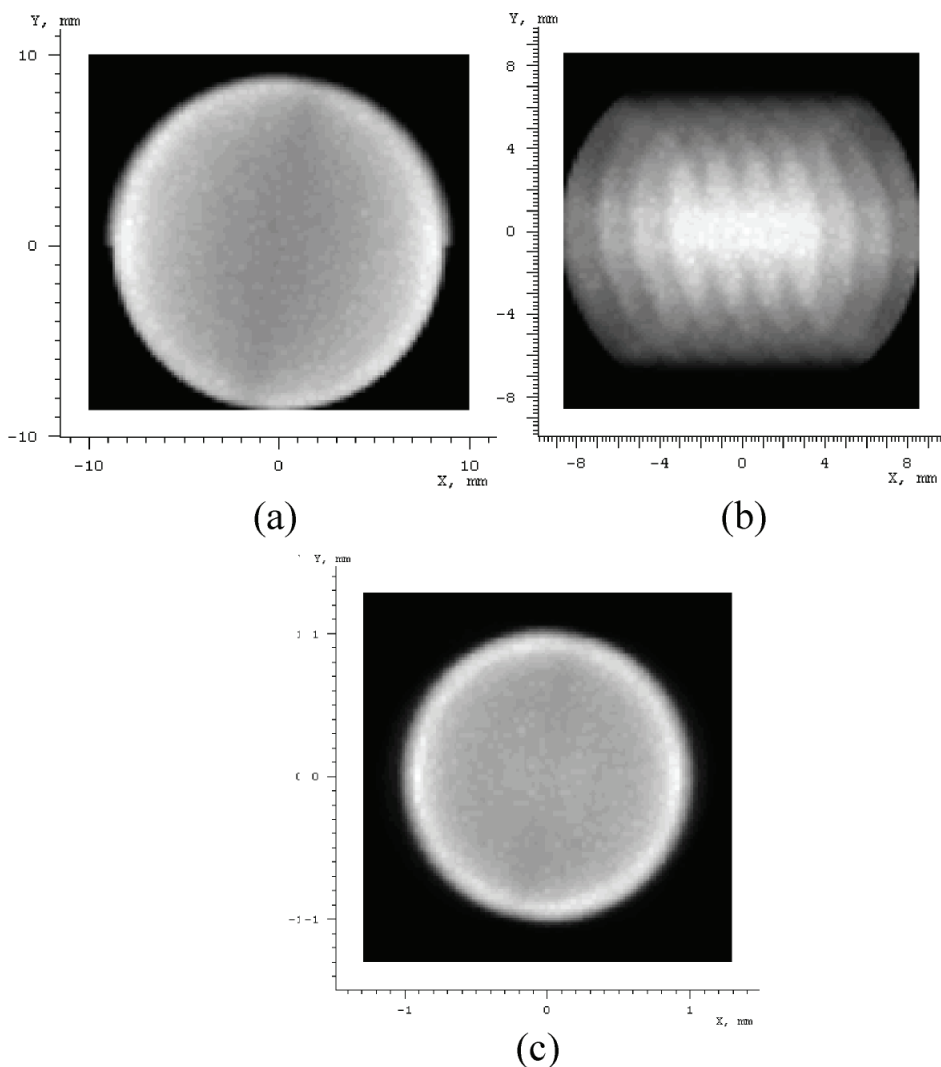


Figure 4.41 Irradiance plot at the (a) field aperture, (b) aperture stop, and (c) object plane of the Köhler illumination system in Fig. 4.40.

approach is to couple the light into a multimode fiber or fiber bundle, then the output surface of the fiber or fiber bundle is treated as a light source. Both spatial and angular distribution of the illumination light can be improved significantly. Two advantages of using fibers in illumination systems are: (1) the bulky light source can be placed outside of the microscope for better heat management, and (2) the coherence of illumination light can be reduced through fiber vibration when a laser is used as the light source.

4.5.2 Dark-field illumination

For some objects, conventional imaging methods cannot provide a clear image because the object's features do not have enough contrast. For example, the refractive indices of unstained transparent specimens are very close to those of their surroundings. The signal from the scattering, diffraction, refraction, and reflection at the tissue boundaries is much weaker than the directly transmitted or reflected light. Therefore, image contrast is poor under bright-field illumination.

If the directly transmitted or reflected light is blocked, the scattered, diffracted, and refracted light from the optical discontinuities (such as the cell membrane, nucleus, and internal organelles) can be observed against a black background. This can be achieved through dark-field illumination. The resolving power of the objective lens is the same in dark-field illumination and in bright-field illumination, but the optical character of the image is not as faithfully reproduced.

To prevent the directly transmitted or reflected light from entering the objective lens, the object should be illuminated with an angle larger than the acceptance angle of the objective lens. If no object or homogenized object is present, the illumination rays will not enter the objective lens and the FOV will appear dark.

Both reflective and refractive systems have been developed for dark-field illumination. The configurations of refractive dark-field illumination systems are similar to those of bright-field systems, except that a ring aperture is used to block the light having an illumination angle smaller than the acceptance angle of the objective lens. Reflective dark-field systems are used more commonly than refractive systems because they are effective at achieving large-angle illumination. There are various designs of reflective dark-field condensers, including paraboloid, ellipse, cardioid, and spherical surface, amongst others.

Figure 4.42 shows two diagrams of simple dark-field condensers. The condenser in Fig. 4.42(a) is based on a parabolic reflector. The input illumination light should be collimated to obtain an aberration-free spot. The illumination angle is determined by the geometrical parameters of the parabolic reflector and the ring size of the input light. The smallest illumination angle can be finely tuned by adjusting the position or the size of the beam blocker. The collimation degree of the input light determines the illumination field.

The condenser in Fig. 4.42(b) is based on an elliptical reflector. When the light source is placed at one focal point of the elliptical reflector, the reflector will focus the light toward another focal point without aberration if the light source is a point light source. Any deviation from a point source will increase the illumination spot size. This property can be effectively used to adjust the illumination spot size. The range of the illumination angle is determined by the geometrical parameters of the elliptical reflector, such as the distance between the two focal points. Similarly, the smallest illumination angle can be adjusted by changing the size of the beam blocker or by moving it away from or toward the object plane.

Figure 4.43(a) shows a bispheric, double-reflection condenser. The first spherical reflector diverges the collimated light, and the second spherical reflector focuses the light onto a small spot. Because of the aberrations introduced by the two reflective mirrors, the spot size cannot be infinitely small, even for a perfectly collimated beam.

In practice, the condenser is sometimes designed and fabricated as a monolithic block, as shown in Fig. 4.43(b). A refractive spherical surface, whose

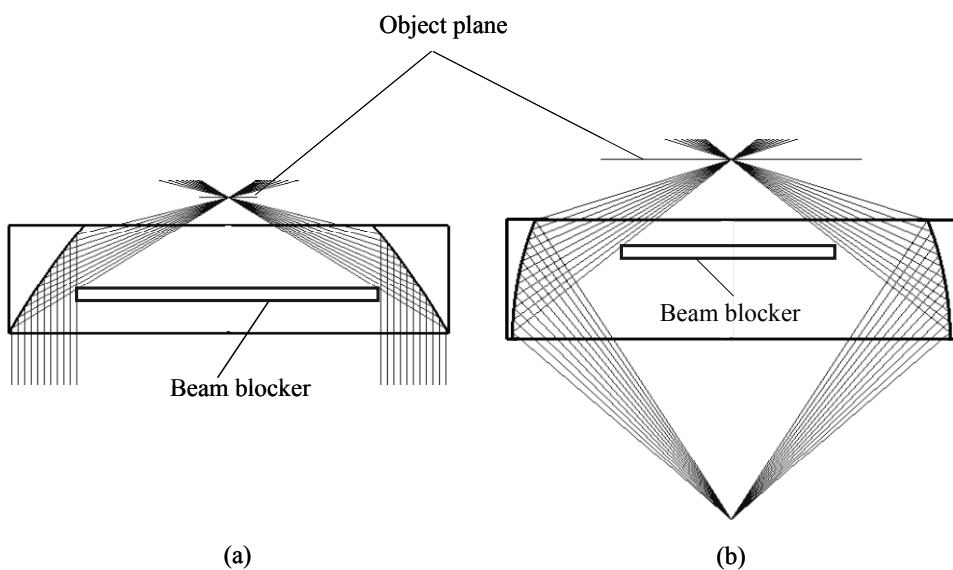


Figure 4.42 Dark-field condensers: (a) parabolic condenser and (b) elliptical condenser.

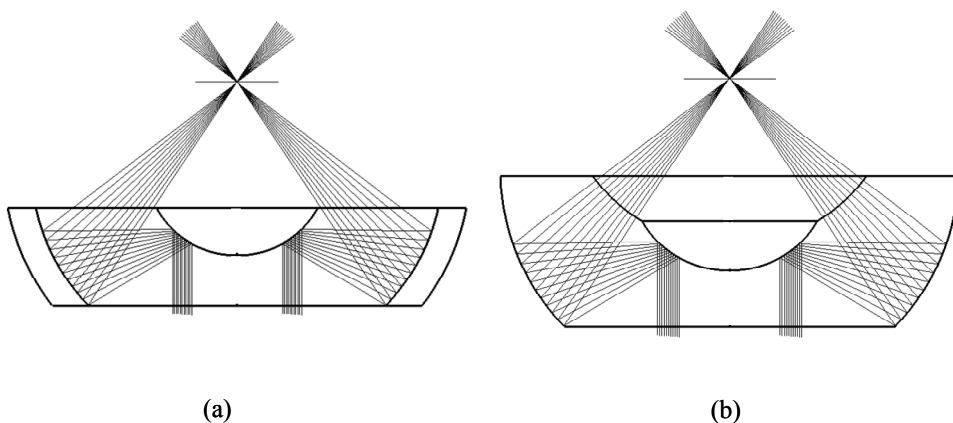


Figure 4.43 Bispheric double-reflection condensers focus the collimated beam onto a small spot: (a) two-element version and (b) monolithic version.

center is at the object plane, is added as the exit surface of the condenser to prevent TIR at the exit surface.

Figure 4.44 shows the simulated irradiance and intensity in the focal plane of the dark-field condenser in Fig. 4.43(a). The plots show that the illumination has a sharp edge in both the spatial and angular distributions. The irradiance plot in Fig. 4.44(a) shows that the illumination at the focal plane is uniform. The range of illumination angles can be found from the intensity plot in Fig. 4.44(b), which is not uniform.

The dark-field condensers discussed above are designed for transillumination only. However, the design principle of a dark-field condenser for epi-illumination is the same, except that the condenser for epi-illumination is subject to more space restrictions.

Dark-field imaging is an effective tool in biomedical imaging applications, such as photographing living bacteria with high magnification and viewing cells and tissues with low magnification. Any featureless region of the object will transmit or reflect the light from a dark-field condenser away from the objective lens, resulting in a dark background. Features in the object will cause the illumination light to scatter so that some of the transmitted or backscattered light will pass through the objective lens and reach the sensor or eye. Those features will show up as bright regions on an otherwise black background. Dark-field imaging can also be used in combination with other imaging techniques, such as fluorescence imaging.

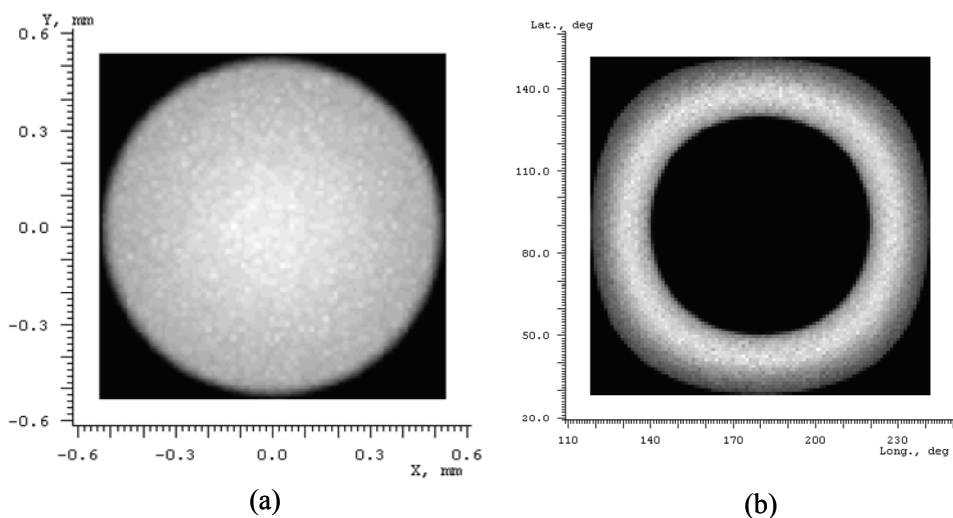


Figure 4.44 Simulation results of the dark-field condenser in Fig. 4.43(a): (a) irradiance and (b) intensity.

References

1. S. Inoue and R. Oldenbourg, "Microscopes," Chapter 17 in *Handbook of Optics*, vol. 2, 2nd ed., M. Bass, Ed., McGraw-Hill, New York (1995).
2. H. Gross, F. Blechinger, and B. Aichtner, *Handbook of Optical Systems*, vol. 4, *Survey of Optical Instruments*, Wiley-VCH, Berlin (2008).
3. H. Gross, H. Zügge, M. Peschka, and F. Blechinger, *Handbook of Optical Systems*, vol. 3, *Aberration Theory and Correction of Optical Systems*, Wiley-VCH, Berlin (2007).
4. K. Uhlenndorf, H. Gross, L. Schreiber, and R. Wartmann, "Developments and design of optical systems for microscopes at Carl Zeiss," *Proc. SPIE* **5962**, 59620P (2005).
5. M. J. Kidger, *Fundamental Optical Design*, SPIE Press, Bellingham, WA (2002) [doi: 10.1117/3.397107].
6. M. J. Kidger, *Intermediate Optical Design*, SPIE Press, Bellingham, WA (2004) [doi: 10.1117/3.540692].
7. D. M. Aikens, "Beginning the design of a UV microscope objective," *Proc. SPIE* **1690**, 20–29 (1992) [doi: 10.1117/12.138010].
8. P. Erdős, "Mirror anastigmat with two concentric spherical surfaces," *J. Opt. Soc. Am.* **49**, 877–883 (1959).
9. D. S. Grey, "A new series of microscope objectives: III. Ultraviolet objectives of intermediate numerical aperture," *J. Opt. Soc. Am.* **40**, 283–290 (1950).
10. D. S. Grey, "Computed aberrations of spherical Schwarzschild reflecting microscope objectives," *J. Opt. Soc. Am.* **41**, 183–192 (1951).
11. K. P. Norris, W. E. Seeds, and M. H. F. Wilkins, "Reflecting microscopes with spherical mirrors," *J. Opt. Soc. Am.* **41**, 111–119 (1951).
12. W. Thornburg, "Reflecting objective for microscopy," *J. Opt. Soc. Am.* **45**, 740–743 (1955).
13. D. S. Grey and P. H. Lee, "A new series of microscope objectives: I. Catadioptric Newtonian systems," *J. Opt. Soc. Am.* **39**, 719–723 (1949).
14. D. S. Grey and P. H. Lee, "A new series of microscope objectives: II. Preliminary investigation of catadioptric Schwarzschild systems," *J. Opt. Soc. Am.* **39**, 723–728 (1949).
15. D. R. Shafer, "Catadioptric imaging system," U.S. Patent No. 5,031,976 (1991).
16. W. J. Smith, *Modern Lens Design*, 2nd ed., McGraw-Hill, New York (2005).
17. A. Nagler, "Ultra wide angle eyepiece," U.S. Patent No. 4,747,675 (1988).

Chapter 5

Fluorescence Imaging

Because of the tremendously sensitive emission profiles and high specificity of fluorescence signals, as well as the selectivity offered by antibody labeling, fluorescence imaging has become one of the most powerful techniques in biomedical research and clinical applications in recent years. Long after the phenomenon of fluorescence was discovered by Stokes in 1852,¹ fluorescence in the biomedical field was first reported in 1911 when fluorescence was observed coming from an animal subjected to UV illumination.² Fluorescence imaging was proposed as a method for cancer diagnostics in 1948.³ Winkelman and Rasmussen-Taxdal performed the first quantitative measurements of *in vivo* fluorescence using exogenous fluorophores in 1960.⁴ Since then, fluorescence technology has attracted extensive research for biomedical applications in both autofluorescence and fluorescence imaging using contrast agents.

In this chapter, the principle of fluorescence imaging and fluorescence imaging techniques will be discussed in Secs. 5.1 and 5.2. Sections 5.3 and 5.4 will focus on the key components of fluorescence imaging systems and fluorescence filters. The configurations and optical design of fluorescence imaging systems will be discussed in Secs. 5.5 and 5.6.

5.1 Introduction to Fluorescence

5.1.1 Fluorescence process

When illuminated with light possessing a suitable spectrum, some specimens, living or nonliving, organic or inorganic, absorb the illumination light and then radiate light with a different wavelength. This phenomenon, called the fluorescence process, was discovered by British scientist Sir George G. Stokes in the middle of the nineteenth century. The fluorescence process is commonly illustrated by the simple electronic-state diagram called the Jablonski energy diagram, as shown in Fig. 5.1.

When molecules absorb light of a suitable wavelength λ_{ex} , electrons may be raised from the ground state S_0 to a higher-energy and vibrationally excited state S_2 . This process may only take 10^{-15} s. Within 10^{-14} to 10^{-11} s, the excited electrons may lose some vibrational energy to the surrounding environment in the form of heat and then relax to the lowest vibrational energy level S_1 within

the electronically excited state from which the fluorescence emission originates. When electrons relax from the excited state to the ground state, light is often emitted at a longer wavelength λ_{em} . This emission process may take 10^{-9} s. The wavelength of the emitted light is determined by E_{em} , the energy difference between the energy levels of the two states:

$$\lambda_{em} = \frac{hc}{E_{em}}, \quad (5.1)$$

where h is Planck's constant and c is the speed of light.

The entire fluorescence process cycles when the fluorescent specimen is illuminated unless the fluorophore is irreversibly destroyed in the excited state by photobleaching. If the exciting radiation is stopped, the fluorescence process ends. Therefore, the same fluorophore can be repeatedly excited and detected, although a maximum number of excitation events exists. This is the fundamental process of the fluorescence detection technique.

Fluorescent molecules, which are capable of emitting fluorescence when illuminated with light of a suitable wavelength, are known as fluorescent probes, fluorochromes, or fluorescent dyes. When the fluorochromes are

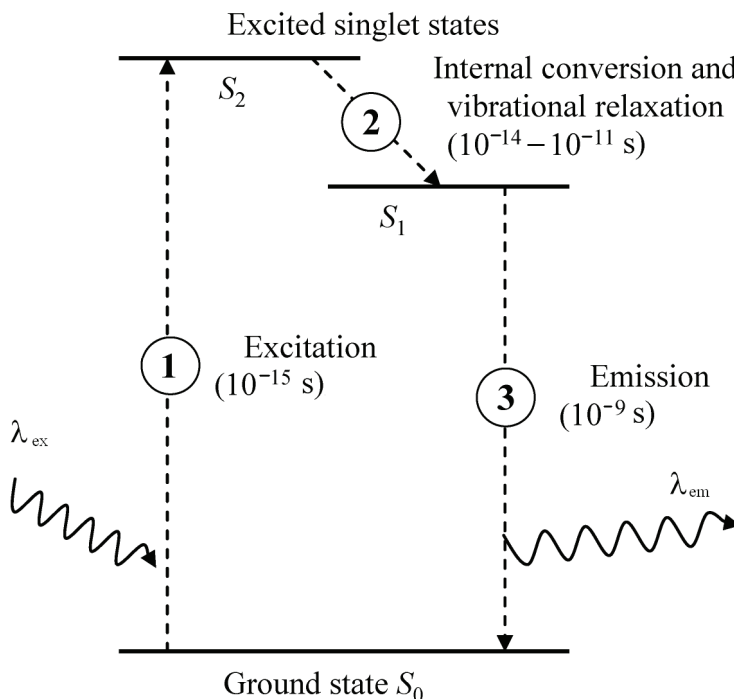


Figure 5.1 Jablonski energy diagram illustrating fluorescence processing. The stages labeled 1, 2, and 3 are excitation, vibrational relaxation, and emission, respectively.

conjugated to a larger macromolecule, they are known as fluorophores. There are two types of fluorophores, intrinsic and extrinsic. Intrinsic fluorophores are those that occur naturally, such as aromatic amino acids, nicotinamide adenine dinucleotide (NADH), flavin adenine dinucleotide (FAD), porphyrins, and green fluorescent protein (GFP). Extrinsic fluorophores are synthetic dyes or modified biochemicals that are added to a sample to provide fluorescence. Examples include dansyl, fluorescein, and rhodamine.

5.1.2 Excitation and emission spectra

Each fluorescent molecule has unique and characteristic spectra of excitation and emission. The normalized intensities of these spectra are usually plotted side by side in the same graph, as shown in Fig. 5.2.

An excitation spectrum describes the relative probability that a fluorophore will be excited by an illumination light of a given wavelength. Light with a wavelength distant from the excitation maximum can also cause excitation, but it does so less efficiently. Therefore, each excitation spectrum has a certain width. For some fluorophores, there is a second peak at a shorter wavelength of the excitation spectrum. This phenomenon occurs because the fluorescent molecule can be in any of several vibrational and rotational energy levels within the ground state and can be excited to any of several vibrational and rotational energy levels.

An emission spectrum is a plot of the relative intensity of the emitted light as a function of wavelength. It shows the relative probability of an emitted photon at a particular wavelength. For many fluorophores, their vibrational energy level spacing is similar for the ground and excited states, so the shape of the emission spectrum is approximately a mirror image of the excitation spectrum, as shown in Fig. 5.2.

Before fluorescent light is emitted, some energy is dissipated as heat during the excited state due to vibrational and rotational relaxation and collisions; as a result, the energy of the emitted photon is lower than that of the excitation photon. Therefore, the emission spectrum is always shifted toward a longer wavelength with respect to the excitation wavelength for one-photon excitation. The difference in wavelength between the emission peak and excitation peak is called Stokes shift. The Stokes shift is fundamental to the sensitivity of fluorescence techniques. By completely filtering out the excited light, it is possible to see weakly fluorescent objects.

In Fig. 5.2, the typical excitation and emission spectra are distinct but often overlap to some degree, more or less for different fluorophores. This overlap must be eliminated in fluorescence imaging by using appropriate filters. Otherwise, the fluorescent light will be overwhelmed because the excitation light is much brighter than the fluorescence.

There are some other characteristics of the emission spectrum that are not shown in excitation and emission spectra. First, although the illumination light of the wavelength at the fluorophore's excitation maximum produces the greatest

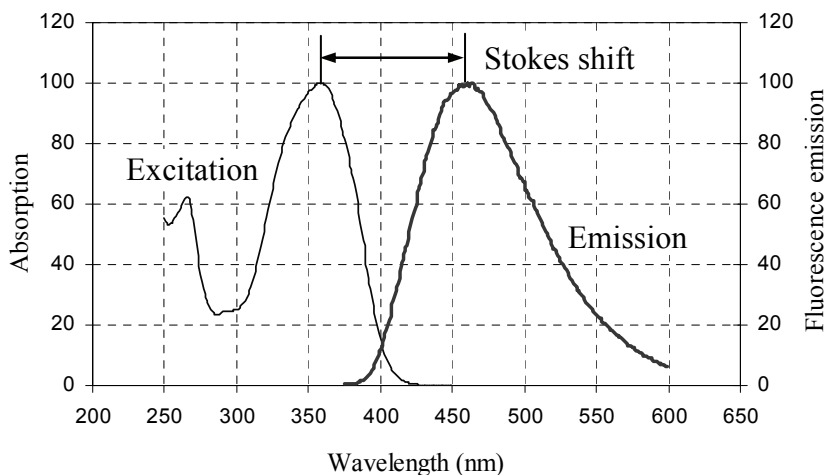


Figure 5.2 Normalized excitation and emission spectra of 4'-6-diamidino-2-phenylindole (DAPI) bound to DNA.

fluorescence output, illumination at shorter or longer wavelengths affects only the intensity of the emitted light; the range and overall shape of the emission profile stay exactly the same. Second, the excitation and emission spectra may shift and change in shape with changes in their cellular environment, such as pH level, dye concentration, and conjugation to other substances. Third, certain photochemical reactions may decrease the fluorescence efficiency of a fluorophore, a phenomenon called photobleaching or fading.

Fluorescence intensity is generally a function of the intensity of the excitation light. A more-intense light excites a greater fluorescence signal. However, as discussed above, photobleaching may occur if the specimen is illuminated by a high-intensity light over a prolonged period of time, especially when UV or near-UV light is used. Fluorescence intensity is also affected by the fluorophore concentration, duration of excitation, temperature, pH, and thickness of the specimen.

When the wavelength and intensity of the excitation light are constant, the intensity of the emitted fluorescent light is a linear function of the concentration of fluorophore. This linearity may not be maintained when the fluorophore concentration is very high, and inner filter effects may take place. The intensity of the emitted fluorescent light is proportional to the product of two properties of the fluorophore, namely, the extinction coefficient, which is the ability to absorb the excitation light, and QE, which is the ability to convert the absorbed light into emitted fluorescent light. Both properties must be considered when evaluating the fluorophores because the intensities of the fluorophores when excited at the same wavelength but with different concentrations and quantum efficiencies may be the same.⁵

5.1.3 Intrinsic and extrinsic fluorescence

Basically, there are two types of fluorescence: intrinsic fluorescence (also called autofluorescence and endogenous fluorescence) and extrinsic fluorescence (also called exogenous fluorescence). Autofluorescence emission, arising from intrinsic fluorophores, is an intrinsic property of tissues, while extrinsic fluorescence is obtained by adding or expressing exogenous fluorophores, such as fluorescein isothiocyanate (FITC), GFP, and phycoerythrin (PE).

Autofluorescence in tissues is typically produced by fluorophores that naturally occur in living cells after excitation with a suitable wavelength. These fluorophores include components of connective tissues (collagen, elastin), cellular-metabolism-related coenzymes [reduced NADH, FAD, flavin mononucleotide (FMN)], aromatic amino acids (tryptophan, tyrosine, and phenylalanine), by-products of heme biosynthesis (porphyrins), and lipopigments (lipofuscin and ceroids). The peak excitation and emission wavelengths of some endogenous fluorophores are listed in Table 5.1.⁶ Each fluorophore is characterized by its peak excitation and peak emission wavelengths, as well as its QE and fluorescence intensity decay times. The excitation spectrum is usually quite broad, and any light within the excitation spectrum will excite the same emission spectrum at a different level. Therefore, it is common for a single wavelength to excite many fluorophores and for their emission signals to overlap. Generally, it is difficult to identify individual fluorophores in a given fluorescence spectrum.

One of the reasons that fluorescence imaging is effective in biomedical imaging is that the presence of disease changes tissue fluorescence. Tissue

Table 5.1 Some known endogenous fluorophores and wavelengths of their excitation and emission maxima. (Courtesy of John Wiley & Sons, Inc.)

Tissue fluorophore	Biological source	Wavelength of maximum fluorescence excitation (nm)	Wavelength of maximum fluorescence emission (nm)
Tryptophan	Amino acids	280	350
Collagen	Connective tissues	330	390
Elastin	Connective tissues	350	420
NADH	Metabolic cofactor	340	450
FAD	Metabolic cofactor	450	515
Porphyrins	By-product of heme biosynthesis; bacterial fauna	405	635
Ceroid, lipofuscin	Lipopigment granules; age related; lipid oxidation products	340–395	430–460, 540–640

fluorescence is affected by the following factors: changes in the mucosal-layer thickness, biochemical composition, concentration of the fluorophores, light scattering and absorption properties of the tissue, tissue blood content and blood concentration, nuclear size distribution, collagen content, and the metabolic status of the tissue. The difference between the fluorescence of normal and diseased tissues may be distinguishable.

The major advantage of using autofluorescence is that it does not require extrinsic fluorophores. The disadvantage is that the autofluorescence signal is usually very weak, sometimes requiring the use of a sophisticated imaging system.

The exogenous fluorophore offers an alternative for fluorescence imaging. It is a molecule that is naturally occurring or specially designed and can be used as a probe to label cells or tissues. Probes can be designed to localize the tissues, cells, or proteins within a cell, as well as to respond to a specific stimulus and to monitor the production of a gene product. Fluorophores can be bound to an antibody for delivery to specific targets. GFP and red fluorescence protein (RFP) are two examples of naturally occurring fluorophore proteins that emit fluorescent light in green and red wavelengths.

Most fluorophores are chemically designed to fluoresce in visible or near-IR ranges. Examples include cytochrome-c-mono-2-deoxy-glucose, indocyanine green dye (ICG), and indocyanine dye Cy5.5. A Cy5.5-conjugated biocompatible, protected-graft copolymer (PGC) has been used effectively to investigate therapeutic drug development within tumor microvasculature.⁷

The excitation and emission spectra of exogenous fluorophores can be specifically designed or selected for different applications. For example, in deep-tissue imaging, fluorophores with longer excitation and emission wavelengths (650–1000 nm) are usually used because they have lower scattering in tissue and lower absorption in blood and water. The availability and stability of the probe, target tissue depth, anatomical location, biocompatibility, and cell toxicity are some other factors to consider when selecting fluorophores for specific applications.

There are several advantages of using exogenous fluorophores. The fluorescence signal is much stronger compared with autofluorescence; therefore, a less-complicated imaging system can be used. Generally, when the optimum excitation and emission wavelengths are known, the autofluorescence can be subtracted from the detected fluorescence. However, the use of exogenous fluorophores has some drawbacks: A certain waiting time after application is necessary for the fluorophore to reach its optimal fluorescence intensity; the diagnostic effectiveness of an exogenous fluorophore depends on the degree of selective localization of the fluorophore; the time for the fluorophore to reach its optimal fluorescence intensity may vary from patient to patient and tissue to tissue; and the application of exogenous fluorophores also involves regulatory issues and additional costs. All of these restrictions may make the technique impractical for use in regular screenings of high-risk patient groups.

Fluorescent light is emitted in all directions, but only the light collected by the detector is useful. The captured fluorescent light is usually very weak compared to the excitation light. Tissue autofluorescence may interfere with a weak fluorescence signal. For example, the autofluorescence from collagens overlaps with the fluorescence from GFP.

One more type of optical signal in tissue is bioluminescence, the light emitted by a living organism as the result of a chemical reaction. Bioluminescence is different from fluorescence in that the excitation energy is supplied by a chemical reaction rather than from a light source. Bioluminescence imaging has become increasingly important for biomedical research and drug development using mouse models of human diseases.

5.2 Fluorescence Imaging Techniques

Fluorescence detection is a quantitative method widely used in biomedical imaging. One of the advantages, compared to other imaging methods, is its high sensitivity. For example, fluorescent stains and dyes are generally far more sensitive than traditional colorimetric methods in detecting total DNA, RNA, and proteins. Fluorescence imaging is also more sensitive to tissue changes, such as tooth decay, than conventional reflectance imaging.⁸

The fluorescence signal can be recorded and analyzed in a number of different ways. In a spectroscopic system, fluorescence is only collected from one or more points and analyzed through a spectrometer to obtain spectral information. 2D fluorescence images are usually recorded by a point-scanning, a line-scanning, or a 2D imaging system.

Several imaging techniques have been developed since fluorescence imaging technology was first applied to biomedical imaging. In this section, four selected fluorescence imaging techniques will be discussed.

5.2.1 Single-band fluorescence imaging

Single-band fluorescence imaging is the simplest fluorescence imaging technique. Only one band of the excitation spectrum is used, and only one band of fluorescence is detected. The excitation wavelength is typically selected as close to the peak of the excitation spectrum as possible for a stronger fluorescence signal. A longpass filter or a bandpass filter is used in the detection path to block the excitation light and allow the fluorescence to reach the detector. One application example of single-band fluorescence imaging is short-wavelength fluorescence imaging for tumor and dental-caries detection.^{9,10} Light with a short wavelength, for example, 400 nm, is optimal for exciting autofluorescence from tissue.

The imaging system is relatively simple for single-band fluorescence imaging. However, using only one wavelength limits the diagnostic potential of single-band fluorescence imaging because no spectroscopic information is provided. Fluorescence intensity at a single wavelength is not necessarily reliable because the fluorescence signal depends on many factors that are of no diagnostic

relevance, such as detection geometry, illumination uniformity, and surface topography. Therefore, single-band fluorescence imaging is usually combined with other imaging modalities, such as reflectance imaging.

5.2.2 Multicolor fluorescence imaging

When two or more wavelength bands in excitation and emission are used, more advanced and powerful contrast functions can be used and dimensionless images can be obtained.¹¹⁻¹³ Two or more fluorescence images also provide spectroscopic information. In addition, the ability to label multiple, distinct objects of interest in a single sample also greatly enhances the power of fluorescence imaging.

Multicolor imaging can be achieved by illuminating the sample either simultaneously or sequentially with two or more wavelength bands, depending on system configurations. In the sequential mode, two or more sets of filters are necessary so that images in different wavelength bands can be captured sequentially. The individual images are then combined for further analysis. One limitation of this approach is its relative low speed since two or more images are needed. Another potential issue is the pixel shift between images with different wavelengths captured at different times. Sequential imaging is sometimes preferred, especially when more than three excitation wavelengths are used for multispectral imaging.

In the simultaneous mode, several sensors are generally needed, one for each wavelength band. The fluorescent light is separated by dichroic beamsplitters and directed to the sensors. By using two or more separate cameras, it is possible to monitor fluorescence images at different wavelengths simultaneously, eliminating the pixel shift in a sequential mode. With the advance of coating technologies, optical filters with multiple passbands can be fabricated so that more than one fluorophore can be excited or the emission from more than one fluorophore can be observed with a single filter and, possibly, a single camera. When more than three bands are used, imaging systems for simultaneous modes tend to become complex and impractical.

With two or more independent fluorescence signals, it is possible to analyze the fluorescence as a dimensionless quantity by the ratio of the fluorescence signals at different wavelengths. In this case, the ratio of fluorescence is less dependent on measurement geometry and the spatial profile of the illumination.

Multicolor fluorescence detection using spectrally resolved fluorescent labels allows for the detection and analysis of two or more targets in the same sample simultaneously. This method is time and cost effective and eliminates sample-to-sample variation. For example, two-color fluorescence imaging is used in differential gene profiling with microarray analysis by labeling complementary DNAs (cDNAs) from two different samples with two different fluorochromes, such as Cy3 or Cy5. Another example is the use of two-color fluorescence to delineate the morphological events of tumor-induced angiogenesis and other tumor-induced host processes.¹⁴

5.2.3 Hyperspectral fluorescence imaging

The fluorescence spectra of the images captured by a multicolor fluorescence imaging system are discrete and limited by the number of images. Hyperspectral fluorescence imaging provides the opportunity to obtain the fluorescence emission spectrum at each point.

Except for point-scanning fluorescence spectroscopy, the fluorescence hyperspectral images can be obtained through two other approaches: a pushbroom scanning system^{15,16} and a Fourier transform spectroscopic imaging system.¹⁷ In a pushbroom system, a thin strip of the sample is illuminated by the excitation light, the induced fluorescence is collected, and the spectrum of the emitted fluorescence is analyzed through an imaging spectrometer. A Fourier transform spectroscopic imaging system uses an imaging Fourier transform spectrometer to obtain the full spectrum of the entire fluorescence image without any scanning across the target.

5.2.4 Fluorescence lifetime imaging

The imaging techniques discussed above only use integrated fluorescence intensities without considering fluorescence decay time. Because of scattering, conventional fluorescence imaging becomes increasingly difficult when imaging into greater depths. Fluorescence lifetime imaging, which records the lifetime of the fluorescence signal rather than the intensity, minimizes the effect of light scattering in thick layers of tissue.¹⁸ It is possible to probe greater tissue depths compared to the conventional fluorescence imaging. Fluorescence lifetime imaging also offers the potential for improved fluorophore specificity, and it can report on the environment of the fluorophore on the basis of changes in the radiative or nonradiative decay constants.

There are two approaches for fluorescence lifetime imaging: time domain¹⁹ and frequency domain.²⁰ With time-domain fluorescence lifetime imaging, the lifetime is usually measured in nanoseconds by a laser with a repetitive pulse duration of a few hundred femtoseconds or picoseconds and a nanosecond shutter because the lifetime of an excitation state is typically 1–20 ns. Fluorescence lifetime images are usually shown in pseudocolor according to their lifetimes. In frequency-domain fluorescence lifetime imaging, the lifetime is calculated by measuring the phase shift of fluorescence and the reduction in its amplitude using a detector with a gain modulator when the continuous-wave excitation light source is modulated or using the frequency-domain information present in a train of optical excitation pulses.

5.3 Components in Fluorescence Imaging Systems

As shown in Fig. 5.3, most fluorescence imaging systems have the following key elements: a light source to excite fluorescence, illumination optics to deliver the excitation light, light collection or imaging optics to collect the fluorescence signal, fluorescence filters to block excitation photons, and a detector to capture

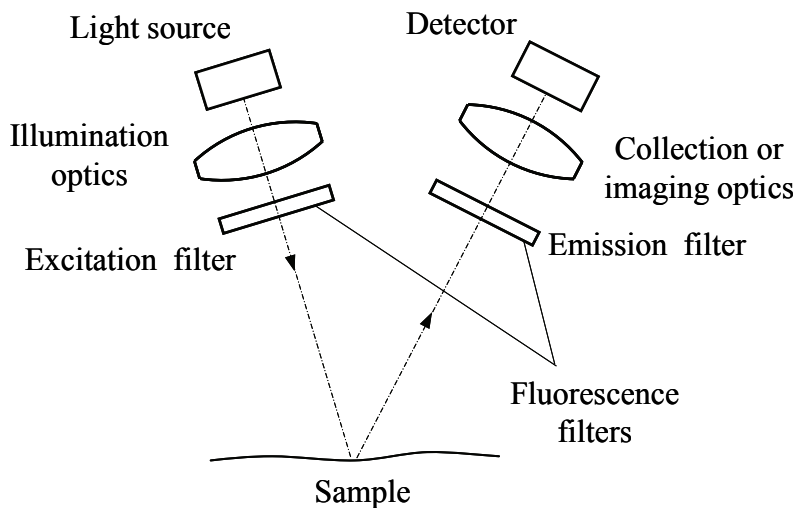


Figure 5.3 Components of a general fluorescence imaging system.

the fluorescence signal. Fluorescence filters include excitation filters, emission filters, and dichroic beamsplitters.

An effective fluorescence imaging system requires careful consideration of the light source, the excitation/emission filters, and the detector, as well as the illumination and detection optics, to optimize the output from the individual probe. In this section, these components will be discussed in detail.

5.3.1 Light source

The most important requirement of the light source in fluorescence imaging is that the wavelength of the light source closely matches the excitation wavelength of the fluorochrome to achieve a high-contrast image. The wavelength of the light source determines which type of fluorophore can be excited and how well that fluorophore can be excited. Any wavelength away from the peak of the excitation spectrum will produce background noise that can overshadow the signal emitted by the object of interest. Wavelength is also the major factor in determining the probed depth given that the penetration depth of light in tissue strongly depends on the wavelength.

Light-source intensity is the other major factor in fluorescence imaging; it determines how much excitation light can reach the sample. Most fluorophores have low quantum yields; therefore, powerful light sources are needed to excite enough fluorescent light to be detected by the eye or other detectors, such as a photomultiplier tube (PMT), avalanche photodiodes (APDs), a CCD, or a CMOS sensor. The human eye is less sensitive than most detection systems, and thus, applications involving visual observation require higher levels of illumination intensity.

There are two types of light sources used in fluorescence imaging: noncoherent and coherent. Noncoherent light sources are usually broadband and

are typically used in fluorescence spectrometers and area imaging systems. The spectral output of noncoherent light sources can be tuned to a narrow band of excitation light or a predefined spectrum by using gratings, filters, and spatial light modulators, such as liquid crystal devices (LCDs) or digital micromirror devices (DMDs).²¹

The most common noncoherent light sources in fluorescence imaging are mercury and xenon arc lamps. Historically, mercury high-pressure arc lamps (commonly referred to as HBO lamps) were the most prevalent light source for fluorescence imaging. The light intensity of a mercury lamp is not evenly distributed from UV to near-IR light, with peaks of intensity at 313, 334, 365, 406, 435, 546, and 578 nm; between the peaks, the excitation energy of the light source is very low. If these lines coincide with or are close to the peak of the excitation spectrum, a mercury lamp is the optimal light source. HBO lamps are suitable for applications that need blue or UV light to excite fluorescence. They typically have a short life span (~300 hours), necessitating relatively frequent bulb changes.

Xenon arc (XBO) lamps have a relatively flat emission spectrum, without emission peaks across the visible spectrum. The even emission levels and lower fluctuations make XBO lamps better suited to applications in ratio imaging and other quantitative applications. XBO lamps also have the advantage that they have stable emission intensity over time. This long stability time relates to the relatively long life span of XBO lamps (400–2000 hours, depending on the bulb).

The xenon-mercury arc lamp has the best characteristics of both xenon lamps and high-pressure mercury lamps. The spectral distribution of a mercury-xenon lamp includes a continuous spectrum from the UV to the IR and the strong mercury line spectrum. It has a bluish-white spectrum and extremely high UV output. The xenon-mercury arc lamp is an ideal source for some applications because it provides extremely high excitation energy.

Metal halide (HXP) white-light sources have an emission spectrum similar to that of mercury lamps. HXP lamps have an emission output featuring pressure-broadened versions of the prominent mercury arc spectral lines in addition to higher radiation levels in the continuous regions between lines. Therefore, HXP lamps usually produce much brighter images of fluorophores, which have absorption bands in the spectral regions between the mercury lines. Commercial HXP lamps have extended arc-lamp lifespans (up to 2000 hours).

To eliminate traditional alignment problems and provide uniform illumination across the entire FOV, a liquid light guide or fiber light guide is often used to couple the light out from the arc lamp. The light guide acts as a scrambler to homogenize the arc output and create an even field of illumination. The light guide also has the advantage of reducing heat transfer from the light source to the sample.

One advantage of arc lamps is that one light source can be used to excite several fluorophores independently or to excite several fluorophores simultaneously. They can also be used as a light source for white-light illumination. One disadvantage of using arc lamps is that one or more filters are

required to select the spectrum for optimal fluorescence excitation. There are several other issues related to arc lamps. The bulbs may explode because of high working temperatures and pressures, although the possibility of explosion is low. For all gas-discharge lamps, reaching optimal working conditions usually takes time, and restarting them after they are turned off requires a certain period of time. The light intensity may fluctuate and drop throughout the lifetime of the lamp.

The laser, the only coherent light source, is monochromatic and is often used in fluorescence scanning systems. Lasers produce a highly coherent light, which can be collimated, expanded, and focused to a diffraction-limited spot. The laser beam can be easily coupled into an optical fiber and delivered to a remote site. Many fluorescence scanning devices for biomedical applications use lasers as their excitation light sources, given that the combination of focused energy and a narrow spectrum width contributes to excellent sensitivity and resolution. A drawback is the laser speckle in the illumination area, which is due to the long coherence length of the laser beam.

The following are commonly used lasers in fluorescence imaging. An argon laser generates a variety of wavelengths from UV to visible light, for example, 364, 457, 488, and 514 nm. Argon lasers are often used as the excitation source for many common fluorophores, such as FITC and Cy2. Neodymium:yttrium aluminum garnet (Nd:YAG) solid-state lasers can generate a strong line at 532 nm using a frequency-doubling method, which is useful to excite Cy3. The helium-neon (HeNe) laser at 633 nm is generally used for exciting Cy5. HeNe lasers with other wavelengths have also become available. A relatively new type of laser used in fluorescence imaging is diode lasers. Diode lasers have a wide selection of wavelengths, from UV to near infrared. They are low cost, compact, and lightweight; they can also be directly modulated and integrated into the fluorescence imaging system.

LEDs have been used in fluorescence imaging since the early 1990s. However, their applications are limited because of low efficiency and limited availability of suitable LED products. With recent advances in efficiency and thermal management, LEDs are finding more and more applications in fluorescence imaging. LEDs produce an output with a spectrum bandwidth wider than the laser but much narrower than the arc lamp. The diverse spectra afforded by LEDs make it possible to select an individual LED to supply the optimum illumination wavelength for fluorescence imaging. Although LEDs are regarded as narrowband light sources, there are tails in the emission spectrum that must be removed by appropriate emission optical filters to minimize background noise.

LEDs have several advantages over other noncoherent light sources, including their compact size, low power consumption, low heat generation, fast switching, high emission stability, and extremely long life spans. LEDs can instantly illuminate at full intensity as soon as the electrical current is applied. LEDs can also be switched on or off instantly with no deleterious effects to their lifespan. The controllable stability of LED intensity is the key benefit for

fluorescence imaging when consistent and repeatable measurement is required for quantitative analyses.

The most important consideration in selecting the light source for fluorescence imaging is whether the spectral intensity peaks of the lamp match the excitation wavelengths of the fluorophores in the sample. The light source intensity is the second important consideration. Other factors include light-induced tissue damage and efficient delivery of light from the light source to the tissue.

5.3.2 Fluorescence filters

Fluorescence filters are essential in fluorescence imaging systems. Without filters, the detector would not be able to distinguish between the desired fluorescence from the scattered excitation light and the autofluorescence from the sample, substrate, and other optics in the system. Generally, one or more of the following filters are necessary in a fluorescence imaging system:

- (1) Excitation filter to select the optimal excitation wavelength from a light source,
- (2) Emission filter to specifically select the emission wavelength of the light emitted from the sample and to block the excitation light, or
- (3) Dichroic beamsplitter to separate the excitation and emission light.

To prevent excitation light from passing through the emission filter and reaching the detector, the passbands of the excitation and emission filters must not overlap.

5.3.3 Illumination optics

For simple fluorescence imaging systems, the excitation light is delivered directly to the sample from a light source without any additional optics. For most fluorescence imaging systems, illumination optics is required in order to deliver the excitation light to the sample more efficiently and uniformly. Two basic requirements of illumination optics are efficiency and uniformity. The fluorescence signal is usually very weak; therefore, the efficiency of the light delivery system becomes much more important in fluorescence imaging systems compared to other imaging systems. Uniform illumination is required because the fluorescence signal at each point is proportional to the illumination light at that location.

Ideally, the optical materials used in the illumination path should not emit autofluorescence, although autofluorescence can be minimized by placing an excitation filter as the last element in the illumination path.

5.3.4 Light collection/imaging optics

Light collection/imaging optics collects and delivers the fluorescent light to the sensor. Because the fluorescence signal is weak, the light collection optics should have a high light-collection efficiency. For imaging optics, in addition to light-

collection efficiency, which is determined by the numerical aperture, there should be adequate aberration control to obtain 2D fluorescence images. Section 5.6 will discuss the imaging optics in detail.

5.3.5 Detector

The detector is one of the most critical components in fluorescence imaging because it determines at what level the fluorescence signal can be detected, what relevant structures can be resolved, and the dynamics of the process that is visualized and recorded. Through the detector, the fluorescent light is converted into electrical energy, thereby producing a measurable signal that is proportional to the number of photons detected. The measured, continuous analogue signal is then converted to a discrete digital signal. Given that there is a limited number of intensity conversion levels, 8-, 10-, 12-, or 16-bit, it is unavoidable that this conversion process introduces a certain amount of error. Because only a small fraction of the fluorescence from a sample is collected by the emission optics, it is essential to have a sensitive and stable detector with a good response over the emission spectrum.

Three commonly used light detectors are the photodiode, PMT, and solid-state imaging sensor. The first two employ a photosensitive surface that captures incident photons and generates an integrated electronic signal, but they lack spatial discrimination. PMTs can be used to detect the incoming photons at high speed because they do not store a charge and can respond to the changes of the input light within a few nanoseconds. They typically generate a low noise signal and have a large dynamic range to accurately reflect the photon flux. PMTs also have large gain without sacrificing bandwidth. Miniature PMTs have recently become available for compact or portable systems.

Silicon photodiodes and avalanche photodiodes (APDs) also respond very quickly to incoming light but without large gain. Compared to PMTs, the spectral sensitivity of photodiodes is relatively flat, and the QE is relatively high over the entire visible spectrum. The major issue with photodiodes is that they produce a considerable amount of noise, resulting in a low SNR under photon-limited conditions, which is the case for fluorescence imaging.

Solid-state imaging sensors consist of a dense matrix of photodiodes incorporating charge-storage regions. CCD and CMOS detectors are two solid-state detectors commonly used in fluorescence imaging. One advantage of these detectors is that their optimum sensitivity is higher at longer wavelengths (700–800 nm) than that of most PMTs, so they are well suited to near-IR fluorescence detection. Solid-state sensors can be characterized by a number of variables, including QE, dynamic range, uniformity, SNR, response speed, and spatial resolution.

QE is a measure of the detector's ability to produce an electronic charge from the incident photons that are detected and is a function of the wavelength. QE is especially important for low-light imaging applications.

Dynamic range is a measure of the maximum and minimum intensities that can be simultaneously detected. Dynamic range is the ratio of the maximum

signal that can be accumulated to the minimum signal that is equivalent to the noise associated with reading the minimum signal.

The uniformity of the response of solid-state imaging sensors has to deal with gain variations across the sensor, regional differences in noise and sampling efficiency, and spatial variation in the efficiency of light collection. The response speed is described by its lag, representing the fraction of the previous image that carries over into the next image within a prescribed time interval. The SNR characterizes the quality of a measurement and determines the ultimate performance of the system.

The resolution of a solid-state sensor is a function of the number of photodiodes and their size relative to the image. The resolution of the entire system is determined by the sensor and the imaging system.

The CCD is the most widely used imaging sensor for fluorescence imaging. While a CCD sensor is usually operated at room temperature, the performance can be significantly enhanced by cooling the sensor to reduce the dark current. Although the CCD consists of silicon photodiodes, its spectral sensitivity is different from that of a simple silicon photodiode detector because the CCD surface has channels used for charge transfer that are shielded by polysilicon gate electrodes. These structures absorb the shorter wavelengths, resulting in a lower blue sensitivity of the device. These losses can be eliminated in the back-illuminated CCD, where light falls onto the back of the CCD in a thinned and transparent region.

Slow-scan CCDs are designed specifically for low-light applications. They have very low background and low noise, allowing for long image-acquisition times. The readout and thermal noise are minimized through slow readout rates and cooling. Slow-scan CCD cameras are generally limited in their frame rates, and the SNR is poor when exposure times are short, unless the specimen has extremely bright fluorescence.

The intensified CCD (ICCD) uses an image intensifier that is fiber-optically coupled to the CCD chip to increase the sensitivity up to a single-photon level. The image intensifier converts incoming photons into electrons at the photocathode, multiplies the electrons with high-voltage acceleration through a microchannel plate, then reconverts the multiplied electrons back to photons through a phosphor-coated window. The photons emitted at the phosphor-coated window are then projected onto a CCD through a fiber-optic plate. Intensified CCD cameras have very fast response times, and the CCD camera readout is the slowest step in the image acquisition. ICCD cameras are frequently used in studying dynamic events and for ratio imaging of ion-sensitive fluorochromes.

To meet the requirements for applications that demand rapid frame-rate capture at extremely low light levels, the electron-multiplying charge-coupled device (EMCCD), in which a gain register is placed between the shift register and the output amplifier, has been developed to amplify weak signals above the read noise floor of the CCD. EMCCD cameras employ an on-chip amplification mechanism to multiply the photoelectrons that are generated in the silicon. As such, the signal from a single photon event can be amplified above the read noise

floor at fast readout speeds. A significant advantage of EMCCDs is that they are less expensive to manufacture than intensified CCDs because the signal amplification stage is incorporated directly into the CCD structure.

A CMOS sensor is a low-cost, compact, versatile detector with the virtues of silicon detection but without the problems of charge transfer. Rapid developments in CMOS sensors put them in a potentially important future role in fluorescence imaging. CMOS sensors have an amplifier and digitizer associated with each photodiode in an integrated on-chip format, which allows manipulation of individual photodiodes, user-selectable ROI readout, high-speed sampling, electronic shuttering, and exposure control.

A new type of CMOS sensor has been developed in recent years: the smart CMOS image sensor, which is a CMOS image sensor that features integrated smart functions (e.g., computation function) on the chip. Smart CMOS sensors bring the benefits of integrated functions and miniaturization. These advantages are further enhanced when combined with on-chip detection.^{22–24}

No single detector will meet all requirements in fluorescence imaging; some trade-offs between the requirements are inevitable. For example, when time is available for image integration, a slow-scan CCD usually outperforms an intensified camera in all areas because of its higher QE and lower noise. However, when time is a critical parameter, an intensified CCD is often the only choice. With future developments, CMOS sensors may replace the CCD in many fluorescence imaging systems.

5.4 Fluorescence Filters

The goals in developing fluorescence imaging systems are to excite the fluorophore as efficiently as possible and to collect as much emitted fluorescent light as possible while blocking as much excitation light as possible in the detection channel. Fluorescence filters are the key components to achieving these goals. The challenge is that the excitation light is usually about 10^4 – 10^6 times brighter than the fluorescence emission.

5.4.1 Types of fluorescence filters

Figure 5.4 shows four types of filters commonly used in fluorescence imaging: a shortpass filter, a bandpass filter, a longpass filter, and a dichroic beamsplitter. The bandpass filter is designed to transmit a discrete band of wavelengths in a specified spectral range, as shown in Fig. 5.4(a). The bandpass filter is characterized by its center wavelength, bandwidth, average passband transmission, cut-on and cut-off wavelength, and slope. The center wavelength is the mean of the wavelengths at 50% of peak transmission. The bandwidth is defined as the full width at half maximum (FWHM) of the passband. The average passband transmission is the average transmission of light passing through the filter within the passband or transmission band. The average passband transmission is usually used to specify the bandpass filter. Some filters also specify the transmission of the reflection band.

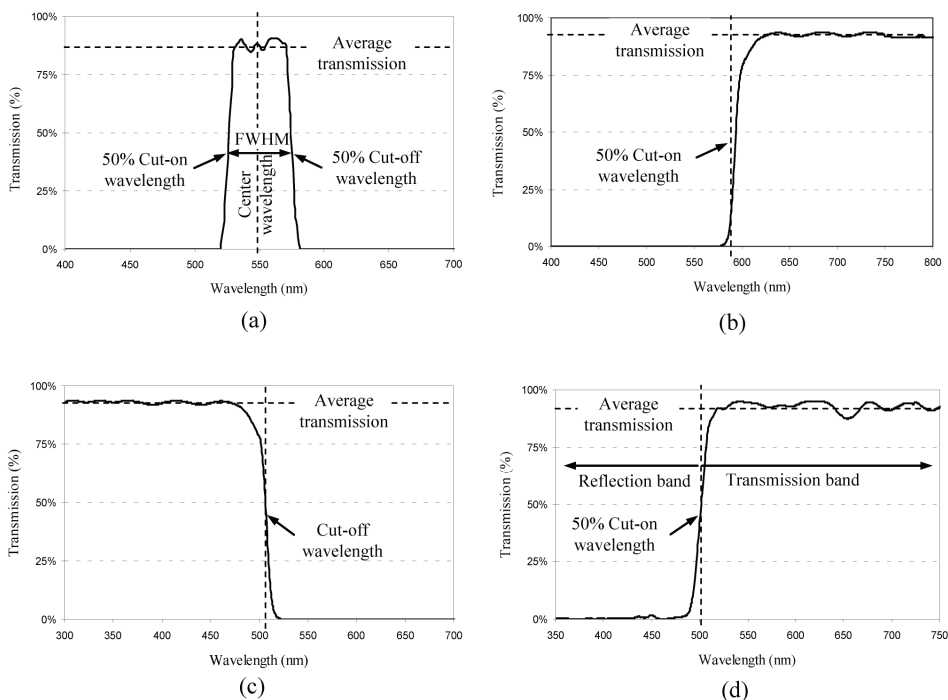


Figure 5.4 Types of filters and their specifications: (a) bandpass filter, (b) longpass filter, (c) shortpass filter, and (d) 45-deg dichroic beamsplitter.

As indicated in the figure, wavelengths of 50% peak transmission are commonly used as cut-on and cut-off wavelengths. The cut-off wavelength defines the transition from the region of high transmission to the adjacent spectral region of low transmission as the wavelength increases. Similarly, the cut-on wavelength defines the transition from the region of low transmission to the adjacent spectral region of high transmission as the wavelength increases.

One important commonly used parameter not shown in Fig. 5.4 is slope, which describes the sharpness of the transition from transmission to reflection or from reflection to transmission.

Longpass filters are designed to transmit light with wavelengths longer than the cut-on wavelengths. Light below the cut-on wavelength is blocked or attenuated. Shortpass filters, on the other hand, transmit light with wavelengths shorter than the cut-off wavelength and block the light above the cut-off wavelength. When these filters have a very sharp transition slope, they are often called edge filters. The specifications for longpass and shortpass filters are the same as for bandpass filter.

Dichroic beamsplitters, also called dichroic mirrors, are usually designed for a 45-deg angle of incidence to separate beams of two different spectral ranges. Dichroic beamsplitters serve a dual function by directing the excitation light to the sample and passing the emission light to the detector. The beamsplitters can be either longpass or shortpass, but a longpass beamsplitter is preferred because

there is generally less passband “ringing” and fewer problems with harmonic effects. In addition to the cut-on wavelength, the reflection band and transmission band are also specified. In general, the dichroic beamsplitter does not reflect all wavelengths below the cut-on wavelength with adequate efficiency. Similarly, the transmission characteristics are not uniform on the long-wavelength side where transmission typically either falls off at longer wavelengths or exhibits other irregularities in the deep red and near-IR regions.

The dichroic beamsplitter can be either a flat plate or a cube. For the dichroic beamsplitter cube, a thin film is coated on the hypotenuse of the right-angle prism that is cemented to a second right-angle prism to form a cube. The advantage of a beamsplitter cube is that there is no spatial shift for the transmitted beam.

Figure 5.4 shows the transmission profiles of the filters only in the region close to the cut-on or cut-off wavelength. Typically, the transmission outside the wavelength range is not a simple extension; it is more complicated and depends on the type of filter, as shown in Fig. 5.5, where the transmission spectrum of a bandpass filter is plotted from 270–1200 nm. There are some spikes in the UV and near-IR regions. Additional filters are necessary to block UV light if those spikes would be harmful to the fluorescence imaging.

There are two major types of filter technologies used in fluorescence imaging: absorption filters and thin-film filters. Other filters infrequently used in fluorescence imaging include acousto-optical tunable filters, liquid crystal tunable filters, and holographic filters.

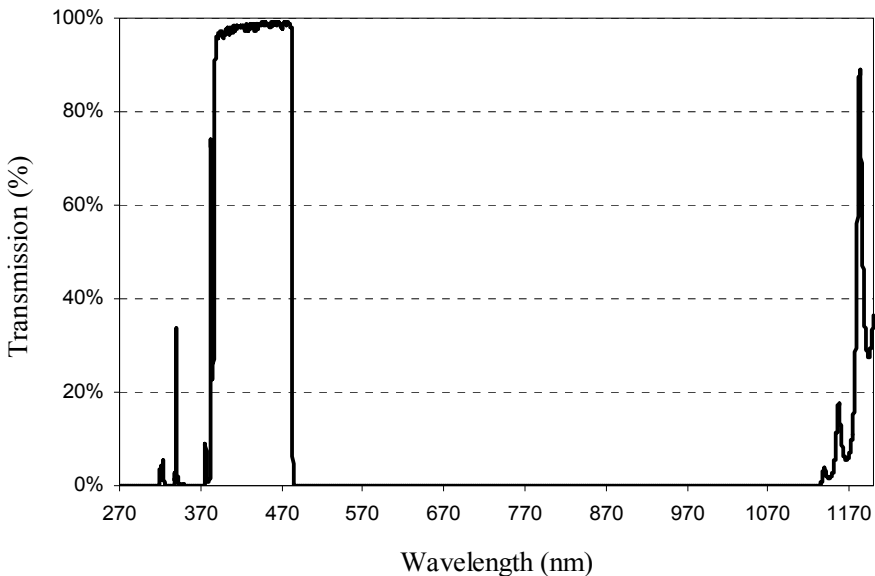


Figure 5.5 Transmission spectrum of a shortpass filter.

5.4.2 Absorption filters

Absorption filters are made primarily from colored filter glass or synthetic gels and are widely used in applications that do not require a precise definition of transmitted wavelengths. Colored filter glasses consist of a base glass and colorants, which are added to the base glass. Typical base glasses are made of potassium, sodium, phosphate, borate, borosilicate, and silicate. The colorants may take the form of metal ions in the base glass (e.g., Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Y, Zr, Mo, La, Ce, Pr, Nd, W, U). They may consist of metal atoms (e.g., Au, Ag, Cu, Pt), and nonmetallic elements or their compounds (e.g., S, sulfides, Se, selenides, Te, P) that are suspended in the glass in the form of submicroscopic crystals. The coloration of the filter may also be due to the base glass alone. Most of the oxides used in glass manufacturing have absorption bands in the UV.²⁵ These materials absorb photons that have energy larger than the bandgap; they do not interact with photons having less energy than the bandgap. The spectral properties of the colored filter glasses depend on the combination of the colorant and the base glass. The cut-off wavelength of colored filter glass can be tuned by choosing different colorants.

When colored filter glass absorbs light, the electron in the colorant is excited to a higher-energy and vibrational state before returning to the ground state by emitting fluorescence. The wavelength of the maximum intensity in the emission spectrum is close to the cut-off wavelength for optical absorption. For example, when Schott glass OC515 is illuminated with UV light at a wavelength of 390 nm, it emits autofluorescence with a maximum wavelength of 540 nm. The QE is about 9.7%.²⁶ Fluorescence generated inside the colored filter glass is distributed equally in all directions. Therefore, these filters should be placed away from the image planes.

Compared to other types of filters, colored filter glass has several advantages. It attenuates light solely by absorption, so the spectral performance is dependent only on the physical thickness of the glass, and it is not sensitive to the angle of incidence except for minor changes in absorption, which are due to an increased effective thickness. For the same reason, the performance is not prone to destruction by minor scratches or abrasions. Colored filter glasses are usually stable and long-lived under normal conditions and have good resistance to chemical attack, fumes, and contamination.

There are also several disadvantages of using colored filter glasses for fluorescence imaging applications. Their transition slopes are usually poor, especially for bandpass filters, and their average transmission is relatively low. Their light transmission properties may also change with prolonged use. Most of the longpass filter glasses have high autofluorescence, which can be a serious issue in fluorescence imaging. One more disadvantage is that there is a limited selection of optical-grade colored filters available for imaging applications.

Colored filter glass can be used in the illumination path to block the UV content of the spectrum from the light source. It can also be used with narrowband optical filters to remove unwanted short-wavelength light.

5.4.3 Thin-film filters

A thin-film filter is composed of a substrate onto which a stack of thin layers of different dielectric materials, with high and low refractive indices, are deposited. The thickness of each layer is about one quarter of the wavelength. Each material is intrinsically colorless and has low light absorption. The reflection and transmission of a thin-film filter is the result of light interference, rather than absorption, between the waves reflected or transmitted from each interface, as shown in Fig. 5.6.

The performance of the thin-film filter is dependent on the layer structure and the materials used in each layer. Theoretically, thin-film filters can provide high transmission over any bandwidth and high blocking of undesired light over the remainder of the wavelength range.

Traditionally, the materials used for high-index layers are zinc sulfide and zinc selenide, whereas sodium aluminum fluoride is used for low-index material. These materials are heated in a chamber until they vaporize and diffuse onto a glass substrate. One of the disadvantages of evaporative coatings, often called soft coatings, is that these materials are both soft and hygroscopic. Therefore, air or water could get into the coating. The coating must be sealed to ward off the undesired effects of moisture on performance and lifetime by laminating the coated surfaces between two glass substrates. In addition, zinc and sodium aluminum fluoride suffer from low-transmission characteristics and temperature instability.

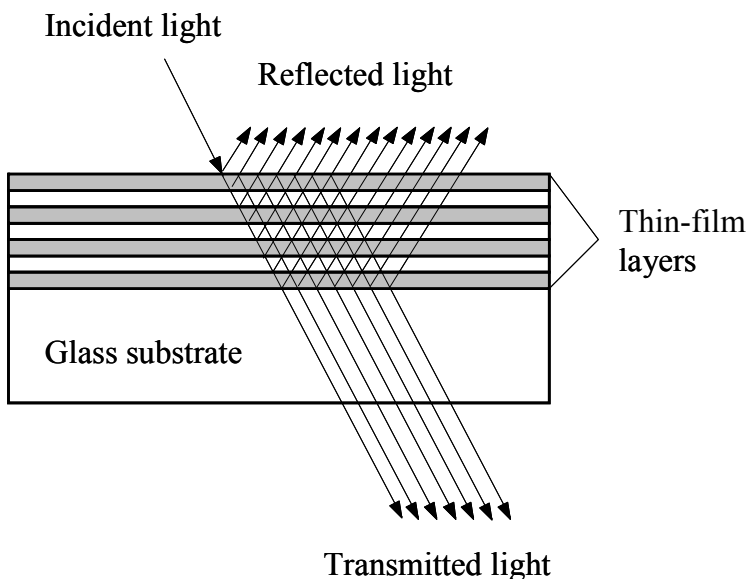


Figure 5.6 Schematic diagram of a plane wave incident onto a stack of thin-film layers. The reflectance is the result of interference between all of the reflected waves, and the transmittance is the result of the interference between all of the transmitted waves.

Figure 5.7(a) shows the structure of a complete traditional thin-film filter. It is a combination of a multilayer bandpass filter and a multilayer blocking filter. The two filters are held together in a protective metal case using an optical epoxy. The blocking filter, which is the thin-film coating applied onto the second substrate, is used to block the transmission of wavelengths farther away from the passband of interest.

A multilayer bandpass filter consists of a number of cavities; each cavity has two reflecting layers separated by a spacer. The thickness of the spacer is an even multiple of a half-wavelength at the desired peak transmission. The reflecting layers consist of several thin-film layers, each of which is a quarter of a wavelength thick. The number of layers in the reflecting layer is adjusted to tailor the width of the bandpass, while the number of the cavities in the bandpass filter is used to tune the slope of the cut-on and cut-off wavelength transmission boundaries. During manufacturing, it is difficult to make all of the flat surfaces parallel to each other, resulting in a wedge angle and, therefore, a pixel shift when light passes through them. In the meantime, the epoxied interfaces may introduce undesirable scattering and autofluorescence.

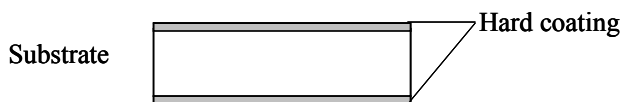
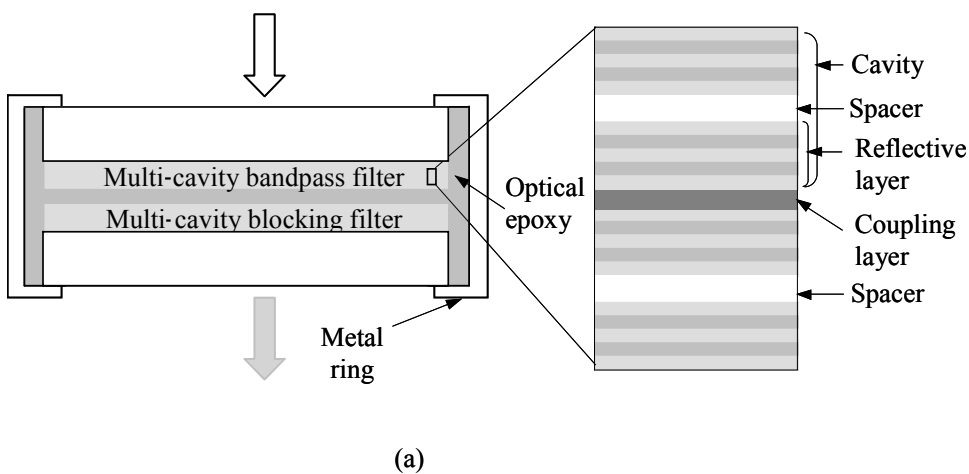


Figure 5.7 Diagrams of (a) a complete traditional thin-film filter and (b) a hard-coated thin-film filter.

Recently, a new coating technology called ion-beam sputtering (IBS) has been developed to deposit semitransparent layers of metal oxides onto the two surfaces of a glass substrate, as shown in Fig. 5.7(b). This new coating is sometimes called a hard coating because the oxide materials are as hard as the glass substrate upon which they are coated. The deposition of metal oxides can be tightly controlled by in-process, optical-transmission monitoring for high- and low-index layer thicknesses. Therefore, denser and more uniform layers can be coated more precisely on a single piece of glass. With an IBS technique, it is possible to deposit many more layers to achieve a sharper edge, greater passband transmission, and deep blocking over the spectral range required for fluorescence applications. Hard-coated filters do not require sealing, nor do they require special handling. Their performance is not affected by water and temperature changes. Because only one substrate is used, epoxies are eliminated, and the wedge error is reduced dramatically, as are scattering and autofluorescence.

Figure 5.8(a) compares the performances of traditional soft-coated and hard-coated bandpass filters. The hard-coated filter has better transmission and sharper edges. More complex filters, such as multiband bandpass filters and multiedge dichroic beamsplitters, are also possible, as shown in Fig. 5.8(b). Multiband filters improve the brightness and sensitivity of multicolor fluorescence imaging systems.

Thin-film filters have several performance limitations; for example, interference coatings are sensitive to the angle of incidence. Most interference coatings are designed for collimated light and normal incidence, except dichroic beamsplitters. For light entering the thin-film layers at an angle, instead of perpendicularly, it is necessary to have a shorter wavelength in order for the optical path between two interfaces to be exactly $\lambda/2$ or $\lambda/4$. In other words, the cut-on and cut-off wavelengths of all types of thin-film filters decrease as the angle of incidence increases. Figure 5.9 shows the “blue-shifted” wavelength with an increased angle of incidence. The shift of almost any spectral feature can be approximately quantified by

$$\lambda(\theta) = \lambda_0 \sqrt{1 - \left(\frac{n_0}{n_{\text{eff}}} \right)^2 \sin^2 \theta}, \quad (5.2)$$

where λ = wavelength at the angle of incidence, λ_0 = wavelength at normal incidence, θ = angle of incidence, n_0 = refractive index of external medium, and n_{eff} = effective refractive index of the filter.

The effective refractive index n_{eff} is determined by the coating materials and the sequence of thin-film layers in the coating. For thin films using zinc sulfide and cryolite as coating materials, the effective refractive index values are typically 1.45 or 2.0, depending on whether the spacer layer is composed of the low- or high-refractive-index material. The effective refractive index is also polarization dependent.

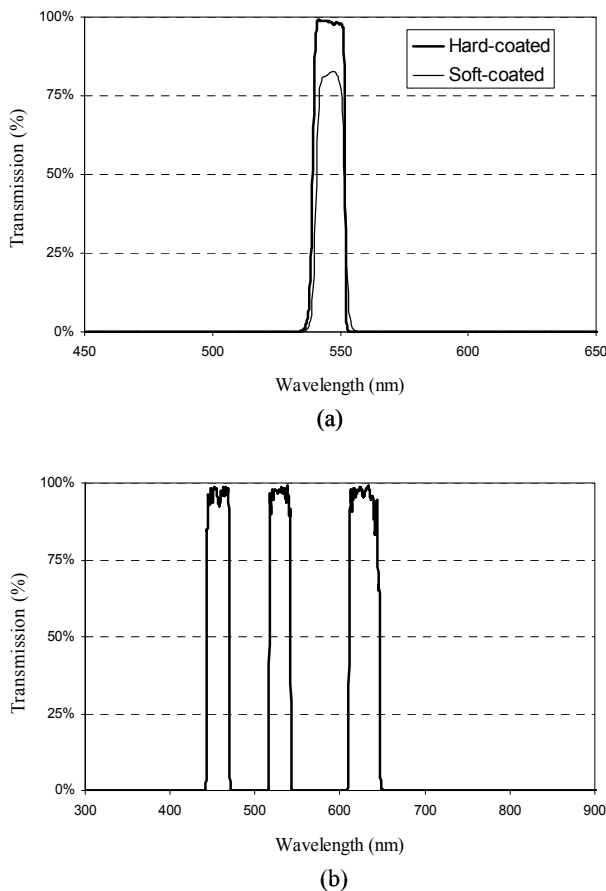


Figure 5.8 (a) Comparison of the traditional soft-coated and hard-coated bandpass filters and (b) transmission of a multiband hard-coated bandpass filter. (Courtesy of Semrock, Inc.)

The second effect that is seen when the angle of incidence is greater than the normal incidence is polarization splitting, as shown in Fig. 5.9. The blue shift for p -polarized light is more significant than that for s -polarized light. This polarization splitting increases as the angle of incidence increases. When necessary, coating designs can minimize polarization effects but not entirely. The bandwidth of s -polarized light becomes narrower than that of p -polarized light as the incident angle increases.

The third effect related to the angle of incidence is the distortion of the transmitted profiles, as clearly shown in Fig. 5.9. As the incident angle increases, the distortion becomes larger.

The fourth effect is that some spikes occur in the blocking band, for example, the spikes between 525 and 600 nm in Fig. 5.9. The increase in the incident angle strengthens the spikes.

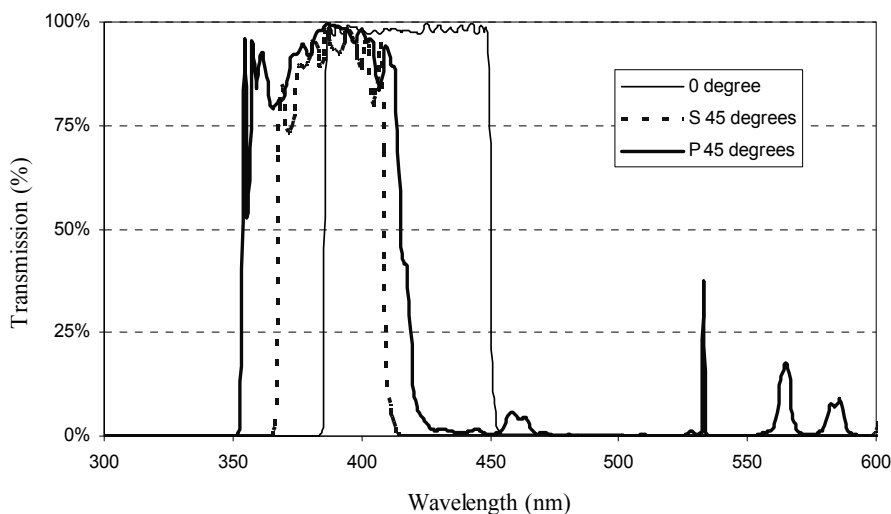


Figure 5.9 Filter performance and angle of incidence. As the angle of incidence increases, the cut-on and cut-off wavelengths for p- and s-polarized light split and shift to shorter wavelengths. (Courtesy of Semrock, Inc.)

The characteristic performance holds only within a finite wavelength range. Additional filters, such as colored glass and wideband blockers, are necessary to extend the blocking wavelength range; however, the peak transmission is reduced with additional filters.

The coating materials used in thin-film filters are limited in the range of transparency. Outside of this range, coatings may become highly absorbing rather than highly transmissive or reflective. As a result, UV filters usually have limited performance and less design flexibility.

5.4.4 Specifications of thin-film filters

In addition to the key performance-related specifications discussed in Sec. 5.3.1, which are the center wavelength, average passband transmission, bandwidth, edge sharpness, and cut-on/cut-off wavelengths, the optical quality specifications of thin-film filters include surface flatness, wedge, transmitted wavefront distortion, scratches and digs, pinholes, and clear aperture.²⁷ The diagrams in Fig. 5.10 show the effects of surface flatness, material inhomogeneity, and wedge on the wavefront of the output beam.

Surface flatness is a measure of the deviation of the surface from a perfectly flat plane. It is usually specified in fractions or multiples of a wavelength. The wavefront distortion upon reflection from the surface is $2/\cos(\theta)$ times the surface flatness, as shown in Fig. 5.10(a), where θ is the angle between the normal of the filter and the optical axis. For example, when the beam reflects at

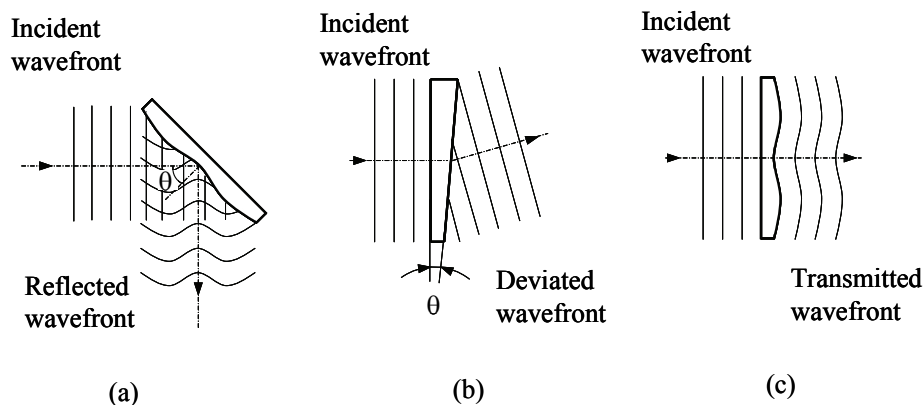


Figure 5.10 Effects of surface flatness and wedge on (a) reflected wavefront, (b) deviated wavefront, and (c) transmitted wavefront.

normal incidence, the wavefront distortion is twice the surface flatness because θ equals zero.

Wedge, measured in arcminutes or arcseconds, defines the degree to which the two outer surfaces of the thin-film filter are parallel. The wedge, mainly from the substrate, causes image shift, as shown in Fig. 5.10(b). For a typical filter with constant refractive index, the magnitude of deviation is approximately one-half the wedge angle.

Transmitted wavefront distortion, measured in fractions or multiples of a wavelength, is the distortion of a plane wave when transmitted through a filter. Transmitted wavefront distortion is the combined effect of surface flatness and the inhomogeneity of the refractive index, as shown in Fig. 5.10(c).

Scratches and digs specify the surface quality of a thin-film filter. A scratch is a defect on an optical surface whose length is many times its width. A dig is a defect nearly equal in terms of its length and width. A comprehensive surface-quality specification typically is stated as a scratch-dig combination, such as 60-40. Scratches and digs originate from either improper handling of filters or improper preparation of substrates.

Pinholes are small breaks in the coating of a thin-film filter. They are usually caused by dust particles on the substrate. Typically, pinholes occupy a relatively small area of the total coated surface. The significance of pinholes in a filter is highly dependent on the application of the filter.

Clear aperture is the maximum usable diameter of a thin-film filter. It should be large enough that it does not restrict the overall aperture of the imaging system, and there should be no leakage of the unfiltered light around the edge of the clear aperture. The requirements for different fluorescence filters will be discussed in the next section.

5.4.5 Filters in fluorescence imaging

5.4.5.1 Requirements for filters in fluorescence imaging

Generally, there are three filters in a fluorescence imaging system: an excitation filter, a dichroic beamsplitter, and an emission filter. Figure 5.11 shows the basic configuration of fluorescence filters for fluorescence imaging. The thickly dashed box in Fig. 5.11 indicates the illumination path, and the thinly dashed box is the detection path. The excitation filter selects a range of wavelengths from a light source to excite the sample. The emission filter transmits the emission light and rejects the excitation light. A dichroic beamsplitter serves a dual function, reflecting the excitation light to the sample and transmitting the emission light to the detector or the eyepiece. For a laser-based fluorescence system, an excitation filter is not required given that the laser usually has a very narrow bandwidth. A dichroic beamsplitter is not necessary for transmitted illumination or off-axis illumination.

To obtain the most optimal filter performance, the fluorescence filters should be oriented properly. The dichroic mirror should be oriented such that the reflective coating side faces toward the incoming excitation light in order to avoid a ghost image and image shift. On the other hand, for excitation and emission filters, the coating surface separating the spectrum should not face the incoming light so that it can block any autofluorescence from the coating and substrate.

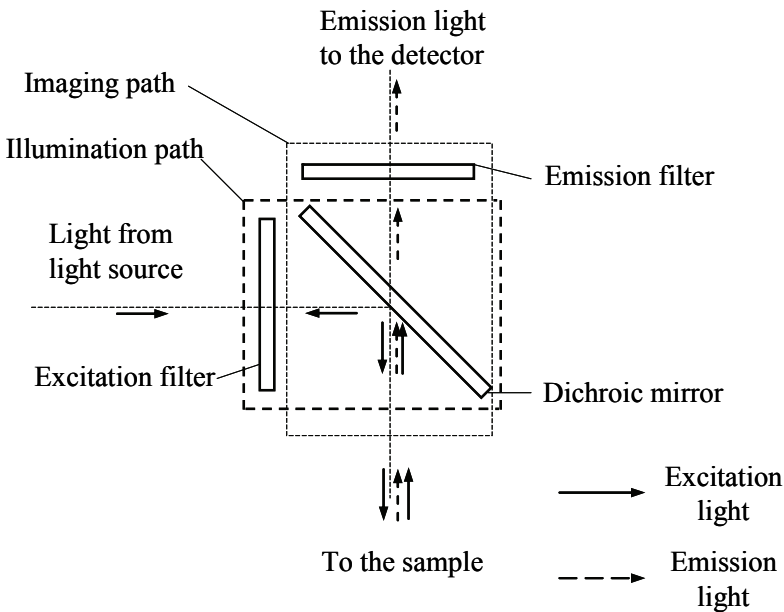


Figure 5.11 The basic configuration of fluorescence filters for fluorescence imaging systems. The thickly dashed box indicates the illumination path, and the thinly dashed box is the detection path.

The illumination system in a fluorescence imaging system should be designed for uniform illumination on the tissue with high efficiency and minimal flare. It is not necessarily designed or preferred for the true imaging of a light source. Therefore, the industry standard for commercial quality is enough for the excitation filters. Specifically, the wedge is less than ± 3 arcmin, the scratch-dig is 80-50, the thickness tolerance is ± 0.15 mm, and the transmitted wavefront error is less than $5\lambda/\text{in}$. When the excitation filter is placed at or near the conjugate plane of the sample, for example, the field diaphragm of the Köhler illumination system, pinholes in the excitation filter must be avoided because leaked light outside the excitation spectrum will be focused on the sample and will cause crosstalk.

The emission filter and dichroic beamsplitter in a detection path generally require an industry standard for precision quality where the wedge is less than ± 1 arcmin, the scratch-dig is 60-40, the thickness tolerance is ± 0.05 mm, and the transmitted wavefront error is less than $1\lambda/\text{inch}$. However, one special case is a point-scanning (nonconfocal) fluorescence imaging system with a PMT, where the sensitive area of the detector is large and the industry standard for commercial quality is enough for optical elements in the detection path.

A dichroic beamsplitter is not required for some fluorescence imaging systems, for example, a system with transmitted illumination or the configuration shown in Fig. 5.3. For a system requiring a dichroic beamsplitter to direct excitation light to the sample, autofluorescence from the dichroic beamsplitter should be minimized because the sample is fully illuminated by the excitation light and is in the detection path.

The above discussion on the requirements for the components in an illumination path usually is not applicable to a laser confocal fluorescence imaging system where the laser beam must remain undistorted in order to achieve the best focus on the sample plane. Both the excitation filter and the dichroic beamsplitter require precision quality.

5.4.5.2 Filter selection

Optical filters must be chosen such that particular fluorophores have the best brightness and contrast performance. The ratio of the emitted fluorescence intensity to the excitation light intensity in a typical application is between 10^{-4} (for very highly fluorescent samples) and 10^{-6} . Therefore, the system must attenuate the excitation light by as much as 10^{-7} (weak fluorescence) without diminishing the fluorescence signal. A good combination of optical filters can reduce both excitation light noise and stray light from the excitation source and other components, as well as instrument autofluorescence outside the emission band.

When choosing an excitation filter, both the spectrum of the light source and the excitation spectrum have to be considered. In many applications, a light source with a wide spectrum, such as a xenon arc lamp, is used for illumination. Both shortpass and bandpass filters can be used in the illumination path to condition the illumination spectrum, but a bandpass filter is usually preferred

because the shortpass filter also passes the UV light, which can cause photobleaching and autofluorescence noise. In contrast, a bandpass filter allows only the excitation light to reach the object. An excitation filter is usually selected to transmit only the light with the bandwidth matching the FWHM of the excitation spectrum. Although highly desired, the peak of the excitation spectrum does not need to exactly match the fluorophore's major absorption peak for efficient excitation.

As discussed earlier, the range and overall shape of the emission profile are the same for any wavelength inside the excitation spectrum. Therefore, it is advantageous to use as much light within the excitation spectrum as possible from the light source, provided that all of the excitation light can be blocked by the longpass or bandpass filter in the detection path. However, allowing a wide range of wavelengths away from the peak of the excitation wavelengths further increases the background noise and may damage the object. The background noise includes the nonspecific fluorescence from the other fluorophores and autofluorescence from other organic matters.

Because the observed background autofluorescence increases roughly in proportion to the bandwidth of the excitation spectrum, narrowing the bandwidth of the excitation filter can reduce background noise and enhance the image contrast; however, this act reduces the image brightness. Therefore, a balance must be struck between the level of the contrast and the brightness for a given application. For many applications, the optimal bandwidth of an excitation filter is approximately 40 nm and is centered on the absorption maximum of the fluorophore.

Emission filters can be either longpass or bandpass filters. A longpass filter may be preferred when the application requires a maximum emission signal and when spectral discrimination is not necessary. The longpass filter transmits fluorescence from all fluorophores with an emission spectrum longer than its cut-on wavelength. It is also useful for simultaneous detection of spectrally distinct multiple emissions.

Bandpass filters should be selected to maximize the SNR for applications where discrimination of signal components is more important than overall image brightness. For a high-resolution fluorescence imaging system, a bandpass filter that transmits a band at or near the emission peak of the fluorophore generally minimizes the background noise and improves the sensitivity and linear range of the measurement. Dual-, triple-, and quadruple-band filter sets make it possible to excite and detect two or more fluorophores simultaneously instead of performing sequential image acquisitions with intervening filter changes.

The dichroic mirror is mounted at a 45-deg angle to the optical axis of the objective lens to reflect light in the excitation band and to transmit light in the emission band. The transmission cut-on of the dichroic mirror lies between the fluorophore's excitation spectrum and its emission spectrum such that the excitation and emission wavelengths are effectively separated. The sharper the slope of the transition from transmission to blocking or from blocking to transmission, the higher the light efficiency and image contrast. To collect as

much fluorescence signal as possible, a shorter cut-on wavelength is usually chosen when the excitation and emission spectra are close and the Stokes shift is small. Since the dichroic beamsplitter is in both illumination and detection paths, the quality requirement should be precision grade. In addition, both coating and substrate materials should have minimal autofluorescence.

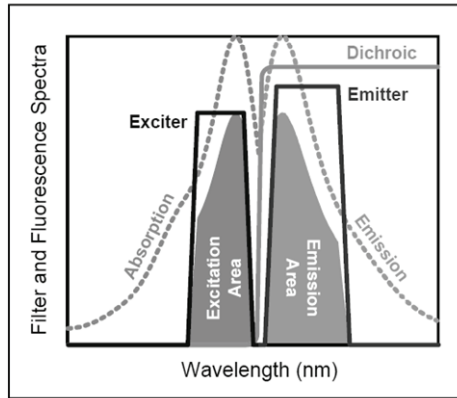
Brightness and contrast are the two criteria used in evaluating a selected set of fluorescence filters. Ideally, the selection of fluorescence filters will maximize both brightness and contrast. However, a trade-off must be made because brightness and contrast are usually contradictory.

The brightness, or signal intensity, is proportional to the product of the “excitation area” and the “emission area.”²⁸ As shown in Fig. 5.12(a), the excitation area is the integrated area under the product of the source spectrum, the excitation filter spectrum, and the fluorophore absorption spectrum, while the emission area is the integrated area under the product of the detector response, the emission filter spectrum, and the fluorophore emission spectrum. From Fig. 5.12(a), it is clear that when the passband of the excitation filter matches the peak of the excitation spectrum and the passband of the emission filter aligns with the peak of the emission spectrum, the fluorescence intensity is maximized. The source spectrum and the detector response are not plotted and are treated as having an ideal value of 1.0 across the entire spectrum in Fig. 5.12(a). When a dichroic mirror is used, the excitation area should include the reflection of the dichroic mirror, and the emission area should multiply the transmission of the dichroic mirror.

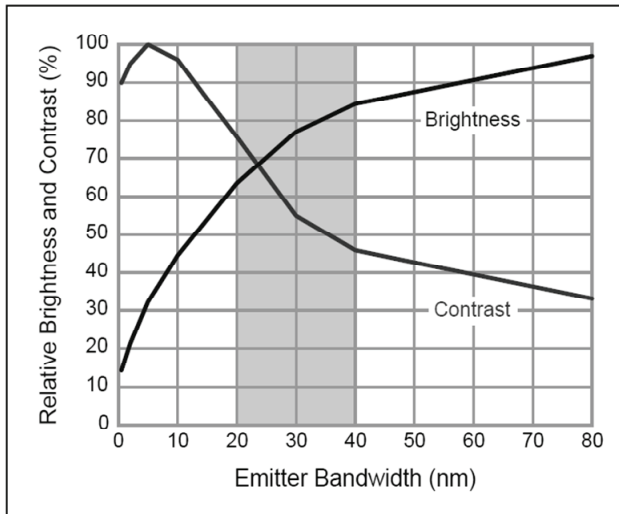
To calculate image contrast, the intensity of the noise, which is the sum of the background fluorescence and the excitation light noise, should be estimated. Background fluorescence noise can be estimated by replacing the fluorophore absorption and emission spectra with very broadband background fluorescence absorption and emission spectra, the relative efficiency of which can be determined empirically. The excitation light noise can be determined by replacing the fluorophore’s absorption and emission spectra with a wavelength-independent factor that represents the efficiency with which the excitation light remains in the optical path.²⁸

The bandwidth of the emission filter has a major impact on the brightness and contrast of the fluorescence image. Figure 5.12(b) shows the trade-off between brightness and contrast for different bandwidths of the emission filter.²⁸ Brightness increases with the increasing bandwidth of the emission filter, while contrast is higher when the bandwidth is narrower. The narrower the bandwidth of the emission filter, the more selectively it transmits the desired fluorescence signal and further reduces background noise, resulting in a higher contrast. But if the bandwidth of the emission filter is too narrow, the signal itself becomes too small for detection. Therefore, a balance must be found between the levels of signal and contrast for a given application.

Bandpass filters with a bandwidth of 20–40 nm are optimal for most fluorescence imaging. Filters with a bandwidth greater than 40 nm allow for the



(a)



(b)

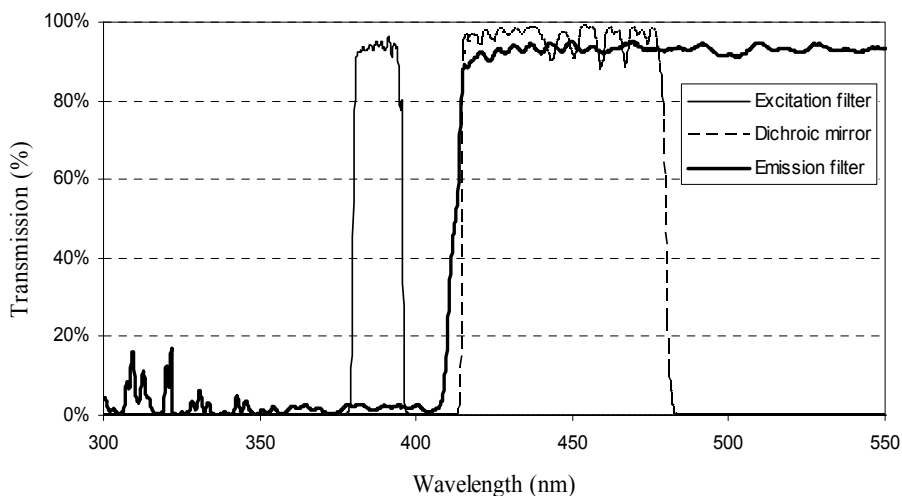
Figure 5.12 (a) Estimation of the fluorescence intensity and (b) the relative brightness and contrast of the fluorescence signal as a function of the bandwidth of the emission filter. (Courtesy of Semrock, Inc.)

collection of light at a wider spectral range and give a higher total signal; however, it is more difficult to discriminate between closely spaced, overlapping emission spectra. Filters with bandwidths narrower than 20 nm transmit fewer signals and are the most useful for fluorophores with very narrow emission spectra.

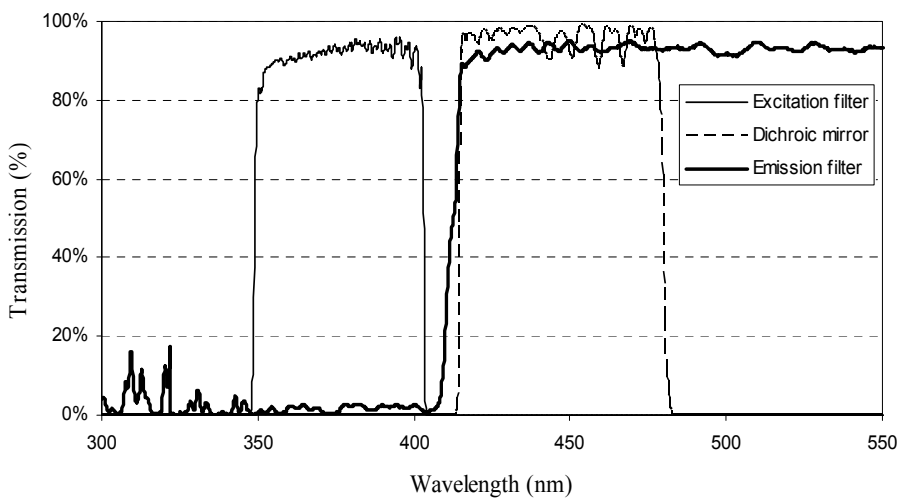
Figure 5.13 illustrates two sets of filters. One is optimized for a high-contrast image; the other is optimized for a high-brightness image. In Fig. 5.13(a), the bandwidth of the excitation filter is narrower to reduce noise and obtain a high-contrast fluorescence image. The bandwidth of the excitation filter in Fig. 5.13(b) is wider so that more fluorescence is generated, resulting in a fluorescence image

with higher brightness. This is a good choice when it is crucial to achieve the highest-brightness image for the shortest possible exposure time or for very low fluorophore concentrations. The image brightness can be further improved by using a bandpass emission filter with a wider passband or a longpass filter.

As a general rule, excitation filters are chosen to maximize blocking in the emission filter passband, while emission filters are chosen to maximize blocking in the corresponding excitation filter passband. The dichroic beamsplitter is is



(a)



(b)

Figure 5.13 (a) The filter set is optimized for achieving a high-contrast fluorescence image, and (b) the filter is set for a high-brightness fluorescence image. (Courtesy of Semrock, Inc.)

selected to maximize both the excitation and emission light intensities. Also, it is preferable to block out-of-band light with an excitation filter instead of an emission filter. Additional considerations when selecting fluorescence filters include the following: (1) fewer components inside the emission filter generally improve its optical imaging quality, and (2) the quality requirement on filters is higher in the imaging path than in the illumination path.

5.5 Fluorescence Imaging Systems

A number of fluorescence imaging systems have been developed to meet various needs. In this section, the configurations and designs of five selected fluorescence imaging systems, namely, lensless, microimaging, macroimaging, scan imaging, and fiber-optic imaging systems, will be discussed.

5.5.1 Lensless fluorescence imaging systems

In a lensless fluorescence imaging system, there is no lens between the sample and the detection element, which precludes any magnification of the image. The resolution of a lensless fluorescence imaging system is determined either by the size of the excitation spot in the scanning system or by the gap between the sample and the detector for an area imaging system. The gap includes an air gap or the thickness of the immersion media, the thickness of the substrate, and the emission filter in front of the detector.

There are three different configurations for lensless imaging. The first configuration is a scanning, lensless fluorescence imaging system, shown in Fig. 5.14, where a nonimaging detector, such as a PMT, is used. The emission filter prevents the excitation light from reaching the detector's photosensitive layer. One or more nonimaging, light-collecting components are necessary to capture and deliver the emission light to the detector because the detector cannot be too close to the object. The resolution of this imaging system is determined by the spot size of the excitation beam and two optical properties, namely, scattering and absorption, of the sample. This configuration has high light efficiency, usually more than 50%. Because of the PMT's inherent ability to provide gain and high light collection efficiency, it can be very sensitive and well suited to detecting low quantities of light. Another advantage is that the FOV is only limited by the scanning system or the moving stage. However, there are some disadvantages associated with this imaging method, including the low speed in obtaining a two-dimensional image and relatively low resolution and low contrast. The low contrast occurs because all of the light collected by the PMT is treated as light from the excitation point. The light detected by the PMT includes the fluorescence signal from the excitation spot and the unwanted fluorescence signals, such as the fluorescence from the light-collecting components or the fluorescence from locations other than the excitation spot, the unwanted fluorescence is excited by the excitation light reflected from the light-collecting components and the emission filter.

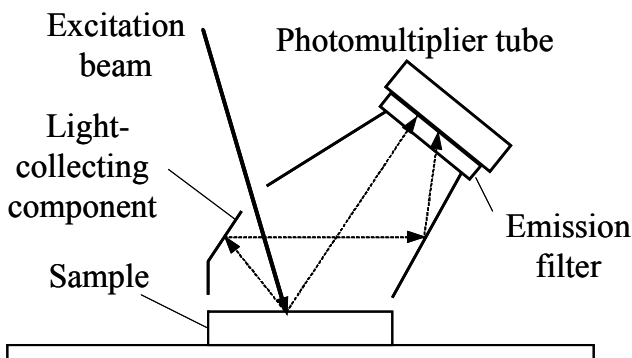


Figure 5.14 Scanning lensless fluorescence imaging system. The light-collecting component directs the fluorescent light to the detector.

The second configuration is a lensless transmission fluorescence imaging system, shown in Fig. 5.15. The sensor array is usually fabricated using standard CMOS technology. Given that the image sensor is a spatial sampling device, its spatial resolution is governed by the Nyquist sampling theorem and the distance between the sample and the photosensitive layer. The larger the gap is, the lower is the resolution of the imaging system. Therefore, the sample is placed as close to the sensor array as possible. The advantages of this imaging method include its compact size and its ability to quickly obtain a 2D image because the entire object can be illuminated and imaged at one time. Limitations of this method include that it can be used only in transmission mode, the resolution is relatively low, and the field size is limited by the sensor size. This configuration is particularly suitable for bioluminescence imaging because no emission filter is required.²⁹ Another application of this configuration is the scanning DNA microarray.³⁰

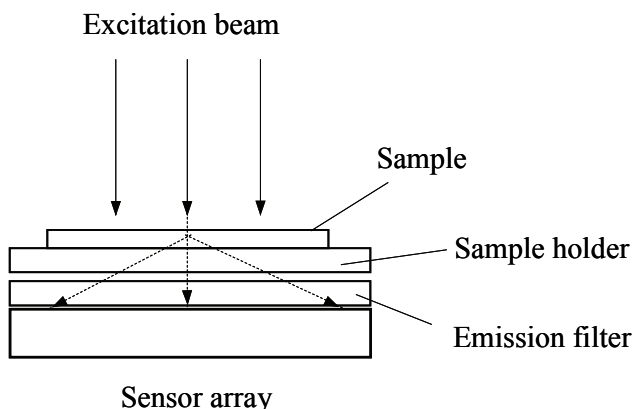


Figure 5.15 A lensless transmission fluorescence imaging system. The sample is illuminated with a broad beam, and the emission spectrum is detected with a sensor array through an emission filter.

The third configuration of a lensless fluorescence imaging system is based on a fiber bundle. The fiber bundle can be used to obtain fluorescence images from locations that are otherwise difficult to access. Usually a custom fiber bundle is required. For example, some of the fibers in the fiber bundle shown in Fig. 5.16(a) deliver excitation light, and the rest of the fibers capture the fluorescence signals and transfer them to the detector. The excitation and detection fibers are interlaced so that the tissue can be excited uniformly. The other end of the detection fibers is attached to the detector, with the emission filter placed between them to block the excitation light. The resolution of this configuration is limited by the distance between the fiber cores in the distal end and the gap between the proximal end of the fiber and the sensor's photosensitive layer.

Figure 5.16(b) shows a variation of the fiber-bundle lensless fluorescence imaging system. Each fiber in the fiber bundle is used for both excitation and detection. A dichroic beamsplitter directs excitation light into the fiber bundle and transmits the fluorescence signal to the detector. This is not a truly lensless system—a coupling lens is necessary to couple the light from the light source to the fiber bundle, and an imaging lens is also required to image the fiber surface to the detector. One advantage is that the individual fiber elements of the bundle are small enough to make high-resolution imaging possible.

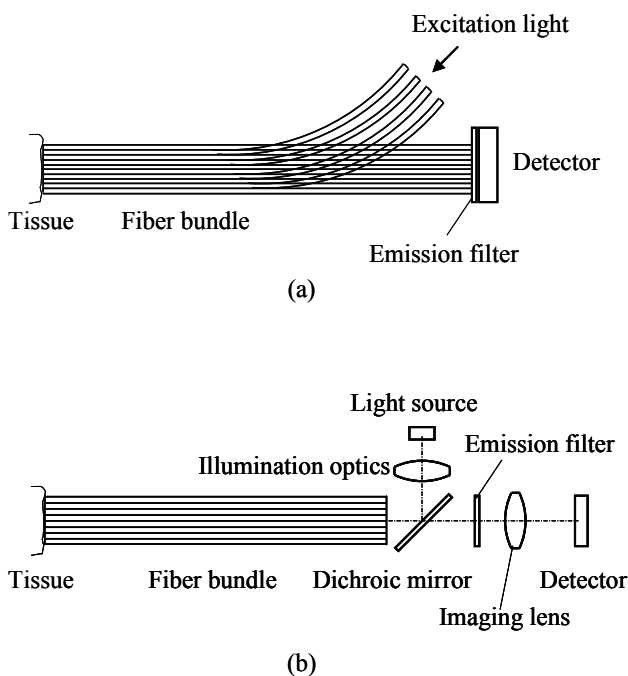


Figure 5.16 Fiber-bundle lensless fluorescence imaging system. (a) Some fibers deliver the excitation light, and the rest of the fibers capture the fluorescence signals and deliver them to the detector. (b) Each fiber in the fiber bundle is used for both excitation and detection.

5.5.2 Microimaging systems

A microimaging system refers to the imaging system that uses an imaging lens with a high NA and large magnification. It is generally used for imaging small features with high resolution. A fluorescence microscope is a typical example of a microimaging system. The fluorescence microscope has become an essential tool in biomedical imaging because it provides unique imaging capabilities that are not available in other optical microscopic techniques. With exogenous fluorophores, it is possible to identify cells and submicroscopic cellular components.

Figure 5.17 illustrates the principle of a fluorescence microscope. The excitation light is delivered to the sample by the illumination system through the filter cube and the objective lens. Inside the filter cube, the dichroic mirror is mounted at 45 deg with respect to the optical axis of the objective lens. The excitation filter is placed in front of the dichroic mirror in the illumination path, and the emission filter is placed after the dichroic mirror in the detection path. A portion of the light passes the excitation filter and is then reflected by the dichroic mirror such that it passes through the microscope's objective lens and illuminates the sample at the focal plane. When a laser or other monochromatic light source is used, the excitation filter is usually not required. When the excitation filter is necessary, it should be tilted slightly to prevent the reflected light from entering the laser cavity.

The sample under investigation absorbs some of the excitation then emits the fluorescent light with a longer wavelength. A portion of the fluorescent light is collected by the microscope objective lens and is transmitted through the dichroic mirror and the emission filter to the eyepiece or the sensor. Any excitation light leaking from the dichroic mirror will be further blocked by the emission filter. This configuration separates the excitation and emission spectra both physically and spectrally and is essential in obtaining a high-contrast fluorescence image.

In general, a bandpass filter is used as the excitation filter so that only the excitation wavelengths can reach the object to minimize background noise. For many applications, the optimal bandwidth of an excitation filter is approximately 40 nm, centered on the absorption maximum of the fluorophore. The dichroic mirror is mounted at a 45-deg angle to the optical axis of the microscope's objective lens. To collect as much fluorescence signal as possible, a shorter cut-off wavelength is usually chosen when the excitation and emission spectra are close to one another and the Stokes shift is small. Since the dichroic beamsplitter is in both the illumination and detection paths, the quality requirement for the beamsplitter should be precision grade. A dichroic beamsplitter is not necessary in some configurations of fluorescence microscopes, for example, in transmitted-illumination modes or fluorescence imaging systems with off-axis illumination.

The emission filter is required to block the excitation light. Both bandpass filters and longpass filters can be used to transmit the emission wavelength. A longpass filter enables the observer, through the eyepiece, to see all of the visible

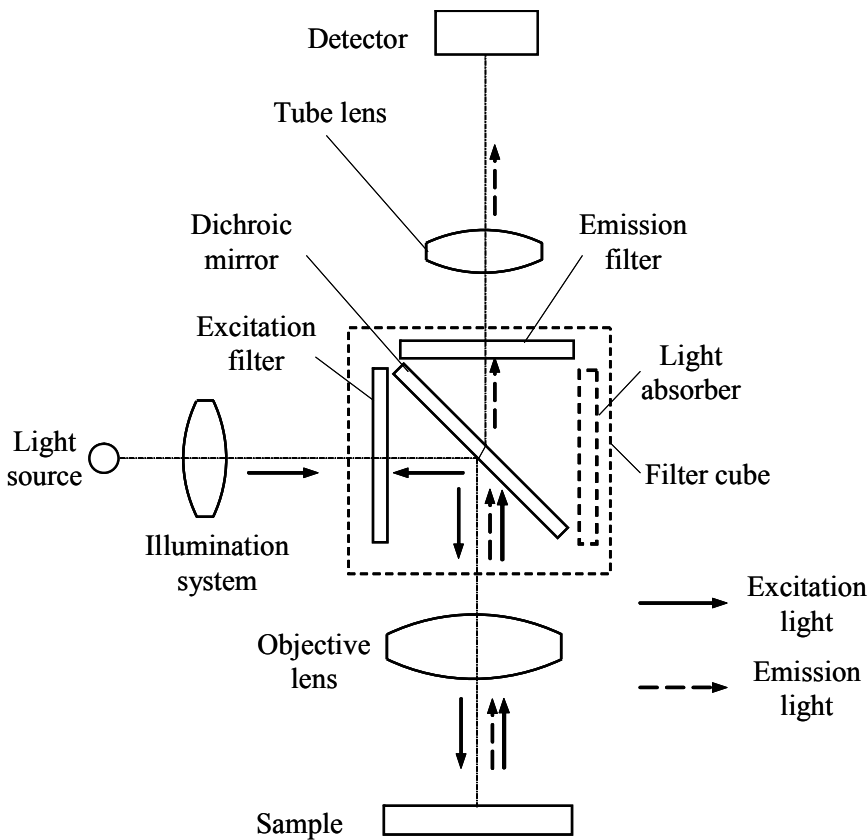


Figure 5.17 Diagram of a fluorescence microscope.

fluorescence emitted from the other fluorophores. However, when a sensor that is sensitive to a longer wavelength is used, or when the object is a multiply stained sample, a bandpass filter is preferred for better image contrast. For a high-resolution fluorescence imaging system, a bandpass filter that transmits a band at or near the emission peak of the fluorophore generally minimizes the background noise and improves the sensitivity and linear range of the measurement. The quality requirement of the bandpass filter is similar to that of the dichroic beamsplitter (precision grade) because it is in the fluorescence detection path.

The fluorescent light radiates from a fluorophore in all directions; it is typically collected from only a relatively small cone angle from one side of the sample. The relatively low light level in fluorescence microscopy makes the choice of filters critical for achieving adequate SNR levels. For the applications where signal bleed-through, caused by overlapping fluorophore emission profiles, is of concern, bandpass excitation and emission filters are usually selected but often only to minimize crosstalk as much as possible.

A light absorber is sometimes used at the opposite side of the excitation filter, as shown in Fig. 5.17. It can be a simple element made from an absorbing material, beam dumper, or other component. The light absorber prevents the light

transmitted through the dichroic mirror from scattering within the filter block and leaking into the detection path. It helps to eliminate the background noise, therefore producing a fluorescence image with higher contrast. The light absorber is particularly useful when the fluorescence signal is very weak.

The objective lens is one of the key components in a fluorescence microscope. Generally, it is defined by the NA, magnification, and the degree of aberration correction. In a reflectance fluorescence microscope, the objective lens serves as both an image-forming lens and a condenser lens. It delivers a strong excitation light onto the object surface, collects weak fluorescent light from the object, and forms an image on the image plane. Generally, the basic requirements for the objective lens used in a fluorescence microscope include:

- A high NA to collect fluorescent light. The image brightness is proportional to the fourth power of the NA of the objective lens. For some applications with weak fluorescence signals, the NA is increased for collecting more fluorescent light instead of increasing the resolution.
- Sufficient transmittance for near-UV light. For some fluorophores, near-UV light is used to excite fluorescence. Therefore, the objective lens should have a high transmission in the near UV so that the light from the light source can be effectively delivered to the object. Also, the transmission in the near UV should not change over time.
- Optical materials with low autofluorescence. Since the objective lens also works as a condenser, the strong excitation light passes all of the optical elements inside the objective lens. Therefore, the materials used in the objective lens should have low autofluorescence, especially the materials used in the front lenses close to the object, where the excitation light converges at the front lenses and the irradiance becomes much higher.

While the brightness of the image is proportional to the fourth power of the NA, it is inversely proportional to the second power of the magnification. Therefore, low-magnification objective lenses are often used when imaging dim samples. When the SNR is more critical than the resolution, a high-NA objective lens with lower aberration correction is sometimes used to gain a higher light throughput.

The aberration correction of objective lenses depends on the application. When the imaging system is required to have perfect alignment between the excitation and fluorescent light in the visible range, an apochromatic lens is necessary. However, the apochromatic lens may not be suitable for fluorescence imaging with UV excitation because it requires additional lens elements to achieve a high degree of aberration correction, resulting in a low UV transmission.

The fluorescence microscope discussed above uses epi-illumination. Figure 5.18 shows two configurations of fluorescence microscopes that use transillumination. Excitation and emission filters are required to obtain a high-

contrast fluorescence image, while a dichroic beamsplitter is not necessary in a transmission fluorescence microscope.

In bright-field transmission fluorescence imaging, as shown in Fig. 5.18(a), both excitation and emission light directly enter the objective lens. Because significant amounts of the excitation light enter the objective lens, the optical materials used in an objective lens should have minimal autofluorescence.

In a dark-field transmission fluorescence imaging system, shown in Fig. 5.18(b), a dark-field condenser with a high NA focuses the illumination light onto the sample. Most of the excitation light never enters the objective lens because the angle of the illumination light is larger than the acceptance angle of the objective lens. The NA of the objective lens must be reduced to prevent the excitation light from directly entering the objective lens. Given that most of the excitation light does not enter the objective lens, the requirement for the objective lens is similar to that of a conventional microscope objective lens. Excitation light that is diffusely scattered from the sample cannot be blocked from entering the detection optics, but its intensity is low enough that it does not excite significant emission in the detection path. Any excitation light entering the objective lens is blocked by the emission filter.

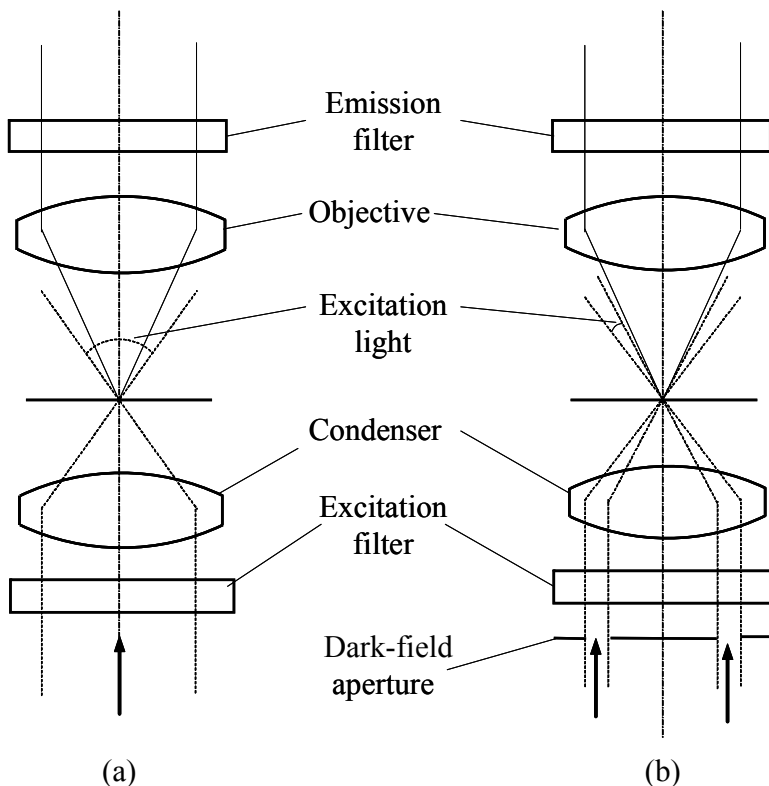


Figure 5.18 Fluorescence microscopes utilizing transmitted light illumination: (a) bright field and (b) dark field.

5.5.3 Macrofluorescence imaging systems

Macrofluorescence imaging systems, also called wide-field imaging systems, are usually used when the imaging field is larger than the detector size, for example, in small-animal imaging systems and image-guide surgery.^{31–33} Macroimaging systems are also commonly chosen when a single detector is used to detect light from regions widely spread out in space, for example, a microarray. The macroimaging system is a simple approach to imaging the fluorescence emitted from a large field. It is relatively fast and highly stable; it also has a low probability of false signals.

Figure 5.19(a) is a diagram of a fluorescence imaging system using off-axis illumination. The illumination is provided by a broadband arc lamp or LEDs. Given that the FOV is large compared to a fluorescence microscope, a powerful light source is needed to excite the entire field. The illumination spectrum can be selected through color filters, thin-film filters, tunable filters, or other mechanisms, as are used in systems based on a spatial light modulator. The excitation filter can be placed in various locations in the illumination path; however, the preferred method places the filter as the last element in the illumination path so that the autofluorescence generated from the components in the illumination system can be blocked.

Efficient illumination optics is necessary to distribute the excitation illumination uniformly across an entire field. Depending on the requirement for uniformity, a number of illumination systems can be used. The simplest

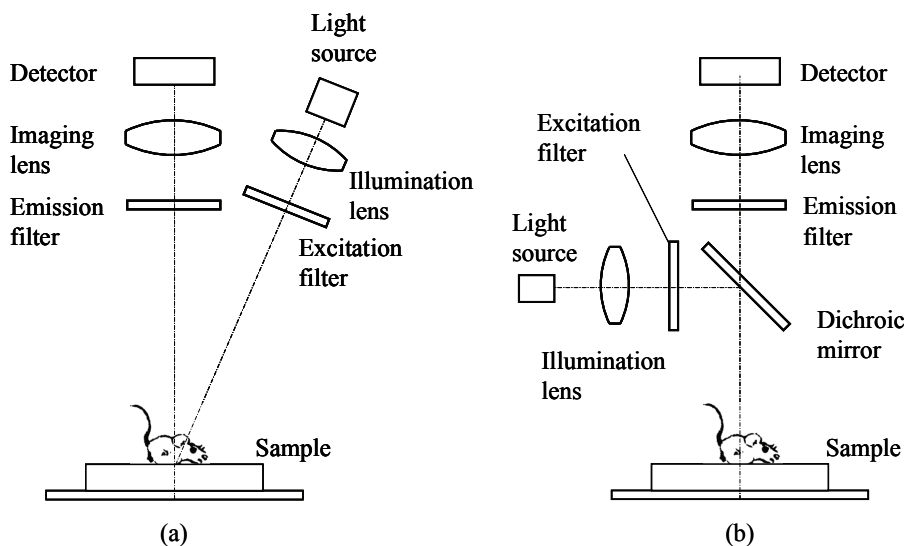


Figure 5.19 Macrofluorescence imaging system: (a) off-axis illumination and (b) on-axis illumination.

illumination system is direct illumination from a light source. This method does not require any additional optical components; the illumination uniformity is determined by the angular distribution of the light source. One way to improve the illumination uniformity is to use a group of light sources, such as LEDs, and arrange them carefully to obtain uniform illumination. The second illumination configuration is to use a lenslet array to subsample the light from the light source then use a field lens to overlap the sampled beams onto the sample. This method has better uniformity and light efficiency. Other illumination methods include critical illumination, Köhler illumination, and a light-pipe illumination system. The illumination configuration in Fig. 5.19(a) is an asymmetric, off-axis reflective illumination. To obtain symmetric illumination, another illumination arm can be added symmetrically.

Figure 5.19(b) shows an on-axis epi-illumination system with a dichroic mirror separating the excitation and emission beams. The dichroic mirror can be placed between the imaging lens and the detector for high-NA and large-FOV systems. The illumination and imaging systems share a common optical axis; therefore, it is relatively easier to control the illumination profile. The light from the object is collected by an imaging lens and focused onto an area detector, such as a CCD or CMOS device. In order to collect as much emission light as possible and ensure enough image resolution, the imaging lens should have a large aperture and adequate aberration correction. Some falloff in light intensity at the edges of the field can be expected when the FOV is large because of the cosine-fourth falloff and vignetting.

The emission filter is preferred as the first element in the fluorescence detection path to minimize the autofluorescence from the lens elements in the imaging path by blocking the excitation light. However, this approach may not be possible when the imaging lens has a large FOV and high NA because the front aperture of the imaging lens is fairly large and an emission filter with a large aperture is expensive. Another issue is that the performance of the emission filter will be degraded because the incident angle of the off-axis ray is large. As an alternative, the emission filter is usually placed between the lens elements where the ray bundle is limited and the ray angle is small, or it is placed in front of the sensor. When designing such an imaging lens, materials with very low autofluorescence should be selected for the elements in front of the emission filter.

One application of macrofluorescence imaging system is a microarray reader. The illumination system uniformly illuminates the microarray, and a CCD-based imaging system is used for high-throughput parallel detection and quantitative analysis of the fluorescence signal. With uniform illumination of the sample, a significant portion of the light falls onto the area on the chip where there is no specimen. That light may generate undesirable background fluorescence and hence may degrade image contrast. To increase the light efficiency and reduce the background fluorescence, a special illumination system that illuminates only the microarray is desired. Beam-shaping optics, such as a diffractive optical element (DOE), a lenslet array, or a special fiber bundle, can be designed and

fabricated to transform excitation light into a dot matrix that matches the relative position of the microarray. Each focused spot excites the fluorescence signal from each dye dot. If LEDs are used as the excitation light source, the LEDs can be packed directly as a matrix, and can then be imaged onto the sample to excite the dye dots.

The performance of the macroimaging system is specified by resolution, sensitivity, linearity, and dynamic range. The resolution of a macrofluorescence imaging system is determined either by the imaging lens or by the pixel size of the sensor. When the RMS diameter of the focused spot is smaller than the pixel size of the sensor, the resolution is limited by the sensor. On the other hand, when the pixel size is smaller than the RMS spot diameter, the system resolution is limited by the imaging lens. The image resolution is reduced when charges from adjacent pixels are “binned” during image acquisition. The resolution can be increased by stitching multiple images that are captured by moving the lens and sensor relative to the sample.

System sensitivity, linearity, and dynamic range are often detector related. Area sensors, CCDs, or CMOS devices are sensitive to light and temperature. The sensitivity and linearity of the system can be significantly improved by cooling the sensor. Dynamic range is the signal range over which the system has a linear response to the fluorophore concentration. The dynamic ranges of macroimaging systems are often limited by electronic background noise.

5.5.4 Scanning fluorescence imaging systems

To improve the dynamic range of a fluorescence imaging system, a more-sensitive detector, such as a PMT, can be used. The samples are analyzed by scanning the excitation beam, and the digitized images are generated by sequentially reading the detector output. The sample can be scanned either by deflecting the excitation beam (off-axis scan or beam scan) through the flat-field scanning lens or by laterally shifting the sample (on-axis scan or stage scan).

A scanning system usually comprises one or more light sources with variable beam-shaping optics to allow for easy changes in resolution and throughput, a high-NA scanning lens to capture as much fluorescent light as possible and thereby increase system sensitivity, and appropriate fluorescence filters to separate the fluorescent light from the excitation illumination.

Figure 5.20 shows diagrams of off-axis and on-axis scanning configurations. In the off-axis scanning system as shown in Fig. 5.20(a), the scanner with either one axis or two axes scans the excitation beam across the sample through the scanning lens. The emission light is collected by the scanning lens and recorded by the detector. The excitation light captured by the scanning lens passes through the dichroic mirror and returns to the illumination system. The excitation light leaking from the dichroic mirror will be further blocked by the emission filter. Different configurations of scanners and scanning lenses will be discussed in Chapter 7.

Figure 5.20(b) is a schematic diagram of an on-axis scanning system. On-axis scanning means that the illumination beam is always on axis. Either the

objective lens or the stage moves relative to the other. The angle and path length of the excitation beam are identical at any point on the object. On-axis scanning eliminates variation in power density, fluorescence collection efficiency, and image distortion of the scanning lens. Also, because the scanning lens only works

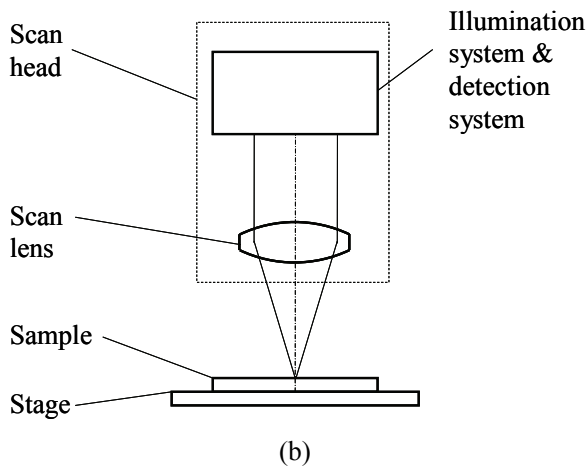
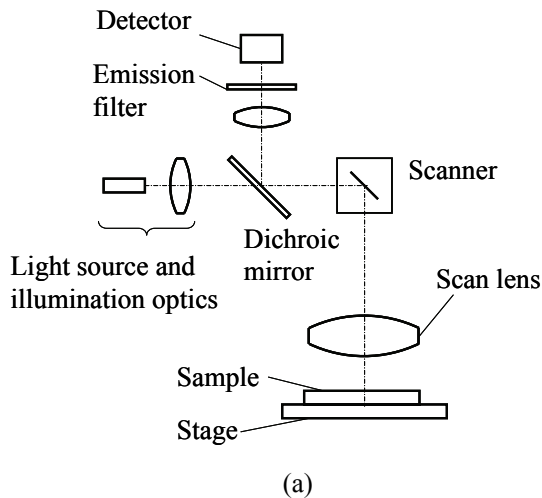


Figure 5.20 (a) Off-axis scanning system. The excitation light is scanned across the sample by the scanner. The fluorescence signal is collected by the scanning lens and directed to the detector by a dichroic mirror. (b) On-axis scanning system. The light from the light source is focused onto the sample. The entire sample is imaged by moving either the sample or the scanning lens relative to the other.

on axis, the lens can be much simpler than the lens used for off-axis scanning. One additional advantage is that the dynamic autofocus can be integrated into the scan head to accommodate for the subtle variations in glass-slide curvature, thickness, or other aberrations that can degrade image quality. The disadvantage of an on-axis scanning system is its low speed in acquiring 2D images.

For an off-axis scanning system, the ideal scanning lens is a telecentric F - θ lens with a flat field. If the scanning lens is not telecentric, the collection efficiency of the fluorescent light will not be uniform across the entire FOV. Also, the excitation beam does not illuminate the sample at the same angle in every location, as shown in Fig. 5.21(a); therefore, the spot size is not the same across the entire field. The best-focused plane formed by a scanning lens without corrected field curvature is a circular arc, producing some variations in the effective excitation energy reaching the object at a different point. The above issues can be minimized with a telecentric F - θ lens with a corrected field curvature, as shown in Fig. 5.21(b). The chief ray is parallel to the optical axis of the scanning lens.

The performance of a scanning system is assessed in terms of resolution, dynamic range, uniformity, and sensitivity. Resolution is the scanning system's ability to distinguish between two close objects and is a function of the NA of the scanning lens for a diffraction-limited system. A wide dynamic range enables the scanning system to detect and quantify signals from very low and very high intensity targets in the same scan. Uniformity is the measurement consistency of a given fluorescence signal at any position within the imaging field. Uniformity is critical for reliable quantitative measurement. Sensitivity is the minimum amount of the fluorescence signal that can be detected by the scanning system. A system with good sensitivity is more cost effective than less-sensitive systems because a smaller fluorescent sample is needed for analysis and the critical time spent on a sample is reduced, limiting photobleaching and phototoxicity.

5.5.5 Fiber-optic fluorescence imaging systems

Because of their ability to guide light between separate locations, their flexibility, and their small size, fibers have found more and more applications in fluorescence imaging. Fiber-optic fluorescence imaging can be used in situations where a conventional fluorescence microscope cannot be used. Optical fibers might be implanted within live subjects for long-term imaging studies; they can also be used in minimally invasive clinical diagnostics and surgical procedures.

The primary function of optical fibers in fiber-optic fluorescence imaging is the remote delivery of excitation and emission light. The excitation light source, detector, and other related optical components, such as the excitation and emission filters, reside apart from the observation location. One or more lenses at the distal end of the fiber focus the light onto the tissue and then collect light from the tissue and couple it back to the fiber. Other components at the distal end may include a focusing mechanism and a scanning mechanism. The fluorescence

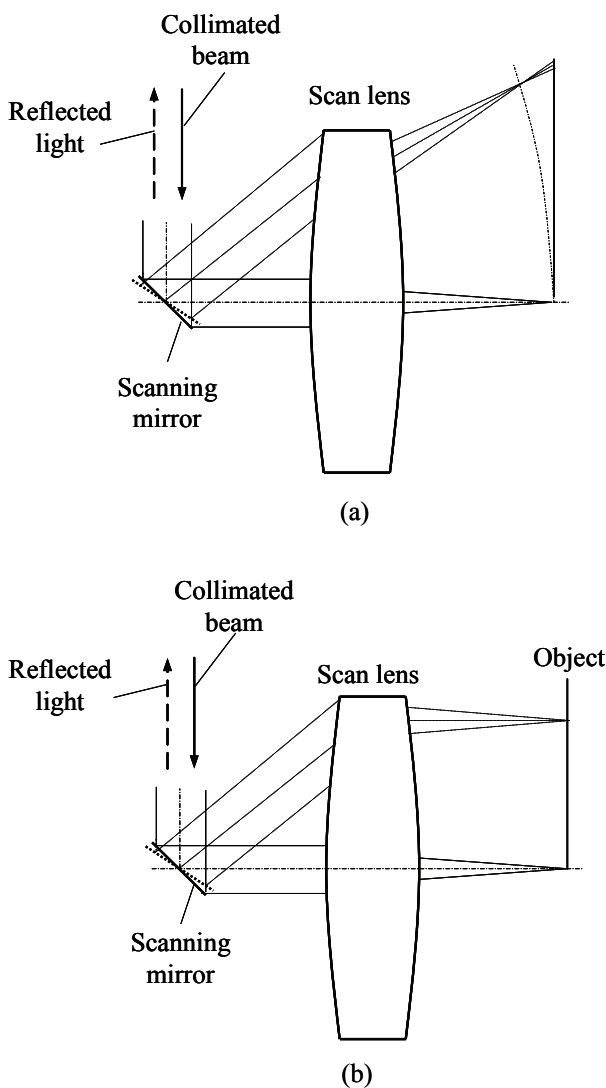


Figure 5.21 Off-axis scanning system. (a) A nontelecentric scanning lens without corrected field curvature focuses the excitation light to a circular arc. (b) A telecentric F - θ lens with corrected field curvature focuses the excitation light onto a flat surface.

filters, namely, a dichroic beamsplitter and excitation and emission filters, are placed in the proximal end of the fiber to separate the excitation and emission lights.

Both single-mode and multimode fibers have been used in fluorescence imaging. Single-mode fibers are well suited for illumination light delivery in laser scanning systems because the single-mode light can be focused to a near-diffraction-limited spot in the object plane for high-resolution imaging. They also

act as a pinhole to reject out-of-focus fluorescence emissions for confocal imaging. The limitation of a single-mode fiber is its collection efficiency of fluorescent light, which is low because of its low NA and small core diameter.

Figure 5.22 shows a typical configuration of fluorescence filters and other components in the proximal end of single-fiber fluorescence imaging systems. The focusing lens focuses the excitation light onto the input surface of the fiber. The emission light transmitted through the fiber is collected by the same focusing lens and is directed to the detection channel through a dichroic mirror, and a highly sensitive photodiode or a PMT is generally used for signal detection. When the light source is monochromatic, the excitation filter is required.

A fluorescence imaging system using single-mode fibers usually employs one of the following distal scanning mechanisms: an electromagnetic actuator, piezoelectric driven actuator, or microfabricated micro-electromechanical system (MEMS) mirror. Two typical examples of distal scanning methods are shown in Fig. 5.23. In Fig. 5.23(a), a piezoelectrically or electromagnetically excited actuator is used to drive mechanical resonance vibrations to the fiber tip. The chief ray of the scanning lens should be along the optical axis of the fiber. The disadvantage of this scanning method is that it is usually not possible to offset the center of the scanning field from the optical axis of the scanning lens. Another limitation is that the sampling may not be uniform. In Fig. 5.23(b), a MEMS scan mirror scans the excitation beam across the sample. A MEMS scan mirror has the advantages of being compact and having low power consumption. Miniaturized scanning mechanisms and imaging lenses will be discussed in detail in Chapter 7.

Multimode fibers can transmit more than one mode of light. The number of transmitted modes increases approximately quadratically as a function of the NA, as indicated in Eq. (3.5). Multimode fibers have a core diameter of 50 μm or more, and they are more suitable for fluorescence collection than single-mode fibers because multimode fibers have larger core diameters and higher NAs, but they are less flexible than single-mode fibers.

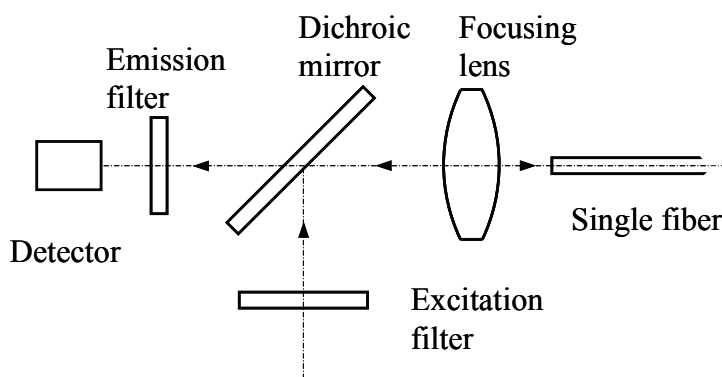
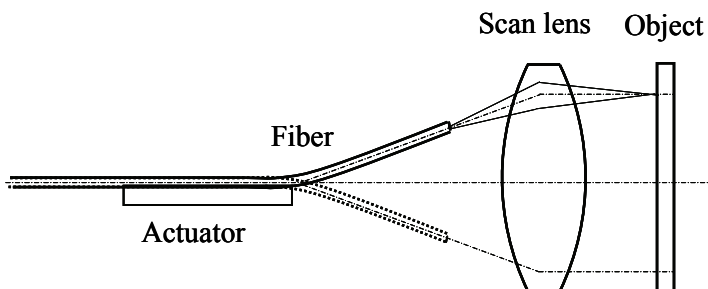
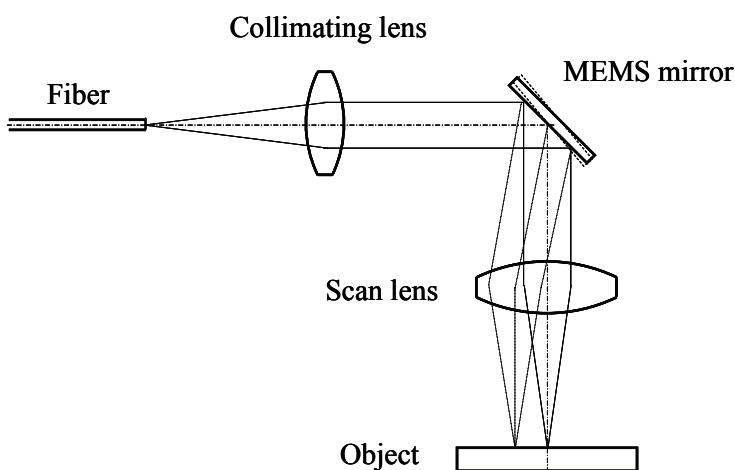


Figure 5.22 Configuration of fluorescence filters and other components in the proximal end of a single-fiber fluorescence imaging system.



(a)



(b)

Figure 5.23 (a) Distal fiber scanning with a magnetically or piezoelectrically driven actuator and (b) distal scanning with a MEMS mirror.

A fiber bundle typically contains thousands of high-index cores in a common cladding with a low refractive index. The relative positions of the fibers at the input and output ends are spatially correlated such that an image transmitted through the fiber bundle can be reconstructed from either end of the fiber in an identical manner. The resolution of the fiber bundle is limited by the core-to-core distance between the fibers and is further degraded as a result of the pixilation nature and crosstalk between individual fibers.^{34,35} In confocal fluorescence imaging systems, the crosstalk issue can be reduced by using noncoherent fiber bundles, where the relative fiber positions are randomized and not maintained throughout the fiber bundle, but the position of each fiber is calibrated.

Figure 5.24 illustrates a typical configuration for a fluorescence imaging system using a fiber bundle. The excitation light is coupled to the fiber bundle by a coupling lens and then is focused onto the tissue surface by an imaging lens. The emission light from the tissue surface is coupled into the fiber bundle by the imaging lens, then the transmitted emission fluorescent light is imaged onto the detector by the coupling lens and the focusing lens. In confocal fluorescence imaging mode, a scanning device is necessary to couple the light into each fiber sequentially and thus scan the focused excitation light across the tissue. The emitted fluorescence transmitted through the fiber bundle is descanned and focused onto the detector, which is usually a PMT or a photodiode.

Proximal scanning is usually used in fiber-bundle confocal fluorescence imaging systems. Galvanometer scan mirrors are typically used in proximal scanning configurations. Figure 5.25 shows two proximal scanning configurations. In Fig. 5.25(a), two scanning mirrors, rotating in orthogonal directions, scan the excitation beam across the proximal end of the fiber bundle and couple the illumination light into the fibers. The configuration in figure 5.25(b) uses a spatial light modulator, such as a digital micromirror device (DMD), to couple the excitation light to the fiber randomly or sequentially. The spatial light modulator is uniformly illuminated. The scanning is performed by turning the pixels on and off in the spatial light modulator. The relay lens images pixels onto the fiber surface and couples the light into fibers. The first scanning method has better light efficiency, while the second method has the potential for multiple-point scanning.

Proximal scanning offers the benefit of separating a bulky scanner from a miniaturized imaging head and provides high image acquisition rates. Video-rate imaging is possible with a fast line-scanning approach. One issue with proximal scanning is that while the light is coupled into the fiber core, it enters the cladding as well. The light in the cladding reduces the light efficiency and degrades image contrast. Optical systems such as the scanning lens in the

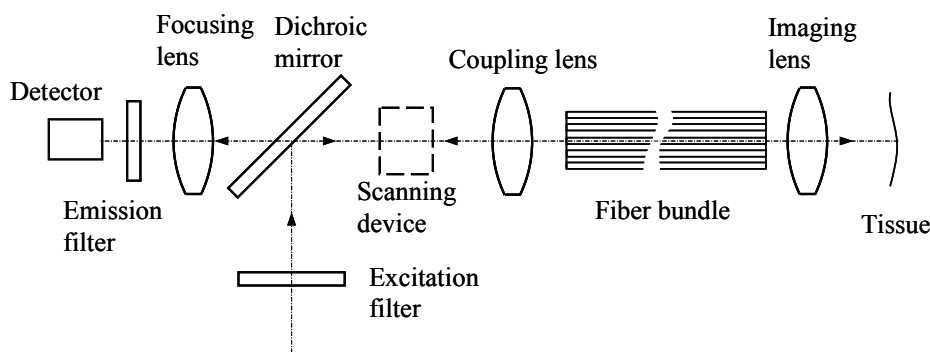


Figure 5.24 Configuration of a fluorescence imaging system using a fiber bundle. The scanning device in this configuration is necessary only in the scanning mode.

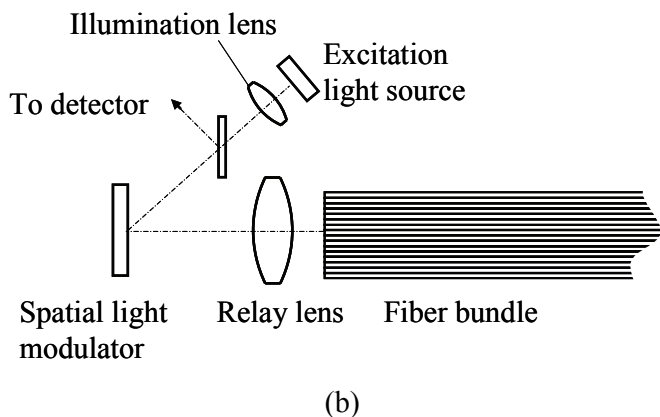
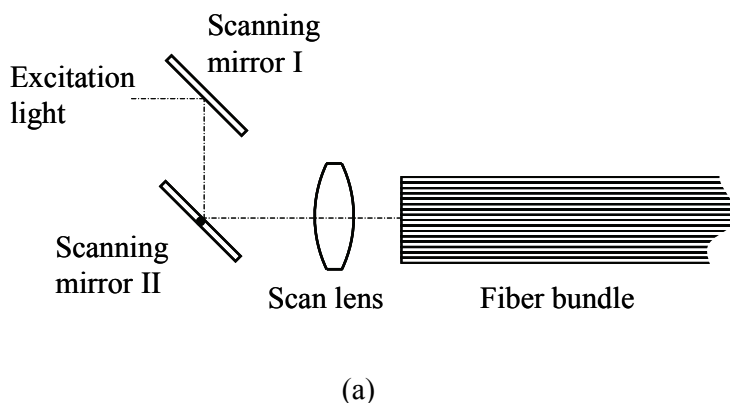


Figure 5.25 (a) Two close-coupled scanning mirrors scan the excitation beam across the proximal end of a fiber bundle. (b) The spatial light modulator is uniformly illuminated, and the relay lens images the randomly or sequentially active pixels onto the proximal end of the fiber bundle without sweeping the beam.

proximal end and the imaging lens in the distal end will be discussed in Chapter 7.

Like other imaging systems using fiber bundles, the performance of the fiber bundle in fluorescence imaging depends on the NA, core-cladding fill factor, reflection at the air/fiber interfaces, and fiber absorption. In fluorescence imaging systems, the autofluorescence from the components other than the sample is background noise, and it reduces image contrast. [The excitation-emission matrices (EEM) of three different fiber bundles from three different manufacturers were illustrated in Fig. 3.11.³⁶] Therefore, it is necessary to measure the autofluorescence from the fiber bundle at the working wavelength and evaluate its impact on the fluorescence image before the fiber bundle is

integrated into the system. The EEM measurements, as well as the background autofluorescence images of the fiber bundle at a working spectrum, provide useful information for optimizing the excitation/emission wavelengths, fluorophores, and configuration of fluorescence filters for specific imaging applications. Filters and fluorophores can be chosen to minimize the impact of the background fluorescence from a fiber bundle. By measuring the background fiber bundle autofluorescence, the effect of the autofluorescence of the fiber bundle can be reduced by subtracting the background autofluorescence from the fluorescence images.

5.5.6 Multicolor fluorescence imaging systems

As discussed in Sec. 5.2.2, if two or more independent fluorescence images excited by different wavelength bands are obtained, it is possible to evaluate the fluorescence as a dimensionless quantity by forming a ratio of intensities. As a result, the measurement will be less dependent on the measurement geometry, the illumination uniformity, and the surface topography.

The fluorescence images excited by different wavelength bands can be measured either sequentially or in parallel. In sequential imaging, multicolor fluorescence images of the same tissue are recorded one by one by shifting emission (and excitation, depending on the configurations) filters between measurements. The advantage of sequential illumination is that there is no limit to the number of fluorescence bands. One limitation with sequential imaging is possible tissue movement, producing increased difficulty in correlating images. Parallel imaging is sometimes preferred to avoid movement during the measurement. The limitation of parallel imaging is that the number of fluorescence images that can be captured simultaneously is restricted.

Figure 5.26 shows two configurations of excitation paths. Figure 5.26(a) shows an off-axis illumination system without a dichroic mirror. Several excitation light sources or one light source with a set of excitation filters are used to illuminate the sample sequentially or simultaneously. The filter wheel selects the excitation spectrum for sequential imaging, and the multiband excitation filter is used for parallel imaging.

Figure 5.26(b) shows an on-axis illumination system, where a multiband dichroic mirror is required to deliver the excitation light to the sample. Similar to the off-axis configuration, several excitation light sources or one light source with a set of excitation filters illuminate the sample sequentially or simultaneously.

Figure 5.27 shows the configurations of the detection systems for multicolor fluorescence imaging systems. Figure 5.27(a) shows a sequential detection system. The filter wheel in the detection path rotates according to the rotation of the filter wheel in the illumination path. Figure 5.27(b) shows a parallel imaging detection configuration. A multiband emission filter is necessary to block the multiband excitation light and pass the multicolor fluorescent light. A color

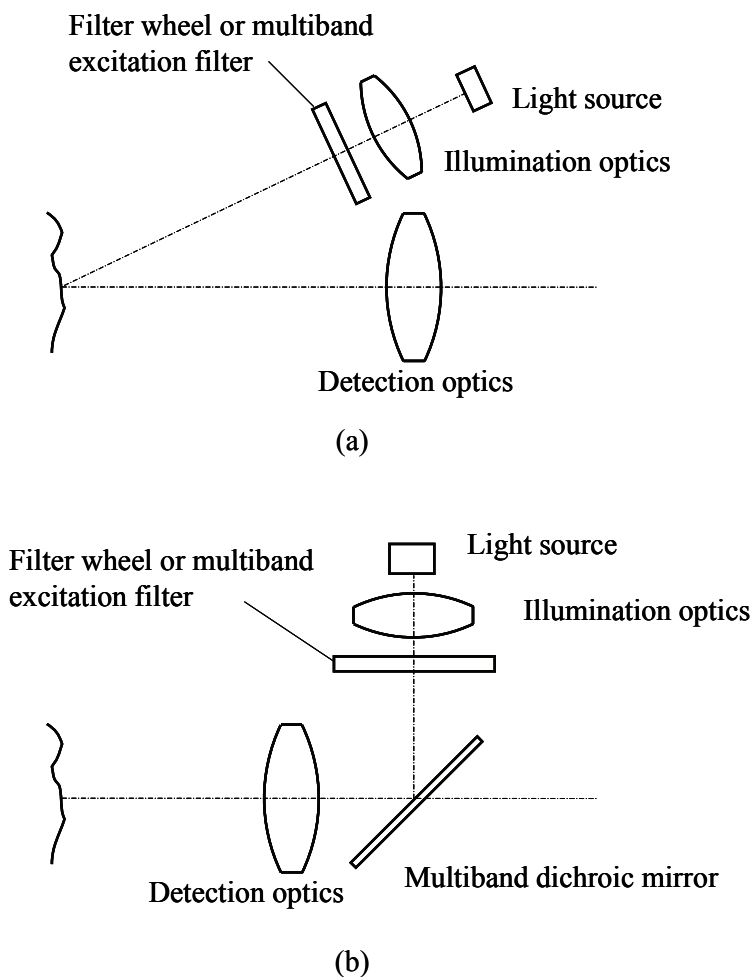


Figure 5.26 Illumination configurations for multicolor fluorescence imaging systems. (a) Off-axis illumination and (b) on-axis illumination. The filter wheel and multiband excitation filter are not necessary for monochromatic light sources.

detector is required in this configuration to capture the fluorescence light at different wavelengths. Parallel imaging with one color detector requires that the transmission bands of the color filter array inside the color detector not overlap. Figure 5.27(c) illustrates a configuration with three detectors to capture the fluorescence images at three spectral bands simultaneously. The dichroic mirrors direct the fluorescence signals to different detectors. This method efficiently captures fluorescence images in three spectral bands. However, it is not practical to add more than three detectors to capture fluorescence images because of the difficulties in the optical systems and package.

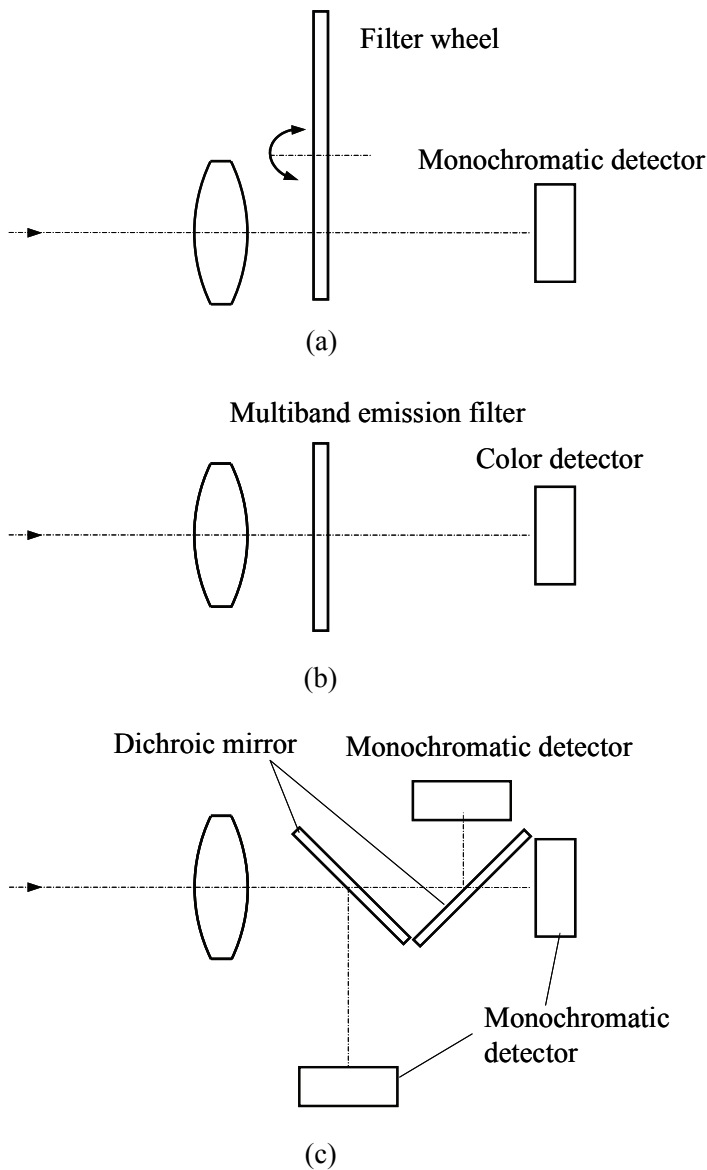


Figure 5.27 Configurations of detection systems used for multicolor fluorescence imaging systems: (a) filter wheel and monochromatic detector, (b) multiband emission filter and color detector, and (c) dichroic mirrors and monochromatic detectors.

5.5.7 Hybrid imaging systems

While the fluorescence technique offers good sensitivity in detecting tissue changes such as early lesions and tooth decay, it can suffer from low specificity, which can be attributed to the presence of blood, food, bacteria, etc. In addition,

ambient light also degrades the quality of the fluorescence image. Because of tissue scattering, a fluorescence image does not provide high-resolution information of the tissue surface, nor does it provide depth information.

Fluorescence images can be combined advantageously with polarized or nonpolarized reflectance images to improve specificity. They can also be combined with the optical coherence tomography (OCT) imaging technique to gain depth information.

Figure 5.28(a) is a diagram of a hybrid tissue imaging system combining fluorescence and polarized reflectance imaging.³⁷ A polarizer is placed in front of the light source, either broadband or monochromatic, to provide polarized light illumination. An analyzer, whose transmission axis is orthogonal to that of the polarizer, is placed in the imaging path to block the reflected or scattered light with the same polarization as the illumination beam. Therefore, all of the reflected light from the specular reflection is filtered out and will not reach the detector. Another light source, usually with a blue or UV wavelength, is used to excite fluorescence from the sample. In the imaging path, a longpass emission filter, instead of a bandpass filter, is chosen to block the excitation light and pass the emission light as well as the polarized reflectance light. By switching the light sources on and off for reflectance imaging and fluorescence imaging alternatively, both polarized reflectance and fluorescence images are available for further analysis. When a white-light source, such as a white LED, is used, the reflectance image provides a clinical view of the tissue surface. Another advantage is that there is no “hot” spot from the specular reflection in the reflectance image. The disadvantage of this imaging system is that the fluorescence signal is cut to half by the analyzer. This drawback can be remedied by moving the analyzer out of the image path when capturing fluorescence images.

Figure 5.28(b) is a diagram of a hybrid imaging system that combines fluorescence and OCT imaging.³⁸ A dichroic mirror is used to combine the light from the OCT and fluorescence imaging paths. The OCT scanning system is usually designed to be telecentric on the tissue side so that the magnifications are the same for points with the same lateral position but at different depths. This imaging system combines the merits of two imaging modalities for better sensitivity and specificity: fluorescence imaging is sensitive to the early stage of tissue changes, and OCT imaging provides depth information. Since OCT imaging is relatively slow, the fluorescence image can be used to guide the OCT imaging so that the OCT image is only obtained from the suspicious region that is identified from the fluorescence image.

The imaging modalities in Figs. 5.28(a) and 5.28(b), namely, polarized reflectance imaging, fluorescence imaging, and OCT imaging, can also be integrated to further improve sensitivity and specificity.^{37,39} With wide-field-of-view visible reflectance and fluorescence imaging, the suspicious regions of the tissue can be easily identified, and detailed information of those regions can be further obtained by OCT scanning.

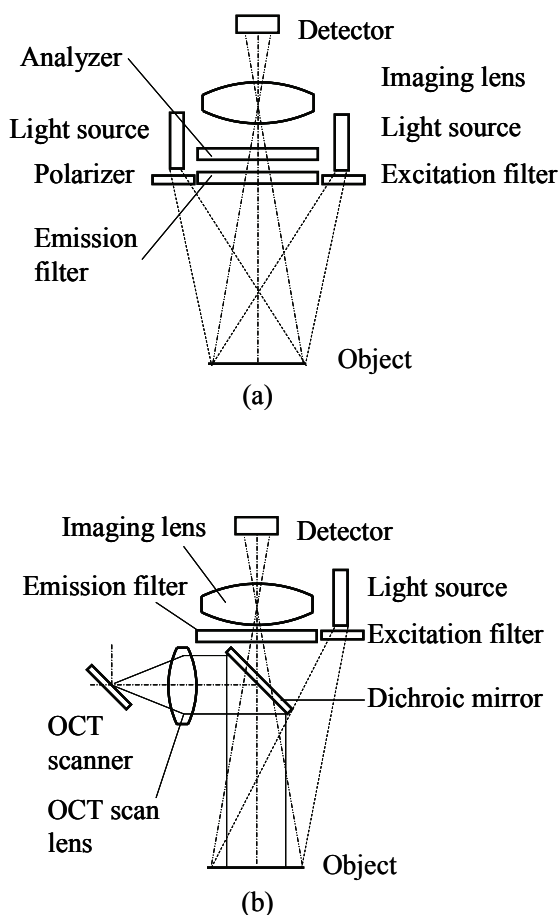


Figure 5.28 Hybrid imaging systems: (a) polarized reflectance imaging and fluorescence imaging and (b) fluorescence imaging and OCT imaging.

5.6 Optical Designs of Fluorescence Imaging Systems

5.6.1 System consideration

The design parameters for a fluorescence imaging system include the power and homogeneity of the excitation light, the FOV and NA of the detection system that collects the emission light, the sensitivity and noise characteristics of the detector, and the transmission and blocking capabilities of the fluorescence filters.

Fluorescence signals are usually very weak and require long exposure times because of the low concentration of fluorophores. Therefore, the design of an optical system for fluorescence analysis must consider the entire optical path. Low fluorescence signal requires a highly efficient optical system to improve light-capturing abilities for a higher throughput, to provide a higher dynamic

range that accommodates the vast differences in fluorophore concentrations across the sample, and to reduce crosstalk between sample spots through improved optical resolution.

To optimize the detection of a fluorescence signal, excitation filters are selected to maximize blocking in the transmission passband of the emission filter in the illumination path and to maximize blocking in the corresponding transmission passband of the excitation filter in the detection path. In general, it is preferable to block out-of-band light with an excitation filter instead of an emission filter so that the sample will be exposed to less radiation. In addition, fewer components and less-complicated optical systems are required in the detection path.

The optimal position of the excitation filter is where the range of the ray angle is small and away from the light source to reduce the angular effect of the optical coatings and autofluorescence of the components in the illumination path. The emission filter should be placed in front of other optical components in the detection path to reduce autofluorescence from those components. However, in many applications, it is not practical to place the emission filter as the first element. For example, in a fluorescence microscope, there is not enough space for the emission filter in front of the objective lens; in this case, the next optimal location is where the range of the ray angle is small.

Special attention should be given to selecting optical materials for fluorescence imaging systems whose excitation and emission paths share common optical elements, such as a microscope objective lens, and for systems whose excitation and emission paths do not share the same elements but where it is not appropriate to place the emission filter in front of the detection path.

As a general guideline, the optical elements in the detection path should be as few as possible to increase light transmission and minimize the autofluorescence of the optical components. This guideline also applies for the excitation path.

5.6.2 Illumination system design

The basic requirement of the illumination system for fluorescence imaging is to closely match the excitation peak of the fluorophore in order to excite the fluorescence effectively. Other considerations include illumination intensity and uniformity. The intensity must be sufficient to excite the fluorescence from the object. Illumination intensity and fluorescence signal intensity do not follow a linear correlation when the illumination intensity is too high or too low.

Fluorescence emission is generally less sensitive to the angle of illumination than irradiance. Therefore, there is no special requirement on the uniformity of the angular distribution. However, spatial uniformity is required in order to achieve high-contrast fluorescence images within the FOV.

For a conventional fluorescence microscope, Köhler illumination is employed to achieve uniform illumination. When there is no beam-shaping element to improve the spatial distribution of the light from the light source, the arc of the light source or the emitting surface should be imaged onto the entrance pupil of the objective lens for better illumination uniformity. The aperture may

not be fully filled, and the angular distribution may not be uniform, but the spatial distribution is better than when the light source is not well focused on the entrance pupil of the objective lens. Even when a liquid or fiber light guide is used, the output surface should also be imaged onto the entrance pupil of the objective lens for better light efficiency and illumination uniformity.

For other fluorescence imaging systems, such as macroimaging systems, beam-shaping elements (such as a lenslet array, light pipe, or aspheric elements) are usually used to provide spatially uniform excitation light to the samples. For some applications, an array of light sources, typically LEDs, is enough to provide excitation light with reasonable uniformity.

One more requirement for illumination is protecting the sample, especially in live-cell imaging applications. The sample must be protected from overexposure by attenuating the light intensity and by limiting the duration of the illumination to exactly the exposure time of the sensor; the sample must also be protected from the heat generated by the light source.

5.6.3 Detection systems

5.6.3.1 General requirements

As discussed in Sec. 5.5, there are a number of detection systems for fluorescence imaging. The basic requirements for the detection systems in a fluorescence imaging system include high-resolution, high-fluorescence signal collection and transmission. When the excitation and detection share the same objective lens, high transmission of the excitation light is an additional requirement.

Generally, high resolution and high light-collection efficiency are related through the NA. The larger the NA is, the higher the resolution and light-collection efficiency. In cases where the fluorescence signal is more critical than the resolution, it is often a better choice to design the imaging lens with a large NA but a lower aberration correction (fewer optical elements).

An infinite-conjugate imaging system is preferred, but not required, when the excitation and emission share the same objective lens, so that the dichroic mirror and other components can be inserted or removed between the objective lens and the tube lens without introducing aberrations and image shift. Generally, a telecentric imaging system is desired, especially for off-axis point scanning, so that the collection efficiency of the fluorescence signal is uniform across the entire FOV.

The transmission of both the excitation and emission light should be as high as possible. Antireflection (AR) coatings are essential for achieving high transmission. Depending on the excitation and emission wavelengths, as well as the locations of the optical elements, various types of coatings can be applied. For optical elements used in both the excitation and emission paths, either a W-shaped AR coating with two minimal reflection bands or a broadband AR coating is required. For elements used only in the excitation path, a V-shaped AR coating is enough to minimize the reflection of the excitation light. Similarly, a V-shaped

AR coating with minimal reflection at the emission wavelength can be sufficient for the elements in the detection path.

5.6.3.2 Optical Materials

5.6.3.2.1 Transmission of optical materials

UV light is often used in fluorescence imaging systems for some biomedical applications. For example, in a fluorescence microscope, UV light at 340 or 365 nm is commonly used. The transmittance of a microscope objective lens a short-wavelength region should be high enough, for example, 20% or more. In a short-wavelength region, there is a limitation on the optical materials having favorable transmittance. Also, the number of lenses in the objective lens should be limited to secure a reasonable transmittance.

Optical glasses are typically optimized for excellent transmittance throughout the total visible spectrum from 400–800 nm, and they usually do not transmit near-UV light well enough, especially dense flint glasses. For example, N-SF6, commonly used in negative optical elements for chromatic aberration correction, has an internal transmittance of 0.04 at 365 nm for a 10-mm slab. Therefore, those glasses with low transmission in the UV should be avoided in fluorescence imaging systems. Classical optical glasses, such as BK7, usually transmit near-UV light well, so they are often used in fluorescence imaging systems. Generally, glasses with a lower refractive index have higher transmittance far down to short wavelengths in the UV. According to the general theory of absorbing dielectric media, the UV absorption edge for higher-index glass moves closer to the visible range.⁴⁰

Heavy elements such as lead, barium, and titanium in the composition of the glass and the melting process largely determine the UV transmittance characteristics of glasses. Glasses with a high refractive index usually contain these heavy elements; therefore, the transmittance in the blue region is low. The UV transmittance characteristic depends on which heavy elements are used. For example, the transmittance of regular SF glasses with lead is better than N-SF glass types with titanium or niobium. Under the same melting processing, the eco version of heavy flint always exhibits a lower transmittance in the blue region. Figure 5.29 plots the transmittance of glasses F2, N-F2, SF6, and N-SF6. The UV absorption edge moves to the visible for N-type glasses.

Fused silica, CaF₂, FK5 HT, LLF1 HT, and LF5 HT are some glasses with excellent homogeneity and UV transmittance. For example, FK5 HT has a transmittance of 99% at 365 nm, and the transmittance of fused silica at 193 nm is 98%.

The transmittance of optical glasses may decrease with light radiation, especially UV light, because of the color centers generated by UV radiation. This effect is called solarization.⁴⁰ Generally, glass with low UV transmittance, such as F and SF types, have small solarization effects. For crown glasses with a

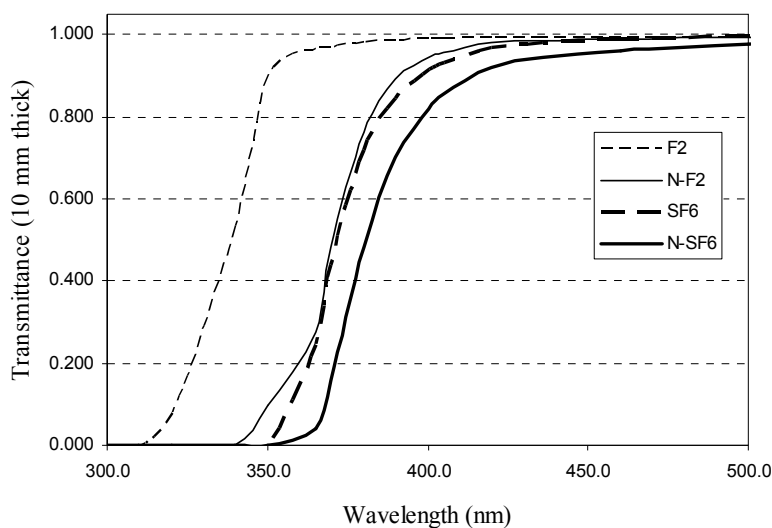


Figure 5.29 Transmittance of regular flint glasses with lead and their eco-versions with titanium or niobium.

higher UV transmittance, such as PSK, BaLK, K, ZK, BaK, SK, and LaK, the steepness of the transmittance becomes slightly smaller. Increasing the temperature also shifts the UV transmittance edge to a longer wavelength.

Figure 5.30 compares the UV transmittance of equivalent glasses from different manufacturers. N-BK7 from Schott and S-BSL7 from Ohara have similar transmittance characteristics, except that there is a slight drop for N-BK7 around 330 nm. However, the difference between N-PSK53 and S-PHM52 is significant: N-PSK53 has much better UV transmittance. When designing an optical system for fluorescence imaging, it may be helpful to compare the transmittance characteristics of the glass when substituting glasses with the same refractive index and dispersion.

5.6.3.2.2 Autofluorescence of optical materials

Another critical optical property of optical materials for fluorescence imaging is autofluorescence. Autofluorescence of optical components is considered an isotropic generation of secondary stray light inside the system; it is undesirable and should be minimized because it is the major source of unwanted light in fluorescence imaging systems.

Autofluorescence in optical glass is generated by color centers, which are originated by rare-earth elements and other critical impurities in optical glasses.⁴¹ Fluorescence excited by UV light correlates strongly with the content of UV-absorbing ions in materials such as the raw earth elements iron, zinc, and vanadium. Fluorescence characteristics are also influenced by existing process technology.

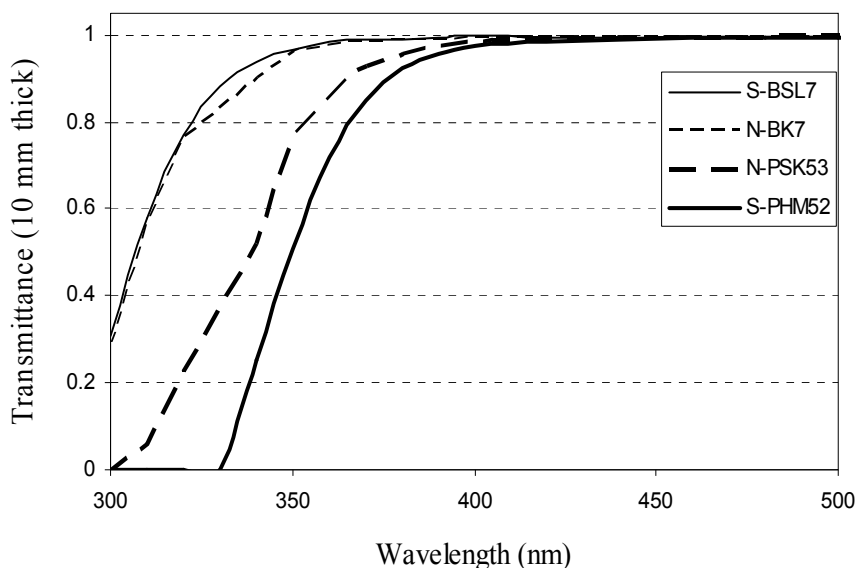


Figure 5.30 Transmittance of equivalent glasses from Ohara Corp. and Schott AG.

The autofluorescence spectra of most optical glasses are continuous with a similar shape. Figure 5.31(a) illustrates the emission spectra of selected low-autofluorescence optical glasses using mercury excitation light at 365 nm.⁴¹ The emission spectra have typical maxima at 435 or 525 nm, depending on the material composition. For N-type glasses, these features are driven by the elimination of PbO and As₂O₃ and the substitution of Sb₂O₃ or ZnO. Figure 5.31(b) shows the emission spectra of some glasses using a 532-nm excitation wavelength. Compared with Fig. 5.31(a), there is no correlation in fluorescence values between 365- and 532-nm excitation wavelengths because the UV excitation and absorption depend mainly on impurities.⁴¹

In practice, it is common to scale the fluorescent light power using one of the common glasses as a reference. Figure 5.31(c) shows the relative integral autofluorescence, which is calculated by integrating intensity values of the fluorescence spectrum between 400 and 700 nm then dividing the obtained result by that of the glass SF1.^{41,42} It is clear that flint types of glass have stronger autofluorescence.

In addition to low cost and light weight, plastic optics offers design freedoms that are not achievable or economical with glass optics. Plastic optics is finding more and more applications in biomedical imaging. However, plastic materials have significant autofluorescence when excited by UV or visible radiation.

Studies have been performed to measure the autofluorescence of plastic materials used in biomedical imaging.^{43,44} There are several unique characteristics of autofluorescence from plastic optics. Compared to optical glass

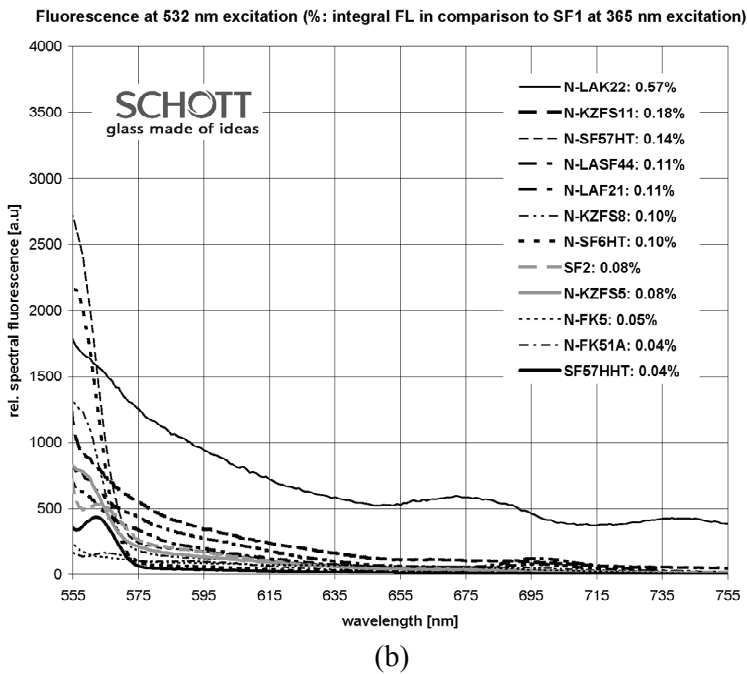
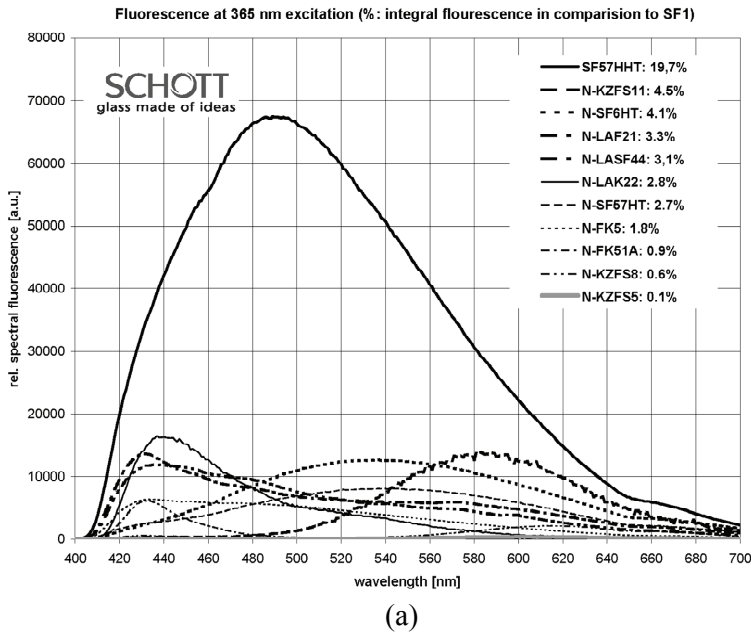
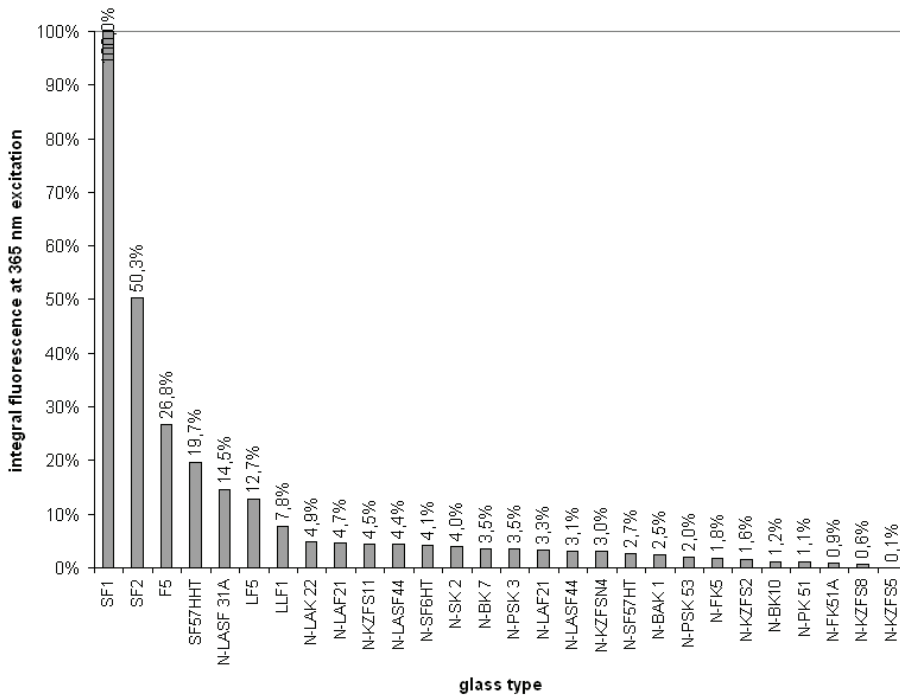


Figure 5.31 (a) Emission spectra of low-fluorescence optical glasses using a 365-nm excitation wavelength, and (b) emission spectra of optical glass using a 532-nm excitation wavelength and without a cut-off filter. (Courtesy of Schott AG.) [Continued on next page.]



(c)

Figure 5.31 (continued) (c) Typical integral fluorescence data relative to SF1 using 365 nm excitation. (Courtesy of Schott AG.)

whose autofluorescence is unaffected by the illumination time, the autofluorescence of plastic optics decreases quickly immediately after exposure and then decays gradually. The autofluorescence intensity approaches a limiting value after a period of a few hundred seconds, and it generally scales linearly with the excitation light power. Typically, the autofluorescence of plastic materials decreases as the excitation wavelength increases.

The autofluorescence of plastic optics is significantly higher than that of optical glass, especially in the short wavelengths. At a short wavelength near 400 nm, the autofluorescence of plastic optics is about 5–10 times higher than Schott Borofloat glass. At longer excitation wavelengths, the difference becomes smaller. The autofluorescence from cyclo olefin copolymers (COCs), as well as cyclic olefin polymers (COPs), is relatively smaller than from other optical plastics. Polycarbonate (PC) exhibits significantly higher autofluorescence. The autofluorescence of polymethyl methacrylate (PMMA) falls between that of COC and PC.

Generally, it is difficult to prevent autofluorescence from optical materials from reaching the detector. Therefore, it is critical to select optical materials with low autofluorescence coefficients in order to obtain high-contrast fluorescence images.

5.6.3.3 Design of a fluorescence microscope objective lens

One of the challenges when designing an objective lens for a wavelength below 350 nm is that glass materials with low autofluorescence and high transmittance in the UV usually have low refractive indices, thereby making it difficult to obtain a flat image surface. The curvature of field is often sacrificed in order to excite fluorescence and collect fluorescent light more efficiently. In addition, cemented elements are not commonly used in UV because cement materials may absorb UV light strongly and exhibit autofluorescence.

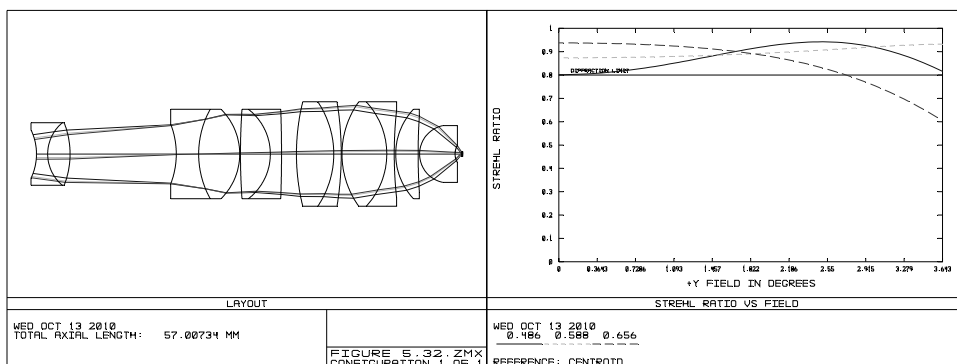
It is also difficult to achromatize optical systems by combining optical glasses having comparatively close dispersions while attaining a large NA. In order to correct the chromatic aberrations using optical glasses with small differences in the refractive index, the radii of the lens element surfaces tend to be small. As a general rule, glasses with an Abbe number of less than 50 and a refractive index greater than 1.70 should be avoided.

Spherical aberration in a standard objective lens for a reflected-light fluorescence microscope is usually corrected for visible light. It is not essential to correct for near-UV light because near-UV light is used for excitation only, and the illuminated area can be slightly larger than the FOV. Additional optical elements can be added in the illumination path to achieve the parfocality of the excitation and emission locations. The axial chromatic aberration in a fluorescence microscope is not very noticeable because the fluorescent light is usually close to being quasi-monochromatic light.

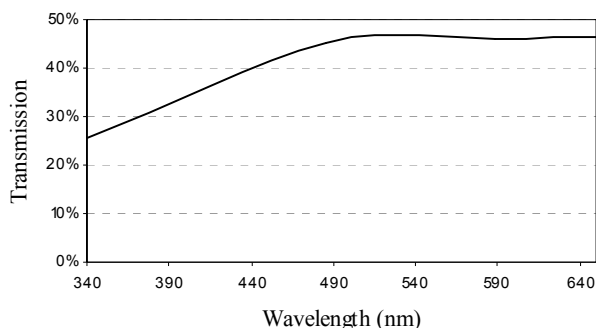
Microscope objective lenses converge the excitation light at the front lens elements; the intensity there becomes much stronger than at other locations. Therefore, the optical materials used in the front lenses must have as little autofluorescence as possible. Also, the material for the first lens must be strong physically and chemically because it is exposed to air or an index-matching liquid.

Figure 5.32(a) shows an objective lens designed for a fluorescence microscope. The lens is optimized for an infinite conjugate and is used with a tube lens to achieve the desired magnification. It is telecentric in the sample space. The maximum chief ray angle is 0.5 deg. The NA is 0.65, and the half FOV is 0.46 mm. The glasses used in the lens meet the above requirements, namely, high transmission and low autofluorescence. The relative autofluorescence coefficients of the glasses used in this objective lens are less than 8% of SF1. N-PK51 and N-SK2, whose relative autofluorescence coefficients are 1.08% and 4% of SF1, are used for the two singlets in the front group. The lens transmittance without coating from 340–650 nm is plotted in Fig. 5.32(b). The transmittance at 390 nm is approximately 33%.

As shown in the plot of the Strehl ratio, the lens has a diffraction-limited performance in the entire field for the blue and green wavelengths and 60% of the field for the red wavelength. The major remaining aberrations include spherical aberration and distortion. This objective lens is not optimized for the



(a)



(b)

Figure 5.32 An objective lens for a fluorescence microscope. (a) The lens layout and its performance in the visible spectrum and (b) the transmittance at 350 nm.

UV wavelength. Since UV light is used as the excitation light, it does not require diffraction-limited performance. The drawback is that the light efficiency is low because the illumination spot is larger than the detection spot. The efficiency can be improved by adjusting the illumination path to compensate for the aberration remaining in the objective lens in the UV range.

In order to minimize the autofluorescence from the objective lens, the materials should have a high excitation light transmission and should generate little autofluorescence. However, the materials meeting the above requirements typically have a low refractive index and a similar Abbe number. Thus, it is difficult to correct the aberrations well enough for high-end applications.

An alternative solution is to divide the objective lens into a front lens group and a rear lens group.⁴⁵ The materials used in the front group should have substantially higher excitation light transmission and should generate less autofluorescence. The excitation light is introduced into the objective lens by a dichroic mirror located between the front and rear lens groups, as shown in Fig. 5.33, and it reaches the sample through the front lens group. Most of the

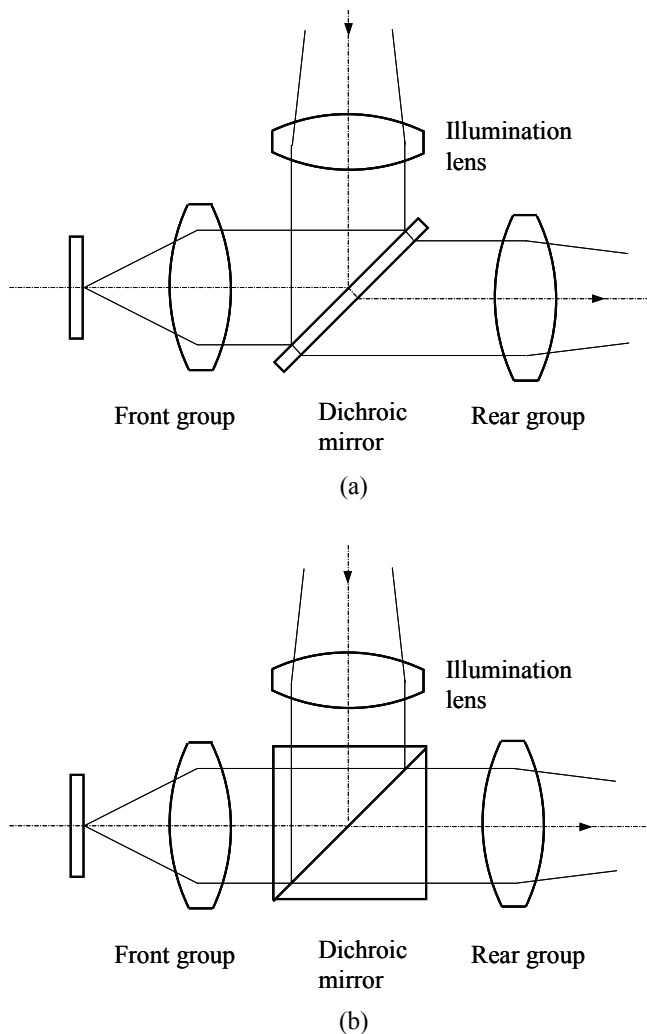


Figure 5.33 An objective lens with two lens groups. The excitation light is introduced by (a) a dichroic mirror plate or (b) a dichroic beamsplitter cube between the lens groups.

reflected excitation light from the sample and lens surfaces in the front group is blocked by the dichroic mirror and cannot reach the rear lens group. Therefore, conventional optical materials can be used in the rear group to correct the various aberrations remaining from the front lens group.

In order to minimize the aberrations introduced by the dichroic mirror plate in Fig. 5.33(a), the front group is preferred to be infinitely conjugated so that the light is approximately collimated between the two lens groups. The alternative is to use a dichroic beamsplitter cube, which does not introduce a beam shift, as shown in Fig. 5.33(b).

Figure 5.34 illustrates a water-immersion objective lens using this design principle. The first three lenses in the front group are made from fused silica, which has excellent transmission in the UV and has low autofluorescence. The ray angle between the front group and the rear group is very small. In order to correct the chromatic aberration remaining in the front group, two special optical glasses, CAF2 and PK50, are used in the rear group. This objective lens has an NA of 0.9 and an FOV of 0.28 mm. Chromatic aberration of this objective lens is well controlled, but spherical aberration and field curvature need further improvement.

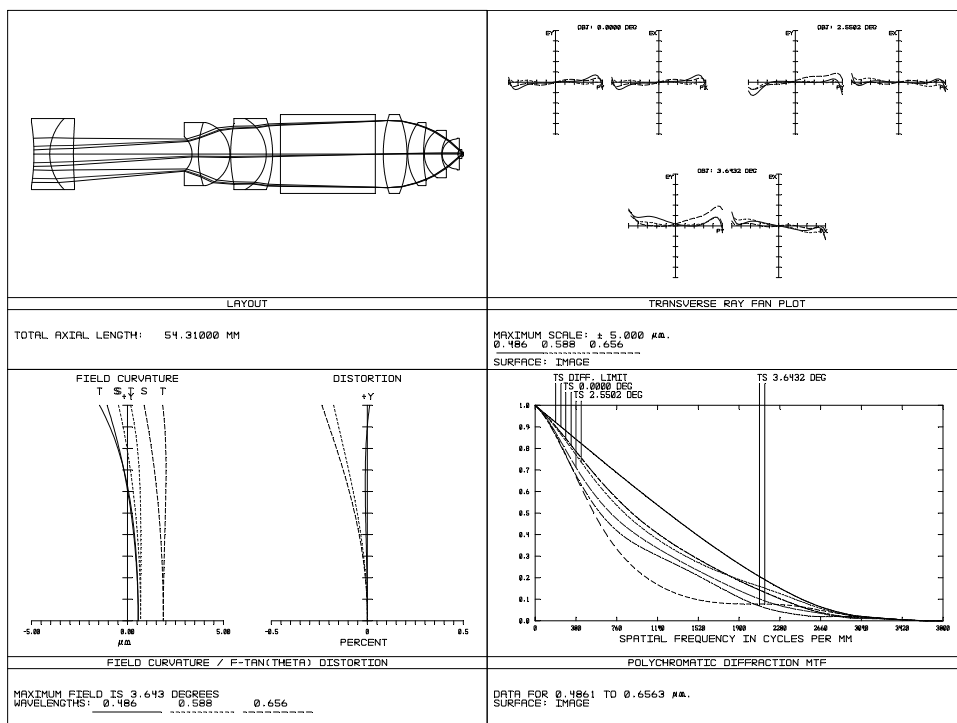


Figure 5.34 Objective lens consisting of a front and a rear group.

References

1. G. Q. Stokes, "On the change of refrangibility of light," *Philos. Trans. R. Soc. Lond.* **142**, 463–562 (1852).
2. T. Stübel, "Die Fluoreszenz tierischer Gewebe in ultravioletten Licht," *Pflügers Arch.* **142**(1), 1–14 (1911).
3. F. H. J. Figue, G. S. Weiland, and L. O. J. Manganiello, "Cancer detection and therapy: Affinity of neoplastic, embryonic and traumatised tissues for porphyrins and metalloporphyrins," *Proc. Soc. Exp. Biol. Med.* **68**, 640–641 (1948).

4. J. Winkelman and D. S. Rasmussen-Taxdal, "Quantitative determination of porphyrin uptake by tumor tissue following parenteral administration," *Bull. Johns Hopkins Hosp.* **107**, 228–233 (1960).
5. Amersham Biosciences, *Fluorescence Imaging Principles and Methods*, Piscataway, NJ (2002).
6. R. S. DaCosta, B. C. Wilson, and N. E. Marcon, "Photodiagnostic techniques for the endoscopic detection of premalignant gastrointestinal lesions," *Dig. Endosc.* **15**(3), 153–173 (2003).
7. A. A. Bogdanov, C. P. Lin, M. Simonova, L. Matuszewski, and R. Weissleder, "Cellular activation of the self-quenched fluorescent reporter probe in tumor microenvironment," *Neoplasia* **4**, 228–236 (2002).
8. E. de Josselin de Jong, F. Sundstrom, H. Westerling, S. Tranaeus, J. J. ten Bosch, and B. Angmar-Mansson, "A new method for *in vivo* quantification of changes in initial enamel caries with laser fluorescence," *Caries Res.* **29**, 2–7 (1995).
9. R. P. Rava, R. Richards-Kortum, M. Fitzmaurice, R. Cothren, R. Petras, M. Sivak, H. Levin, and M. S. Feld, "Early detection of dysplasia in colon and bladder tissue using laser induced fluorescence," *Proc. SPIE* **1426**, 68–78 (1991) [doi: 10.1117/12.44048].
10. D. Roblyer, C. Kurachi, V. Stepanek, M. D. Williams, A. K. El-Naggar, J. J. Lee, A. M. Gillenwater, and R. Richards-Kortum, "Objective detection and delineation of oral neoplasia using autofluorescence imaging," *Cancer Prev. Res.* **2**, 423–431 (2009).
11. S. Montán, K. Svanberg, and S. Svanberg, "Multi-color imaging and contrast enhancement in cancer tumor localization using laser-induced fluorescence in hematoporphyrin derivative (HpD)-bearing tissue," *Opt. Lett.* **10**, 56–58 (1985).
12. P. S. Andersson, S. Montán, and S. Svanberg, "Multi-spectral system for medical fluorescence imaging," *IEEE J. Quantum Electron.* **23**, 1798–1805 (1987).
13. S. Y. Park, T. Collier, J. Aaron, M. Markey, R. Richards-Kortum, K. Sokolov, N. Mackinnon, C. MacAulay, L. Coghlan, A. Milbourne, and M. Follen, "Multispectral digital microscopy for *in vivo* monitoring of oral neoplasia in the hamster cheek pouch model of carcinogenesis," *Opt. Express* **13**, 749–762 (2005).
14. M. Yang, L. Li, P. Jiang, A. R. Moossa, S. Penman, and R. M. Hoffman, "Dual-color fluorescence imaging distinguishes tumor cells from induced host angiogenic vessels and stromal cells," *Proc. Natl. Acad. Sci.* **100**, 14,259–14,262 (2003).
15. U. Gustafsson, E. McLaughlin, E. Jacobson, J. Håkansson, P. Troy, M. J. DeWeert, S. Pålsson, M. Soto Thompson, S. Svanberg, A. Vaitkuviene, and K. Svanberg, "Fluorescence and reflectance monitoring of human cervical

- tissue in vivo: A case study,” *Proc. SPIE* **4959**, 100–110 (2003) [doi: 10.1117/12.479495].
16. U. Gustafsson, S. Pålsson, and S. Svanberg, “Compact fibre-optic fluorosensor using a continuous wave violet diode laser and an integrated spectrometer,” *Rev. Sci. Instrum.* **71**, 3004–3006 (2000).
 17. T. H. Pham, C. Eker, A. Durkin, B. J. Tromberg, and S. Andersson-Engels, “Quantifying the optical properties and chromophore concentrations of turbid media by chemometric analysis of hyperspectral, diffuse reflectance data collected using a Fourier interferometric imaging system,” *Appl. Spectrosc.* **55**, 1035–1045 (2001).
 18. R. Cubeddu, G. Canti, A. Pifferi, P. Taroni, and G. Valentini, “Real time system for fluorescence lifetime imaging,” *Proc. SPIE* **2976**, 98–104 (1997) [doi: 10.1117/12.275518].
 19. D. Elson, J. Requejo-Isidro, I. Munro, F. Reavell, J. Siegel, K. Suhling, P. Tadrous, R. Benninger, P. Lanigan, J. McGinty, C. Talbot, B. Treanor, S. Webb, A. Sandison, A. Wallace, D. Davis, J. Lever, M. Neil, D. Phillips, G. Stamp, and P. French, “Time-domain fluorescence lifetime imaging applied to biological tissue,” *Photochem. Photobiol. Sci.* **3**, 795–801 (2004).
 20. A. D. Elder, S. M. Matthews, J. Swartling, K. Yunus, J. H. Frank, C. M. Brennan, A. C. Fisher, and C. F. Kaminski, “Application of frequency-domain fluorescence lifetime imaging microscopy as a quantitative analytical tool for microfluidic devices,” *Opt. Express* **14**, 5456–5467 (2006).
 21. N. MacKinnon, U. Stange, P. Lane, C. MacAulay, and M. Quatrevalet, “Spectrally programmable light engine for in vitro or in vivo molecular imaging and spectroscopy,” *Appl. Opt.* **44**, 2033–2040 (2005).
 22. B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert, M. Steinhauser, G. Eckstein, D. Schmitt-Landsiedel, and R. Thewes, “A 128 × 128 CMOS biosensor array for extracellular recording of neural activity,” *IEEE J. Solid-State Circuits* **38**(12), 2306–2317 (2003).
 23. U. Lu, B. Hu, Y. Shih, C. Wu, and Y. Yang, “The design of a novel complementary metal oxide semiconductor detection system for biochemical luminescence,” *Biosens. Bioelectron.* **19**(10), 1185–1191 (2004).
 24. H. Ji, P. A. Abshire, M. Urdaneta, and E. Smela, “CMOS contact imager for monitoring cultured cells,” in *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 3491–3495 (2005).
 25. J. A. Dobrowolski, G. E. Marsh, D. G. Charbonneau, J. Eng, and P. D. Josephy, “Colored filter glasses: an intercomparison of glasses made by different manufacturers,” *Appl. Opt.* **16**, 1491–1512 (1977).
 26. W. E. Van Derveer and D. Wolpert, “Optics for microscopy: Fluorescence of colored-glass filters can be stronger than expected,” *Laser Focus World* **44**(10), 86–89 (Oct. 2008).

27. J. Reichman, *Handbook of Optical Filters For Fluorescence Microscopy*, Chroma Technology Corp., Brattleboro, VT (2000).
28. T. Erdogan, A. Pradhan, and V. Mizrahi, "Optical filters impact fluorescence fidelity," *Biophotonics Int.* **10**(11), 38–42 (2003).
29. K. Salama, H. Eltoukhy, A. Hassibi, and A. El Gamal, "Modeling and simulation of luminescence detection platforms," *Biosens. Bioelectron.* **19**, 1377–1386 (2004).
30. S. Parikh, G. Gulak, and P. Chow, "A CMOS image sensor for DNA microarrays," in *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 821–824, (2007).
31. S. Leavesley, Y. Jiang, V. Patsekin, H. Hall, D. Vizard, and P. J. Robinson, "Hyperspectral small animal fluorescence imaging: spectral selection imaging," *Proc. SPIE* **6850**, 68500X (2008) [doi: 10.1117/12.763935].
32. A. M. De Grand and J. V. Frangioni, "An operational near-infrared fluorescence imaging system prototype for large animal surgery," *Technol. Cancer Res. Treat.* **2**, 553–562 (2003).
33. A. Nakayama, F. del Monte, R. J. Hajjar, and J. W. Frangioni, "Functional near-infrared fluorescence imaging for cardiac surgery and targeted gene therapy," *Mol. Imaging* **1**, 365–377 (2002).
34. K. L. Reichenbach and C. Xu, "Numerical analysis of light propagation in image fibers or coherent fiber bundles," *Opt. Express* **15**, 2151–2165 (2007).
35. A. Komiyama and M. Hashimoto, "A new class of crosstalk in image fibers," *Opt. Commun.* **107**, 49–53 (1994).
36. J. A. Udovich, N. D. Kirkpatrick, A. Kano, A. Tanbakuchi, U. Utzinger, and A. F. Gmitro, "Spectral background and transmission characteristics of fiber optic imaging bundles," *Appl. Opt.* **47**, 4560–4568 (2008).
37. R. Liang, V. Wong, M. Marcus, P. Burns, and P. McLaughlin, "Multimodal imaging system for dental caries detection," *Proc. SPIE* **6425**, 642502 (2007) [doi: 10.1117/12.702131].
38. R. J. McNichols, A. Gowda, B. A. Bell, R. M. Johnigan, K. H. Calhoun, and M. Motamedi, "Development of an endoscopic fluorescence image-guided OCT probe for oral cancer detection," *Proc. SPIE* **4254**, 23–30 (2001) [doi: 10.1117/12.427941].
39. R. Liang, V. Wong, and P. McLaughlin, "Optical design of a multimodal imaging system," *Proc. SPIE* **6849**, 684908 (2008) [doi: 10.1117/12.427941].
40. *Schott Technical Information TIE-35: Transmittance of optical glass*, Schott AG, Duryea, PA (2005).
41. *Schott Technical Information TIE-36: Fluorescence of optical glass*, Schott AG, Duryea, PA (2009).

42. H. Gross, F. Blechinger, and B. Aichtner, *Handbook of Optical Systems*, vol. 4, *Survey of Optical Instruments*, Wiley-VCH, Berlin (2008).
43. A. Piruska, I. Nikcevic, S. H. Lee, C. Ahn, W. R. Heineman, P. A. Limbach, and C. J. Seliskar, “The autofluorescence of plastic materials and chips measured under laser irradiation,” *Lab Chip* **5**, 1348–1354 (2005).
44. K. R. Hawkins and P. Yager, “Nonlinear decrease of background fluorescence in polymer thin-films—a survey of materials and how they can complicate fluorescence detection in mTAS,” *Lab Chip* **3**, 248–252 (2003).
45. K. Tomimatsu and Y. Shimizu, “Objective lens unit for reflective fluorescence and reflective fluorescence microscope having an objective lens unit for reflective fluorescence,” U.S. Patent No. 6,219,189.

Chapter 6

Polarization Imaging

Bickel et al. investigated the polarization effect on the scattering of biological scatterers in 1976,¹ but only since 1991 has polarized light started to attract a great deal of attention in the field of biomedical optical imaging.² It has been demonstrated that polarization methods can distinguish less-scattered photons from diffusive ones,^{3,4} improve image quality,⁵ obtain better surface features,⁶⁻⁹ and provide information on the properties of tissues.¹⁰⁻¹²

In a polarization imaging system, the light from the light source is first conditioned by the polarization elements, such as the polarizer and the retarder, then it illuminates the object under investigation. The polarized light interacts with the object through reflection, scattering, and absorption. Part of the light, either the backscattered light with same wavelength or the fluorescence light with a different wavelength, is collected by the collection optics and delivered to the sensor. One or more polarization elements are typically placed in the detection path to select the light with the desired polarization state.

The fundamentals of polarized light and the interactions of polarized light with tissues will be discussed in Secs. 6.1 and 6.2. Sections 6.3 and 6.4 will summarize the polarization imaging systems for biomedical optical imaging and discuss polarization elements, and Sec. 6.5 will discuss optical design for polarization imaging systems.

6.1 Basics of Polarized Light

Light can be described as an electromagnetic wave. The electric field \mathbf{E} of a monochromatic electromagnetic wave can be presented as the vector sum of two electrical fields in the x and y planes, which are perpendicular to each other (as shown in Fig. 6.1):

$$\mathbf{E}(t) = \hat{x} E_x e^{i(kz - \omega t + \phi_x)} + \hat{y} E_y e^{i(kz - \omega t + \phi_y)}. \quad (6.1)$$

During propagation, the electrical field vector changes its orientation and magnitude. The relative magnitudes E_x and E_y and phases ϕ_x and ϕ_y of the

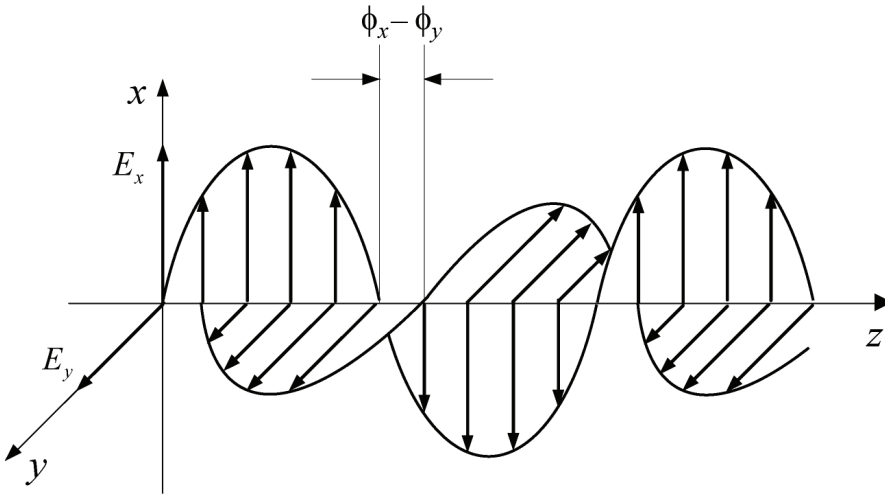


Figure 6.1 Electrical field vectors in the x and y planes.

electrical fields specify the polarization status of the light ranging from completely unpolarized to completely polarized.

The general state of a completely polarized wave is elliptical, which means that the electrical field rotates elliptically around the propagation direction, as shown in Fig. 6.2. The orientation ψ of the polarization ellipse is defined as

$$\tan(2\psi) = \frac{2E_x E_y}{E_x^2 - E_y^2} \cos \delta, \quad 0 \leq \psi \leq \pi, \quad (6.2)$$

where $\delta = \phi_y - \phi_x$. The ellipticity angle χ of the polarization ellipse is

$$\tan 2\chi = \frac{2E_x E_y}{E_x^2 - E_y^2} \sin \delta, \quad \frac{\pi}{4} \leq \chi \leq \frac{\pi}{4}. \quad (6.3)$$

The ellipticity e is defined as

$$e = \pm \frac{b}{a}, \quad (6.4)$$

where a and b are the half-lengths of the long and short axes, respectively. The ellipticity angle χ is the arctangent of ellipticity e , $\chi = \arctan(e)$.

When ϕ_x and ϕ_y are equal, the wave is linearly polarized [Fig. 6.3(a)]. The orientation β is a constant,

$$\beta = \arctan \frac{E_y}{E_x}. \quad (6.5)$$

When ϕ_y advances ϕ_x by $\pi/2$, the wave is left circularly polarized, which means \mathbf{E} traverses a circle counterclockwise when looking toward the direction of propagation [Fig. 6.3(b)]. Conversely, when ϕ_y lags behind ϕ_x by $\pi/2$, the wave is right circularly polarized [Fig. 6.3(c)].

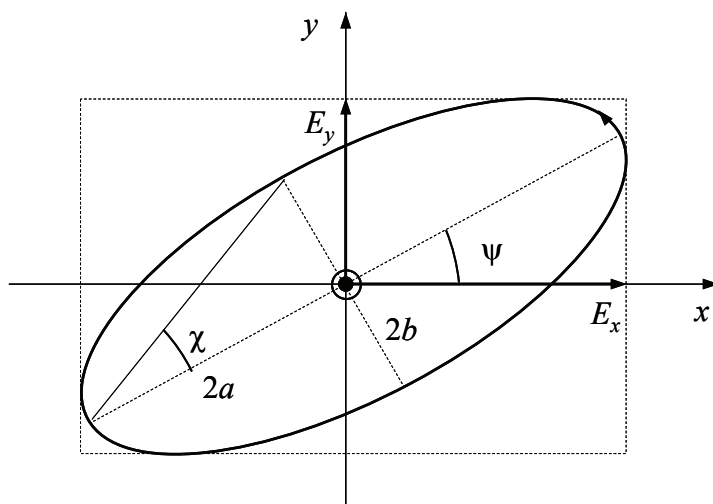


Figure 6.2 Elliptical polarization state. Parameters a and b are the half-lengths of the long and short axes, ψ is the orientation angle, and χ is the ellipticity angle of the ellipse.

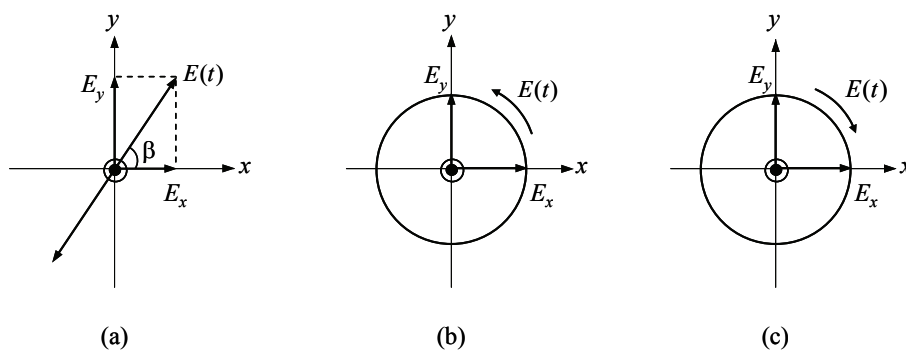


Figure 6.3 Polarization states of light: (a) linearly polarized, (b) left circularly polarized, and (c) right circularly polarized.

The polarization ellipse is an excellent way to visualize polarized light. However, it is difficult to determine the orientation and ellipticity angle by viewing the polarization ellipse. To overcome this limitation, the Poincaré sphere was introduced to represent polarized light.^{13,14} The limitation of both the Poincaré sphere and the polarization ellipse is that neither the rotation angle nor the ellipticity angle is directly measurable. It is also difficult to determine the new angles of a polarized beam that propagates through one or more polarizing elements. The Jones and Stokes vectors are two methods commonly used for describing polarization states with measurable components.

The Jones vector \mathbf{E} , proposed by R. C. Jones in 1941, is a convenient method to represent the polarization state of a monochromatic plane wave propagating along a z axis by a complex two-component vector:

$$\mathbf{E} = \begin{pmatrix} E_x e^{i\phi_x} \\ E_y e^{i\phi_y} \end{pmatrix}. \quad (6.6)$$

The Jones vector contains the state of polarization, the amplitude, and the phase of the wave. Usually, the Jones vector is normalized so that the magnitude of the elements equals 1. Two Jones vectors are orthogonal if their interference contrast is 0, $\mathbf{E}_1 \mathbf{E}_2 = 0$. If the norm of the two orthogonal Jones vectors is 1, they form an orthogonal set, which is the basis for all completely polarized states. Table 6.1 lists the normalized Jones vectors for basic polarization states.

Unpolarized light cannot be represented by a single Jones vector because Jones vectors have only two components to describe completely polarized states. However, unpolarized light can be represented by the incoherent superposition of two orthogonal polarization states with equal intensity.

When the incident and exiting light are described by Jones vectors, a 2×2 matrix, called the Jones matrix, is used to characterize the polarization transformation of a single interaction or a cascade of polarization effects generated by polarization elements. The Jones calculus is a system for performing polarization calculations. The basic Jones matrix equation is

$$\mathbf{E}' = J \mathbf{E} = \begin{pmatrix} J_{xx} & J_{yx} \\ J_{xy} & J_{yy} \end{pmatrix} \begin{pmatrix} E_x e^{i\phi_x} \\ E_y e^{i\phi_y} \end{pmatrix}, \quad (6.7)$$

where \mathbf{E} and \mathbf{E}' are the Jones vectors for the incident and exiting light, respectively. J is the Jones matrix of the polarization element. The elements of the Jones matrix, J_{xx} , J_{xy} , J_{yx} , and J_{yy} , are dimensionless. The diagonal elements of the Jones matrix describe the change in the incident polarization state, while the nondiagonal elements represent cross polarization. Table 6.2 lists the Jones matrices of basic polarizers and retarders.

Table 6.1 Normalized Jones vectors of selected polarization states.

Polarization state	Jones vector
Linear (horizontal)	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$
Linear (vertical)	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$
Linear (45 deg)	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
Linear (135 deg)	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$
Linear (θ)	$\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$
Right circular	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$
Left circular	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$
Right elliptical	$\frac{1}{\sqrt{E_x^2 + E_y^2}} \begin{pmatrix} E_x \\ -iE_y \end{pmatrix}$
Left elliptical	$\frac{1}{\sqrt{E_x^2 + E_y^2}} \begin{pmatrix} E_x \\ iE_y \end{pmatrix}$

The Jones calculus is intended for coherent monochromatic light calculation and is an effective tool for solving polarization problems in optical systems. A series of optical elements can be described by the multiplication of the Jones matrices associated with each element:

$$J_{\text{sys}} = J_n J_{n-1} \dots J_2 J_1, \quad (6.8)$$

where J_{sys} is the Jones matrix of an entire system and J_i is the Jones matrix of the i th element. The first element that the light passes is on the right side of the sequence, and the last element n is on the left side of the sequence.

Table 6.2 Jones matrices of selected polarization elements.

Polarization element	Jones matrix
Linear polarizer (horizontal)	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$
Linear polarizer (vertical)	$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$
Linear polarizer (θ)	$\begin{pmatrix} \cos^2 \theta & \cos \theta \sin \theta \\ \cos \theta \sin \theta & \sin^2 \theta \end{pmatrix}$
Right circular polarizer	$\frac{1}{2} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$
Left circular polarizer	$\frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$
Retarder ($\lambda/4$) with fast axis horizontal	$e^{i\frac{\pi}{4}} \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$
Retarder ($\lambda/2$) with fast axis horizontal	$\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$
Retarder (δ) with fast axis horizontal	$e^{-i\frac{\delta}{2}} \begin{pmatrix} 1 & 0 \\ 0 & e^{i\delta} \end{pmatrix}$
Retarder (δ) with fast axis vertical	$e^{i\frac{\delta}{2}} \begin{pmatrix} 1 & 0 \\ 0 & e^{i\delta} \end{pmatrix}$

Another method to describe a polarization state is the Stokes vector, which was introduced by George Stokes in 1852. Any light, polarized or nonpolarized, coherent or noncoherent, can be represented by a Stokes vector. The Stokes vector \mathbf{S} summarizes the balance of the measured irradiance between horizontal linearly (I_H) versus vertical linearly polarized (I_V) components, +45-deg linearly (I_{45}) versus 135-deg linearly polarized (I_{135}) components, and right circularly (I_R) versus left circularly polarized (I_L) components:

$$\mathbf{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} I_H + I_V \\ I_H - I_V \\ I_{45} - I_{135} \\ I_R - I_L \end{pmatrix}, \quad (6.9)$$

where S_0 is the total irradiance of the beam, S_1 is the difference of the irradiance between the horizontal and vertical linearly polarized components, S_2 is the difference between 45-deg and 135-deg linearly polarized components, and S_3 is the difference between the right and left circularly polarized components. S_1 , S_2 , and S_3 are regarded as polarization-difference images.

The Stokes parameters are related to the orientation and ellipticity angles, ψ and χ , as follows:

$$S_1 = S_0 \cos(2\psi) \cos(2\chi), \quad (6.10a)$$

$$S_2 = S_0 \sin(2\psi) \cos(2\chi), \quad (6.10b)$$

$$S_3 = S_0 \sin(2\chi), \quad (6.10c)$$

$$\psi = \frac{1}{2} \arctan\left(\frac{S_2}{S_1}\right), \quad (6.10d)$$

$$\chi = \frac{1}{2} \arctan\left(\frac{S_3}{S_0}\right). \quad (6.10e)$$

The Stokes vector of a beam can be measured by using six different ideal polarizers (the horizontal and vertical linear polarizers, the 45-deg and 135-deg linear polarizers, and the right and left circular polarizers) in front of a detector. Table 6.3 lists the Stokes vectors of basic polarization states.

The degree of polarization (DOP) is used to describe the extent to which the beam is polarized:

$$\text{DOP} = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0}. \quad (6.11)$$

A DOP of 1 corresponds to completely polarized light, whereas a DOP of 0 corresponds to completely unpolarized light.

The degree of linear polarization (DOP_L) describes the extent to which the polarization ellipse distribution tends to be linear:

Table 6.3 Stokes vectors of selected polarization states.

Polarization state	Stokes vector
Linear (horizontal)	$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$
Linear (vertical)	$\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$
Linear (45 deg)	$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$
Linear (135 deg)	$\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$
Linear (θ)	$\begin{pmatrix} 1 \\ \cos 2\theta \\ \sin 2\theta \\ 0 \end{pmatrix}$
Right circular	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$
Left circular	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$

$$\text{DOP}_L = \frac{\sqrt{S_1^2 + S_2^2}}{S_0}. \quad (6.12)$$

The degree of circular polarization (DOP_C) describes the tendency of the beam to be circularly polarized:

$$\text{DOP}_C = \frac{S_3}{S_0}. \quad (6.13)$$

$\text{DOP}_C = +1$ indicates right circularly polarized light, $\text{DOP}_C = -1$ indicates left circularly polarized light, and $\text{DOP}_C = 0$ indicates either linearly polarized or unpolarized light.

The Mueller matrix M is the matrix that characterizes the transport of polarized light through an optical element or medium by the relationship between the incident Stokes vector \mathbf{S}_{in} and the output Stokes vector \mathbf{S}_{out} :

$$\mathbf{S}_{\text{out}} = M \mathbf{S}_{\text{in}}, \quad (6.14a)$$

$$\begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix}_{\text{out}} = \begin{pmatrix} M_{00} & M_{01} & M_{02} & M_{03} \\ M_{10} & M_{11} & M_{12} & M_{13} \\ M_{20} & M_{21} & M_{22} & M_{23} \\ M_{30} & M_{31} & M_{32} & M_{33} \end{pmatrix} \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix}_{\text{in}}. \quad (6.14b)$$

The Mueller matrix is determined only by the intrinsic properties of the optical element or medium. It can fully characterize the optical polarization properties of the optical element and optical system.

A set of 16 measures is acquired from a combination of four sources and four detectors in order to obtain a data matrix that can be converted into a Mueller matrix. The measurement and method to convert the data matrix into a Mueller matrix is outside the scope of this book; for more information, refer to Ref. 15. Table 6.4 lists the Mueller matrices of basic polarizers and retarders.

The effect of a series of polarization elements is described by a matrix multiplication of their Mueller matrix M_i ,

$$M = M_n M_{n-1} \dots M_2 M_1, \quad (6.15)$$

where i is the index of the polarization element. The first element that the light passes is on the right side of the sequence, and the last element n is on the left side of the sequence.

Table 6.4 Mueller matrices of selected polarization elements.

Polarization element	Mueller matrix
Linear polarizer (horizontal)	$\frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Linear polarizer (vertical)	$\frac{1}{2} \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Linear polarizer (θ)	$\frac{1}{2} \begin{pmatrix} 1 & \cos 2\theta & \sin 2\theta & 0 \\ \cos 2\theta & \cos^2 2\theta & \cos 2\theta \sin 2\theta & 0 \\ \sin 2\theta & \cos 2\theta \sin 2\theta & \sin^2 2\theta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Right circular polarizer	$\frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$
Left circular polarizer	$\frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$
Retarder ($\lambda/4$) with fast axis horizontal	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$
Retarder ($\lambda/2$)	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$
Retarder (δ)	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \delta & \sin \delta \\ 0 & 0 & -\sin \delta & \cos \delta \end{pmatrix}$

A special matrix used in Mueller calculus is the rotation matrix $R(\theta)$, which is used to represent the rotation of coordinate axes or the rotation of the polarization ellipse,

$$R(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & -\sin 2\theta & 0 \\ 0 & \sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (6.16)$$

where θ is the rotation angle.

The above three methods, namely, the polarization ellipse, Jones vector, and Stokes vector, are commonly used in optical design software to specify the polarization elements and describe the polarization state of the beam as it propagates through the system.

6.2 Polarized-Light Interactions with Tissues

Biological tissue is an optically inhomogeneous and absorbing medium. The property of the light propagating within the tissue depends on the scattering and absorption properties of the tissue, which are further determined by particle size, shape, and density.

When light illuminates the tissue normally, approximately 4–7% of the incident light at all wavelengths reflects at the tissue surface because of the large change in the refractive index between the air and the tissue. This component accounts for the “glare” from the tissue surface and is the major cue by which surface texture is judged. Most of the light reflected at the tissue surface maintains its polarization state. The other 93–96% of the incident light enters the tissue, where it is absorbed and scattered. Scattering causes a change in the direction of light propagation and is a physical process by which the light can be detected through the tissue surface. The backscattered light carries few or no visual cues related to the surface texture, yet it carries all of the visual cues related to the internal tissue structure.

Since the randomness of tissue structure results in fast depolarization of light propagating inside the tissues, polarization effects are usually ignored. However, many fibrous tissues, such as the cornea, sclera, tendon, cartilage, as well as some structured tissues, such as the retina, tooth enamel, and dentin, show a wide variety of polarization properties, namely, linear birefringence, optical activity, and diattenuation. The refractive index of the tissue is higher along the length of the fibers than across the width. Those polarization-related properties are primarily defined by the tissue structure or by the intrinsic anisotropic character of the tissue components or metabolic molecules.

When polarized light is used in illumination, tissue birefringence and tissue scattering change its polarization state. The polarization degree of transmitted or reflected light is measurable, even after it propagates a considerable distance

inside the tissue. The depolarization length in tissues is close to the mean transport path length of a photon. This length characterizes the distance within which the direction of the light propagation and the polarization plane of linearly polarized light become totally random after a number of scattering events.¹⁶

The interaction of polarized light with the birefringent tissue is complicated because both the birefringent and the scattering effects can change the polarization state of light.

As shown in Fig. 6.4, light scattering by tissue gradually randomizes the original polarization state of the illuminating light. Unscattered or weakly scattered light preserves its original polarization state, whereas multiply scattered light is randomly polarized and contributes equally to both copolarization and cross-polarization states. Thus, the light backscattered from the tissue surface and the superficial layers beneath the surface has less opportunity to depolarize, and it is mostly polarized parallel to the incident light. The perpendicular polarization component consists predominantly of photons scattered from the tissue in a certain depth. Therefore, the parallel component is more intense. The peak intensity of the perpendicular component is delayed in time because the photons propagate into the tissue to a certain depth and then scatter back to the detection system.

Figure 6.5 shows the temporal profiles of the two polarization components of backscattered light when linearly polarized, 6.5-ps laser pulses with a wavelength of 1064 nm illuminate bovine gray matter brain tissue.¹⁷ $I(t)_{\parallel}$ is the parallel polarization component, $I(t)_{\perp}$ is the perpendicular component, and $I_{\text{norm}}(t)_{\perp}$ is the normalized perpendicular component relative to the peak of the parallel polarization component. The parallel polarization component is more intense

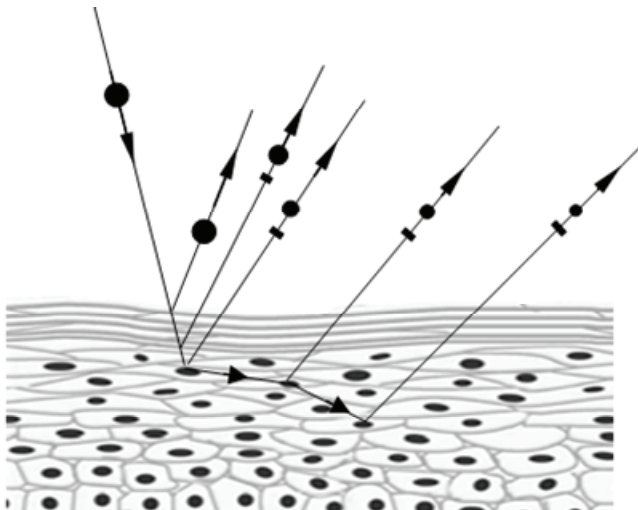


Figure 6.4 Polarized light propagates inside the tissue. As the light travels through the tissue, it is gradually depolarized. Solid circles and lines indicate the two orthogonal polarization directions.

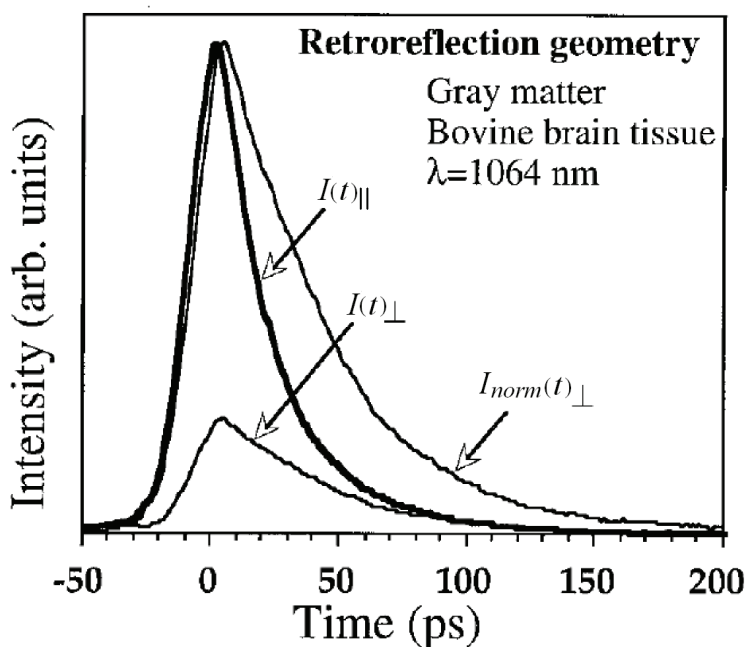


Figure 6.5 Temporal profiles of two polarization components of backscattered light pulses when 1064-nm, 6.5-ps laser pulses are used to illuminate bovine gray matter brain tissue. $I(t)_{||}$ is the parallel polarization component, $I(t)_{\perp}$ is the perpendicular component, and $I_{norm}(t)_{\perp}$ is the normalized intensity perpendicular component.¹⁷ (Reproduced with permission from the Optical Society of America.)

than the perpendicular polarization component by a factor of ~ 4.3 , and the peak of the perpendicular component appears ~ 4 ps later than that of the parallel component. The FWHM of the parallel component is smaller than that of the perpendicular component.

Data about the tissue surface and the superficial layers are contained in the parallel image $I_{||}$, and data about the tissue in depth can be obtained from the perpendicular image I_{\perp} . The parallel image $I_{||}$ also contains photons that penetrate beneath the tissue. Therefore, the polarization imaging technique can be used in biomedical imaging to improve the image quality of the tissue and extract the structural information of the tissue surface and the tissue beneath the surface, as well as the birefringence of its components.

Besides linearly polarized light, circularly polarized light has also been used in biomedical imaging.^{18,19} For highly scattering tissues, linearly polarized light is rapidly depolarized because each scattering event rotates the photon's reference plane. For circularly polarized light, changes in the azimuthal angle do not affect its helicity. Therefore, circular polarization shows its potential in deeper tissue imaging with better contrast.²⁰⁻²² Another advantage of using circularly polarized light for tissue imaging is that circularly polarized light improves image contrast, as proven by a number of researchers.²³⁻²⁵

6.3 Polarization Imaging Systems in Biomedical Imaging

This section will discuss five selected polarization-sensitive techniques that have been applied to the imaging and functional diagnosis of biological tissue.

6.3.1 Conventional polarization imaging

For biomedical imaging systems using reflectance light, part of the light is reflected at an opposite angle to the incident light in the local plane of incidence. This phenomenon is called specular reflection. The light specularly reflected from the tissue surface is generally much stronger than the light backscattered from the tissue, and it may cause local saturation of the sensor.

When polarized light is used in the illumination path and an analyzer with a transmission axis perpendicular to the polarization axis of the incident light is employed in the imaging path, the specularly reflected light will be blocked by the analyzer because it has the same polarization as the illumination light. Thus, only the cross-polarized component of the scattered light contributes to the image formation. The captured image is a perpendicular image I_{\perp} , which does not have a specular reflection component. This technique is an effective tool for functional diagnostics and for the imaging of subcutaneous tissue layers. It has been also applied extensively in biomedical imaging to eliminate specular reflection.

Figure 6.6(a) illustrates the basic configuration of a polarization imaging system built to capture a polarized image. A linear polarizer is placed in the illumination path to ensure a linear polarization illumination, and a second polarizer with an orthogonal transmission axis, used as an analyzer, is positioned between the sensor and the tissue. To avoid a depolarization effect from the optical elements in the illumination and imaging paths, both the polarizer and analyzer should be the elements closest to the tissue. When the angular FOV is large, the optimal positions for the polarizer and the analyzer are where the ray angle is small. Figure 6.6(b) is an on-axis version of Fig. 6.6(a), where a polarization beamsplitter is used to provide polarized light illumination and block the copolarization light. To improve the degree of polarization in the illumination and imaging paths, an additional polarizer can be used in the illumination path to prepolarize the light, and an additional analyzer can be used in the imaging path to clean the leaking copolarized light.

The light specularly reflected from the tissue surface can also be reduced by illuminating the tissue at an angle, as shown in Fig. 6.6(a), and it can be further reduced by using an index-matching liquid.²⁶ This polarization imaging technique is simple and low cost; it can be a powerful tool for functional diagnostics and for imaging diseased fragments of tissues, as well as for detecting tumor borders that are not visible to the naked eye.²⁷ The disadvantage of this technique is that some surface features may be lost because a large portion of the reflected light carrying the surface information is blocked by the analyzer.

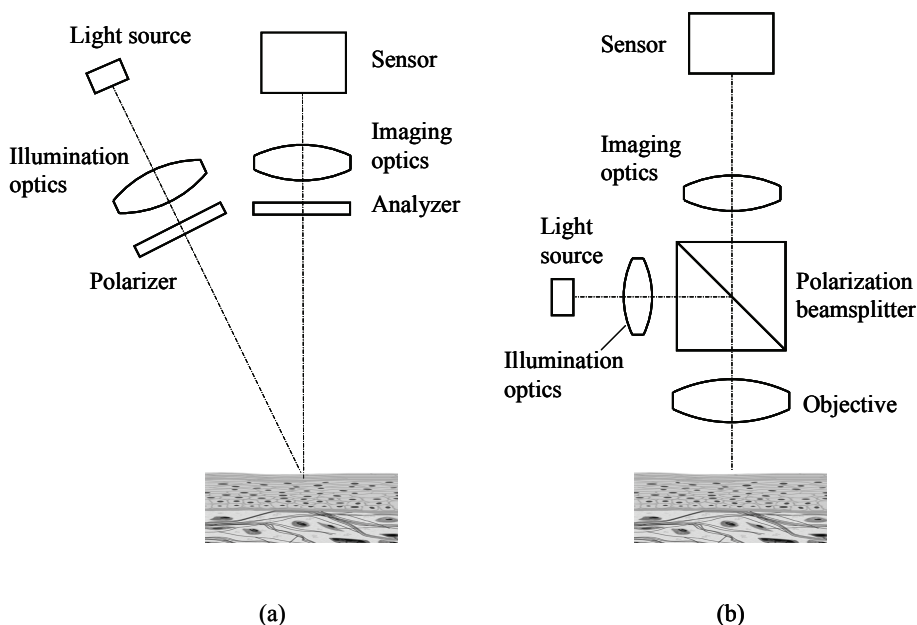


Figure 6.6 Polarization imaging systems to capture a cross-polarization image: (a) off-axis system and (b) on-axis system.

6.3.2 Orthogonal polarized spectral imaging

Orthogonal polarized spectral (OPS) imaging, developed by Cytometrics, Inc. in the early 1990s, is an application that uses the configuration in Fig. 6.6(b).²⁸ It is a relatively new noninvasive method to assess human microcirculation. The polarization state of the light is preserved in specular reflection as well as in a single scattering event. After several scattering events, the light is depolarized. This depolarized scattered light effectively back-illuminates any material in the foreground, making the back-illuminated material visible. The illumination spectrum is selected to highlight tissue structures. For example, when a 548-nm wavelength within the hemoglobin absorption spectrum is chosen, the blood vessels of peripheral microcirculation can be visualized using OPS imaging as in transillumination intravital microscopy.

One recent development of OPS imaging is dark-field orthogonal polarized spectral (DFOPS) imaging.^{29,30} Instead of sharing an optical path for illumination and imaging as in the conventional OPS system, the DFOPS imaging method illuminates the tissue outside the FOV of the imaging system. In conventional OPS, given the birefringent nature of the tissue, even the light from specular reflections and light with a single scattering event undergo some degree of polarization rotation from their original orientation of the source polarization. The analyzer cannot fully block these components, leading to a diminished image contrast. DFOPS can avoid specular reflection and eliminate the light with single

scattering events dramatically, thus enhancing image contrast and improving the OPS image quality.

6.3.3 Polarization ratio imaging

The polarization imaging techniques discussed so far only use backscattered light with orthogonal polarization; the reflected light with the same polarization as the illumination light is abandoned because of strong specular reflection. The reflected or backscattered light in the parallel polarization state contains three parts of the light from the tissue: light from specular reflection, backscattered light with a single scattering event, and backscattered light with multiple scattering events. The light from specular reflection is generally too strong to extract valuable information of the tissue surface, but the light backscattered with a single scattering event and backscattered light with multiple scattering events contain useful structural information of the tissue and tissue surface. Therefore, by capturing only the light with orthogonal polarization to eliminate the specular reflection, the useful light, which reflects from the tissue surface and the superficial layer of the tissue, is also reduced.

After the images I_{\perp} and I_{\parallel} are captured, a new image is calculated pixel by pixel by subtracting I_{\perp} from I_{\parallel} . As a result, the large amount of randomly polarized light due to the multiply scattered light from deep tissue is subtracted out in the new image $(I_{\parallel} - I_{\perp})$.³¹⁻³³ The image removes the depolarized component and retains the surface structure information. To cancel any spatial variation in the illumination light and surface geometry, the subtracted image $I_{\parallel} - I_{\perp}$ is divided by $I_{\parallel} + I_{\perp}$:

$$p = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}. \quad (6.17)$$

This image is sometimes called the polarization ratio image, which is insensitive to variations in illumination light intensity and in surface pigmentation, but it is sensitive to the superficially scattered polarized illumination light.

Images with parallel and perpendicular polarizations can be obtained by rotating the analyzer for a one-sensor system or by using a polarization beamsplitter to direct the light to two sensors, one for each polarization state, as shown in Fig. 6.7. The advantage of the second approach is that two images can be captured simultaneously, minimizing the error from the tissue's instability. Thus, the image can be subtracted directly with no need to reregister the images. In Figs. 6.6(a) and 6.7, the imaging path is outside the optical path of the specular reflection so that less light from specular reflection will enter the imaging path. Another option to minimize the image degradation from specular reflection is to place a window on the tissue surface with the index-matching liquid between them.

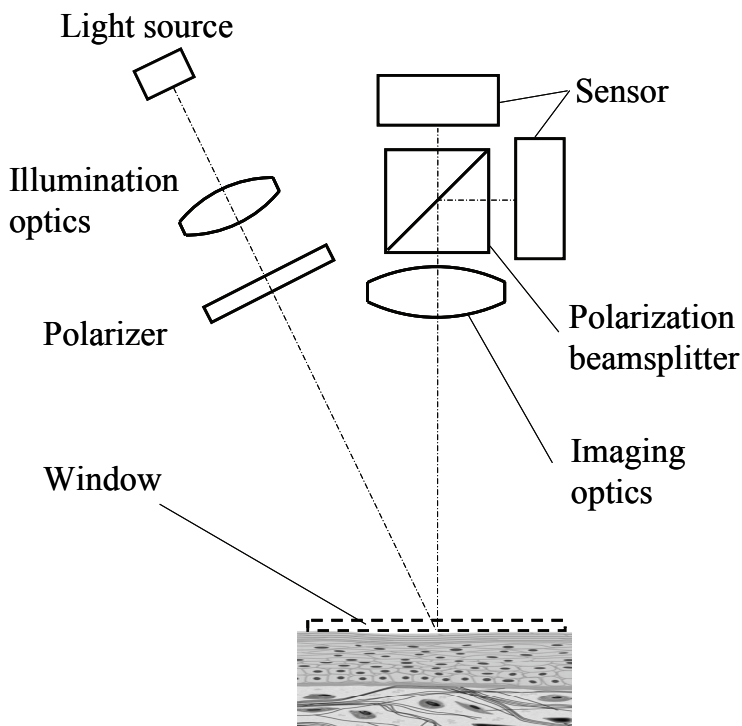


Figure 6.7 Polarization imaging system to capture images with orthogonal polarization states simultaneously.

6.3.4 Spectral polarization difference imaging

The spectral selection of scattered radiation introduces a novel quality to the polarization-sensitive methods and significantly improves their diagnostic potential. OPS is one of the polarization imaging methods that use the spectral dependence of light/tissue interaction.

In conventional polarization imaging, the photons backscattered from the tissue surface and the superficial layer of the tissue dominate the image, while the photons that reach the tissue located below the surface are hidden in a strong “background.” Therefore, the object below the surface is difficult to identify. Demos, Alfano, and Radousky proposed the spectral polarization difference imaging (SPDI) technique to provide subsurface imaging at larger depths, and they demonstrated deep-tissue imaging (1.5 cm beneath the surface) by using both spectral and polarization discrimination of the backscattered photons.³⁴

As discussed in Chapter 1, the penetration depth of light in tissue is determined by absorption and reduced scattering coefficients. In the diagnostic and therapeutic window (600–1300 nm), the penetration depth is larger for the longer wavelengths because the scattering power reduces as the wavelength increases.

The SPDI technique uses different illumination wavelengths to probe different penetration depths inside the tissue. A difference in illumination wavelengths allows for imaging structures at different depth zones. The large difference in illumination wavelengths gives rise to wider depth zones.

Polarization filtering is essential for obtaining images of tissues at different depths. Without polarization filtering, there is no other technique that can simultaneously remove the polarized specular reflection component and the depolarized component from the outer tissue layers.

The cross-polarization images of the tissues are recorded with different spectra. The SPDI image is obtained by subtracting a recorded image using a shorter illumination wavelength from a recorded image using a longer illumination wavelength. Since a large amount of the image information not relevant to the target is removed, the contrast of the image obtained using the SPDI imaging technique is improved. This process enhances the relative intensity of the image of an object located below the tissue surface.

Photons carrying target information undergo multiple scattering inside the tissue; therefore, image resolution is reduced significantly. The maximum imaging depth depends on the scattering and absorption characteristics of the tissue at the illumination wavelengths.

Figure 6.8(a) shows a typical SPDI system, which includes a broadband light source, a spectrum filter, illumination optics with a polarizer, an imaging system with an analyzer, and an image sensor. Basically, the setup is a conventional polarization imaging system with a filter to select the illumination spectrum. The filter can be a tunable filter, such as a liquid crystal tunable filter (LCTF) or an acousto-optical tunable filter (AOTF), or it can be an array of filters mounted on a filter wheel. The light from the light source passes through the illumination optics, spectrum filter, and polarizer and illuminates the object. As the light penetrates the tissue, it interacts with various tissue structures. Some of the illumination light emerges from the tissue in the backscattering direction. The intensity of the backscattered light depends on the optical characteristics of the tissue surface as well as the tissue below its surface at a particular location. Light reflects from the surface specularly, as well as from the superficial layer with few scattering events, and is copolarized and removed with the analyzer. The light that backscatters from somewhere below the tissue surface is depolarized, and consequently, the cross-polarized component can pass through the analyzer to the sensor.

The top four curves in Fig. 6.8(b) are the digitized intensity profiles across a line that contains the target object for the four images obtained under 600, 690, 770, and 970 nm. The bottom curve is the intensity profile of the SPDI image that is obtained by subtracting the images at 770 and 970 nm. In the experiment, the depth of the target object was chosen such that it would be “visible” to light with a longer wavelength.³⁴

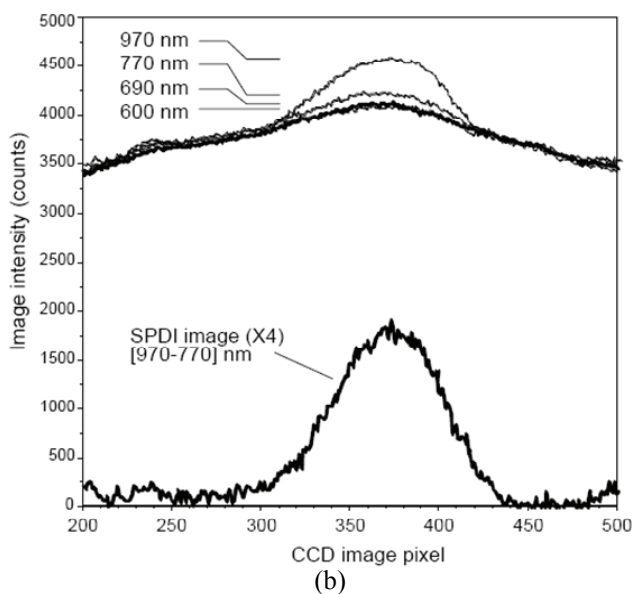
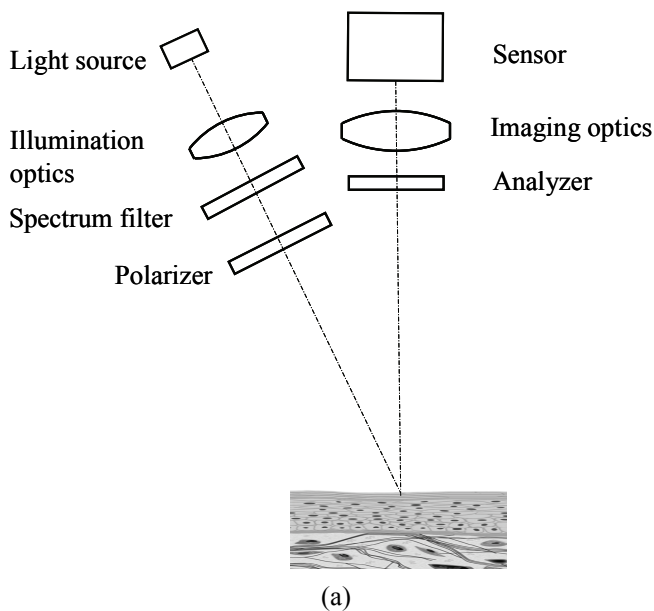


Figure 6.8 (a) Diagram of a spectral polarization difference imaging (SPDI) system. The illumination path includes a light source, illumination optics, filter, and polarizer. The imaging path has an analyzer, imaging optics, and a sensor. (b) The top four curves are the digitized intensity profiles across a line that contains the target object for the four images obtained under 600, 690, 770, and 970 nm. The bottom curve is the digitized intensity profile of the [970–770 nm] SPDI image multiplied by 4.³⁴ Reproduced with permission from the Optical Society of America.

In recent years, with the advances in LEDs and laser diodes, there is a wider selection of wavelengths and powers. Both LEDs and laser diodes can be used to replace the broadband light source and filters.

6.3.5 Polarized light spectroscopy

One potential approach to detecting and diagnosing cancer in its early stage is the spectral analysis of the scattered light perturbed by a tissue structure. Polarized reflectance spectroscopy can be used to detect the scattering properties of the epithelial layer in the presence of both a strong diffusely scattered background and hemoglobin absorption. The nuclear size distribution, which is the key quantitative morphologic characteristic used in the detection of precancerous lesions, can be obtained by best fitting the experimental data.³⁵

When polarized illumination is used, reflected light with polarization parallel to that of the illumination light contains the majority of the light scattered by the superficial layers of the tissue (such as the epithelial cells) and a portion of the diffuse background signal from the structure inside the tissue (such as the stroma), which is modulated by hemoglobin absorption. Conversely, reflected light with polarization perpendicular to that of the incident light consists of only the diffuse background scattering and a small portion of the light scattered on the superficial layers of the tissue. Therefore, the scattering, size-dependent characteristics of the epithelial layer can be obtained from the depolarization ratio spectrum $D_p(\lambda)$,

$$D_p(\lambda) = \frac{I_{\parallel}(\lambda) - I_{\perp}(\lambda)}{I_{\parallel}(\lambda) + I_{\perp}(\lambda)}, \quad (6.18)$$

where $I_{\parallel}(\lambda)$ is the component of the light scattered by the sample with the polarization parallel to the incident light, $I_{\perp}(\lambda)$ is the component with the polarization perpendicular to the incident light, and λ is the wavelength of the incident light. The normalization term in the denominator is the total intensity of light. Normalization accounts for the spectral characteristics of the illumination lamp and the spectrometer. The measured depolarization ratio spectra are then fit by varying the scattering size and size distribution with a fixed refractive index through a theoretical model developed from the Mie theory to extract nuclear sizes and refractive indices.

Figure 6.9 presents two configurations for polarized reflectance spectroscopy. In Fig. 6.9(a), the illumination light from the light source is directed to the tissue surface through the illumination path, which consists of a polarizer, filter, and illumination optics. The filter is used to select the wavelength range for illumination. The light backscattered from the tissue is collected by the collection optics and analyzed using the spectrometer. The collection optics includes a collection lens and an analyzer. The advantages of this configuration include the following:

- The illumination and detection regions, as well as the illumination and detection angles, can be easily adjusted.
- The illumination and detection polarizers can be rotated to study different polarization effects.
- The requirement of the angular performance of the polarizers can be looser because the polarizers can be placed in a region where the light has a small divergent or convergent angle.

The limitation with this configuration is that the light collection efficiency is limited by the collection optics, and the system is relatively bulky and complex. Another limitation is the possible specular reflection from the tissue surface. The specular reflection can be minimized by placing a flat window on the tissue surface with an index-matching liquid between them.

Figure 6.9(b) shows a compact probe, which consists of a central fiber delivering polarized illumination to the tissue and two fibers for simultaneous collection of scattering components with polarizations parallel and perpendicular to the incident light.³⁶ The surfaces of the illumination and detection fibers can be angle polished or tilted to selectively probe scattering at various depths within the epithelium. This modification also allows for collection of the light backscattered with polarization parallel and perpendicular relative to the illumination from the same region within the tissue. One advantage of this configuration is that there is no additional optical component after the polarizer in the illumination path and before the analyzer in the detection path; therefore, depolarization by the other optical component can be avoided. Another advantage is that it is compact and can be used in an endoscope to take images inside the body. One issue, though, is that detected signals with different polarization states may not originate from the exact same point. Other limitations include the inability to adjust the relative positions of the illumination and detection regions or the angles of the detection paths.

There are some other imaging modalities, such as polarization-sensitive OCT and the Mueller matrix polarization imaging system, that use polarized light in illumination and in detecting light with different polarization states. Those systems will not be discussed in this chapter; for more information, please refer to Refs. 37–41.

6.4 Polarization Elements

A polarization element is a component that modifies the polarization state of a light beam. Polarizers, retarders, rotators, and depolarizers are some commonly used polarization optical elements. Polarizers are essential in polarization imaging systems to precondition the illumination light and filter the light from the object. Retarders are used in some biomedical imaging systems, such as the Mueller matrix polarization imaging systems, to change the polarization state. Polarization rotators and depolarizers are not commonly used in biomedical polarization imaging systems. This section will discuss the types and properties

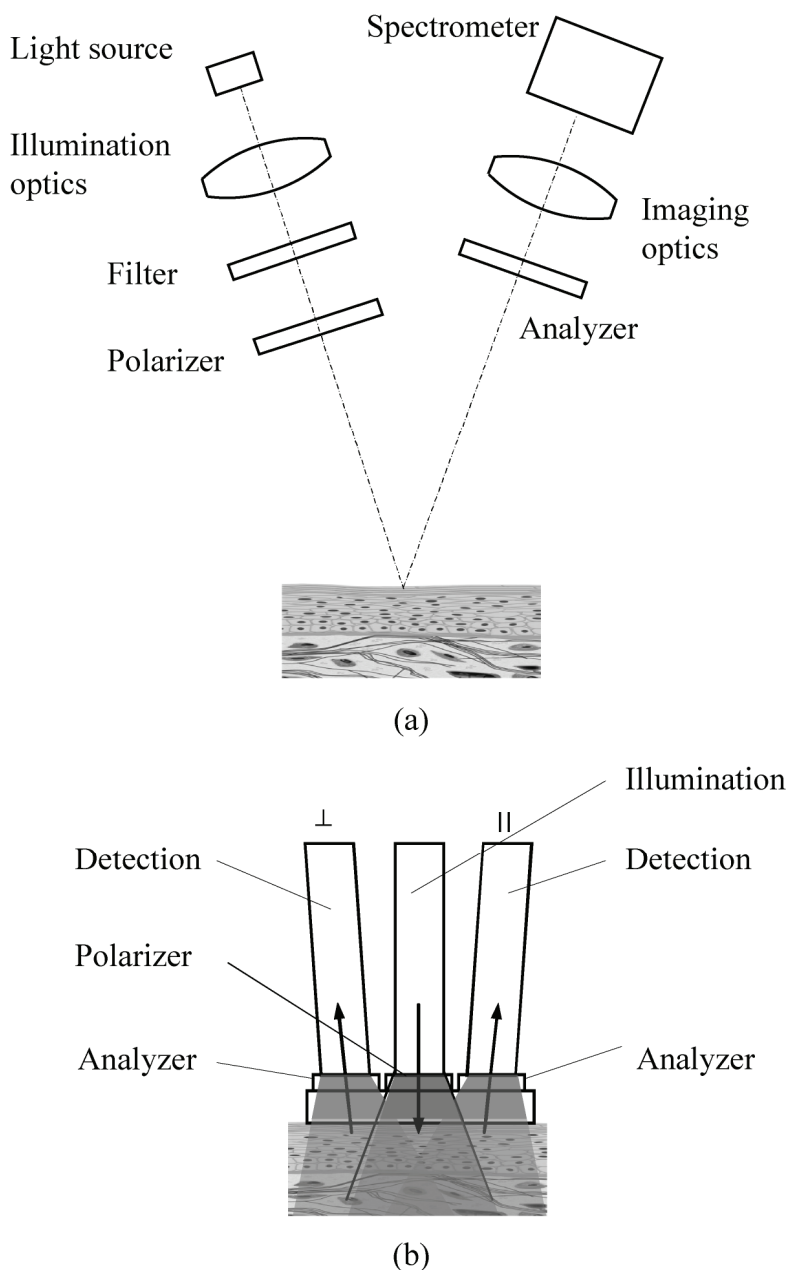


Figure 6.9 Configurations of polarized light spectroscopy: (a) imaging based and (b) nonimaging based.

of polarizers and retarders to provide a basic understanding of their principles, strengths, and weaknesses and, therefore, how to choose among them.

6.4.1 Polarizers

A polarizer is an optical element used to control the polarization state of light. It usually converts unpolarized or mixed-polarization light into light with a single polarization state, or it transforms light between polarization states. A polarizer creates a fixed polarization state independent of the polarization state of the incident light. For a linear polarizer, the plane of polarization can be changed by rotating the polarizer about its axis.

Basically, polarizers can be classified into two general categories: absorptive polarizers and beamsplitting polarizers. Absorptive polarizers absorb light with unwanted polarization state and transmit the polarized light along its transmission axis. Beamsplitting polarizers split the incident beam into two beams that have orthogonal polarization states.

Polarizers are generally specified by the following parameters: transmission, contrast ratio, acceptance angle, wavefront quality, beam deviation, laser damage threshold, temperature range, surface quality, and geometric specifications. Transmission is the percentage of light that passes through the polarizer, and the contrast ratio is the ratio between the transmission or the reflection of the desired polarization and the undesired orthogonal polarization. The acceptance angle is the range of the incidence angle within which the polarizer has acceptable performance. Wavefront quality is the measure of the distortion created when a plane wave passes through or reflects from the polarizer. Beam deviation is an angular measure of the change in light direction when passing through or reflecting from the polarizer. The laser damage threshold is the amount of laser power that the polarizer can handle. Surface quality determines the number of defects that will be allowed on the surface. The geometrical specification includes thickness, outside diameter, and the clear aperture of the polarizer.

Not all of these properties will be discussed for the polarizers described in this chapter. More information about them can be found in Refs. 42–45.

6.4.1.1 Absorptive polarizers

The sheet polarizer is the most common, inexpensive absorptive polarizer. It consists of a matrix of oriented dye molecules or crystals on a polymer substrate. The original material was invented and patented by Edwin H. Land in 1929. The original J-sheet form consists of submicroscopic needles of herapathite oriented parallel to one another in a sheet of cellulose acetate. This type of polarizer, being microcrystalline, has a tendency to scatter light. Its later H-sheet form is made from polyvinyl alcohol (PVA) plastic with iodine doping. During manufacture, the sheet is stretched to ensure that the PVA chains are aligned in one particular direction. Electrons from the iodine dopant are able to travel along the chains so that the sheet absorbs the polarized light parallel to the chains and transmits the light polarized perpendicularly to the chains. This type of polarizer works throughout the visible spectrum but deteriorates in IR light, and it is totally ineffective in UV light because PVA absorbs strongly in the 200- to 400-nm region.

A PVA sheet is generally bonded between two films of triacetyl cellulose (TAC) for support because the PVA sheet is only approximately 35 μm thick and may decompose in a moist environment. Sheet polarizers are typically coated with a peel-off protective layer for safe and easy handling.

Sheet polarizers are also known as dichroic polarizers because they absorb light differently along orthogonal axes, and the unequal absorptions are wavelength dependent. A sheet polarizer has 45–90% transmission for polarized light and a contrast ratio from 100:1 to 1000:1. Its acceptance angle is relatively large, up to 20 deg. The transmitted wavefront is poor, on the order of a few waves, when used alone. A sheet polarizer can be laminated between glass plates for a higher-transmission wavefront. For visible polarizers, this construction produces a total transmitted wavefront distortion of less than $\lambda/5$ per inch. Figure 6.10 shows the typical performance of a sheet polarizer laminated between two glass plates. The thin solid curve plots the transmission of a single polarizer, the thick solid curve illustrates a pair of parallel polarizers, and the thick dashed curve indicates a pair of crossed polarizers. The linear polarizer has about 38% transmittance for unpolarized light, and two crossed sheets have an average transmission of 0.04%.

The advantages of a sheet polarizer include a large angular acceptance and extreme flexibility for custom shapes and sizes. However, the sheet polarizer can

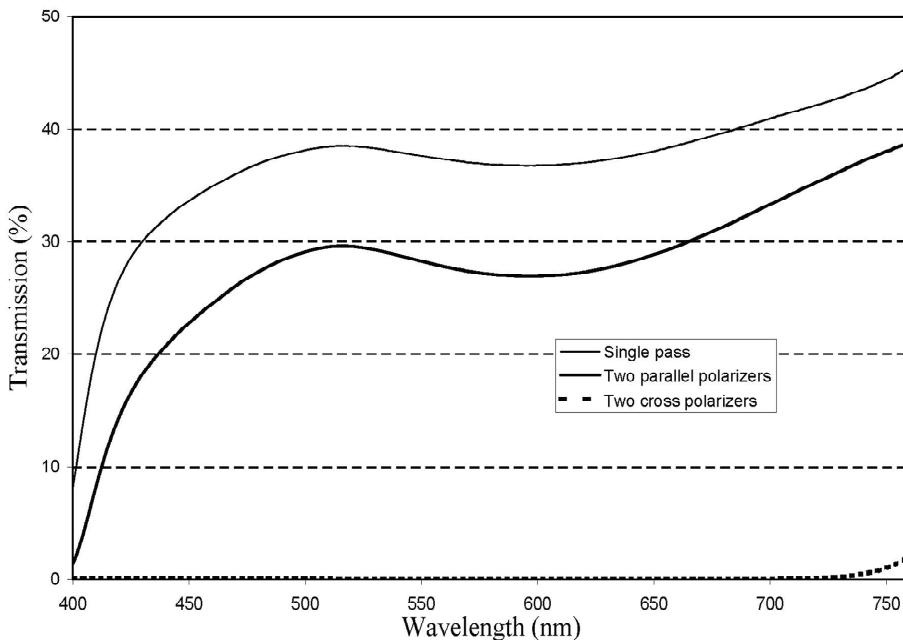


Figure 6.10 Performance of a typical sheet polarizer. This diagram was plotted from original data kindly provided by American Polarizers, Inc.

only be used in low-irradiance applications. The typical power limit is 1 W/cm^2 . Other limitations include relatively poor wavefront quality, low contrast, low working temperature range, and low transmission in short wavelengths. Sheet polarizers may be damaged by prolonged exposure to strong UV radiation.

Another type of absorptive polarizer is made of elongated silver nanoparticles embedded near the surface of a glass. The elongated silver particles are aligned along a common axis. Because of prolate silver particles and the preferential absorption along the particles, the transmitted light becomes linearly polarized. The level of absorption depends on the wavelength.

Because this type of absorptive polarizer is a solid-glass product, it is more durable and can polarize light much better than a sheet polarizer. It also has a lower absorption of correctly polarized light. Other advantages include high contrast, high transmission, more resistance to UV radiation and chemicals, more resistance to temperatures (up to $+400^\circ\text{C}$), and a large acceptance angle, as high as 20° .

Currently, there are two commercial products based on this technology: Polarcor™ by Corning and colorPol® by Codixx. Figure 6.11 shows the transmittance and contrast of a visible colorPol® polarizer. The contrast is better than 1000 from 460–675 nm, with a peak around 490 nm.

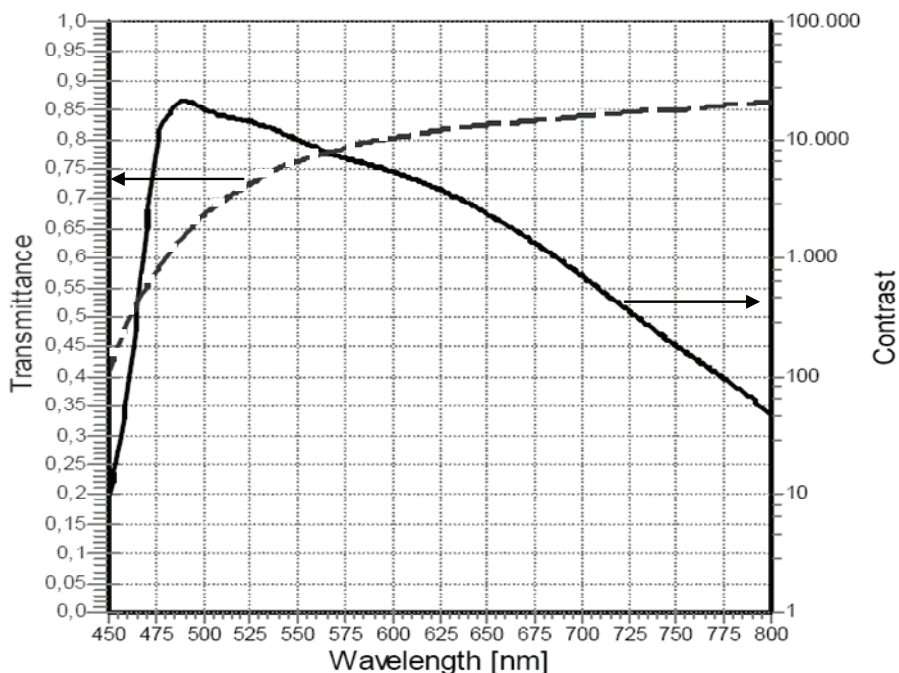


Figure 6.11 Transmittance and contrast of a visible colorPol® polarizer. Courtesy of Codixx AG, Germany.

6.4.1.2 Beamsplitting polarizers

Beamsplitting polarizers split the incident light into two beams with orthogonal polarization states by one of the following mechanisms: reflection, birefringence, thin-film coating, or wire-grid coating. Unlike absorptive polarizers, beamsplitting polarizers do not absorb light; therefore, they do not need to dissipate the energy of a rejected polarization state. Beamsplitting polarizers can be used for high-intensity beams, such as laser beams.

An uncoated flat plate is the simplest polarizing beamsplitter. When a single beam of unpolarized light strikes a flat optical plate at an angle, the reflected beam will be partially or completely linearly polarized. The DOP depends on the angle of incidence and the refractive index of the flat plate. The angle at which the DOP is 100% is defined as Brewster's angle. The DOP can be increased by using a stack of uncoated flat plates.

The thin-film beamsplitting polarizer is based on the interference effect. A thin-film stack of alternating high- and low-index materials of appropriate thickness is coated onto one surface of a plate polarizer or onto the hypotenuse of a right-angle prism that is cemented to a second right-angle prism to form a cube. Figure 6.12(a) shows a plate polarizer with a multilayer coating on the incident side of the plate; the p component of the unpolarized incident light transmits, while the s component reflects. Historically, the plate polarizer is used only at or beyond Brewster's angle. Recent advances in coating design and fabrication, however, have made 45-deg plate polarizer designs more manufacturable. With ion-beam-sputtering coating technology, many hundreds of thin-film coating layers can be deposited on the plate with high precision. The plate polarizer can be designed as a bandpass filter for p -polarized light and a broadband blocking filter for s -polarized light.

Figure 6.12(b) shows a conventional MacNeille polarization beamsplitter cube with a thin-film coating. Like the plate polarizer, the p component of the unpolarized incident light transmits, while the s component reflects. To minimize losses, all entrance and exit surfaces of the polarization cube are coated with an AR coating. As a benefit of its construction, the transmitted or reflected light of the polarization beamsplitter cube has a low wavefront distortion.

The thin-film polarization beamsplitter cube can tolerate relatively high light levels of 500 W/cm^2 . The limiting factor is that the adhesive between the prisms breaks down before the thin-film coating. Absorption, scattering, and outgassing, which can occur within the cement layer, will result in laser damage and loss of energy.

The thin-film polarization beamsplitter cube offers the unique advantage of providing two linearly polarized output beams with negligible absorption: one transmitting straight through (p -polarized light) and the other reflecting off at precisely 90 deg (s -polarized light). Figure 6.13(a) shows the performance of a visible broadband polarization beamsplitter cube at normal incidence.⁴⁶ The performance, such as the contrast ratio and transmission efficiency, and the wavelength range of the thin-film beamsplitting polarizer depend on the design of the thin-film stack. The contrast ratio of the transmitted beam on the

polarization cube beamsplitter is more than 500:1, much better than that of the reflected beam, which is usually less than 100:1.

One major limitation of thin-film polarizers is that their performance has a strong dependence on wavelength and angle of incidence since they operate on the principle of multiwave interference. As shown in Figs. 6.13(a)–6.13(c), the transmission of the p -polarized light and the reflection of the s -polarized light are highly sensitive to the angle of incidence and wavelength. The transmission of

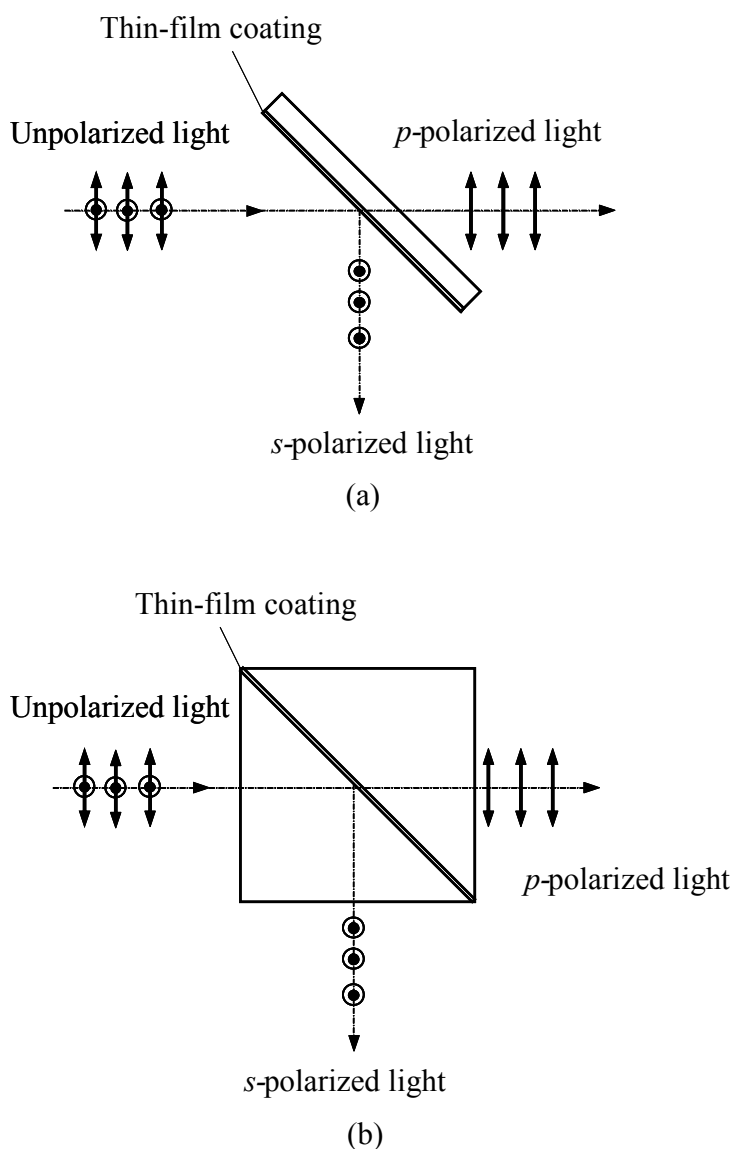


Figure 6.12 Thin-film beamsplitting polarizer: (a) plate polarization beamsplitter and (b) polarization beamsplitter cube.

the p -polarized light drops quickly as the angle of incidence increases, and there is substantial contamination of the p -polarized light in the reflected s -polarized beam.

Another limitation of a conventional polarization beamsplitter cube is that

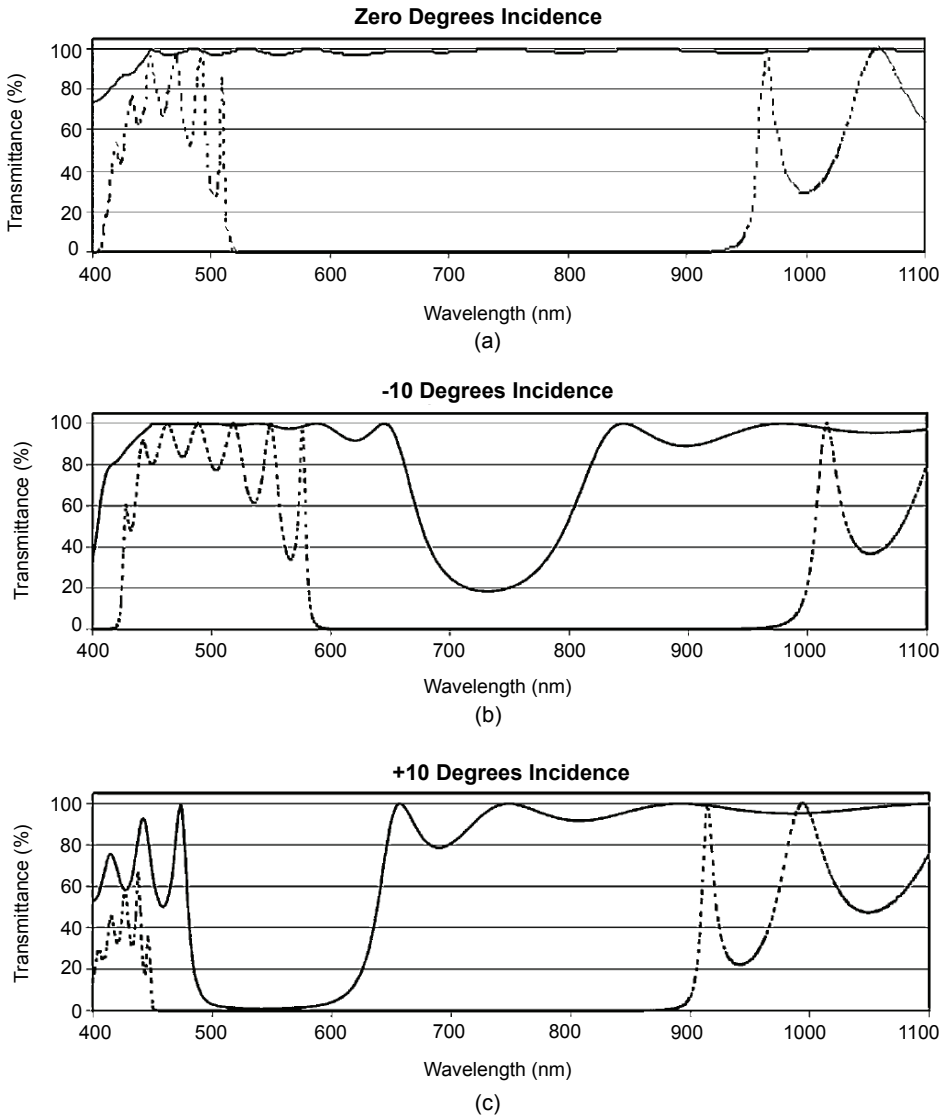


Figure 6.13 Performance of a thin-film polarizing beamsplitter cube: (a) normal incidence, (b) -10 -deg incidence, and (c) $+10$ -deg incidence. The solid line is the transmission of the p -polarized light, and the dashed curve is the s -polarized light.⁴⁶

the polarization axes for skew rays are rotated geometrically with respect to the system coordinate axes.^{47,48} Skew rays experience a relative phase shift, which causes a small degree of ellipticity in the output beams. These effects limit the angular range over which the conventional polarization beamsplitter cube can be used. To address these problems, two special cube-type polarization beamsplitters have been developed. One is based on a plastic film made of hundreds of layers of highly birefringent polymers on the cube hypotenuse,⁴⁹ and the other is based on a very fine wire-grid structure on the cube hypotenuse.⁵⁰

6.4.1.3 Birefringent polarizers

Birefringent polarizers exploit the birefringent properties of crystals such as calcite and quartz. When an unpolarized ray is incident upon the surface of the crystal, it is split by refraction into two linearly polarized rays with orthogonal polarization states: an ordinary ray (o ray) and an extraordinary ray (e ray).

There are several types of birefringent polarizers, such as the Nicol prism, Glan-Thompson prism, Glan-Foucault prism, Glan-Taylor prism, and Wollaston prism. Figure 6.14(a) is a diagram of the well-known Glan-Thompson prism, which consists of two right-angled calcite prisms that are cemented together. The optical axes of the calcite crystals are parallel and aligned perpendicular to the plane of reflection. When unpolarized light enters the prism, the *p*-polarized o ray is totally internally reflected from the calcite-cement interface, leaving the *s*-polarized e ray to be transmitted. The reflected beam is not close to 90 deg to the transmitted beam.

The Glan-Thompson polarizer has an excellent contrast ratio of greater than 100,000. Its wavelength range can be as large as 320–2500 nm. However, its angle of incidence is limited, and it is expensive. The aperture of the Glan-Thompson polarizer is also smaller than 30 mm in diameter because of the limited availability of large optical-quality calcite. Also, the transmission is not consistent across the entire wavelength range and decreases significantly in the UV. In addition, the cemented construction limits both power handling and UV performance. The Glan-Foucault polarizer, as shown in Fig. 6.14(b), has an air gap between two prisms and can be used in much higher irradiance without damage. The Glan-Foucault is also shorter, and its reflected beam is close to perpendicular with the transmitted beam. The limitation is that the reflected beam is not fully polarized.

6.4.1.4 Wire-grid polarizers

As shown in Fig. 6.15, a wire-grid polarizer consists of an array of fine parallel metallic wires deposited on one surface or imbedded into a glass substrate. The wire-grid polarizer has been known in theory and used in certain applications in IR for many years but was not applied to visible optics until the late 1990s with the technical advances in grid-fabrication techniques.

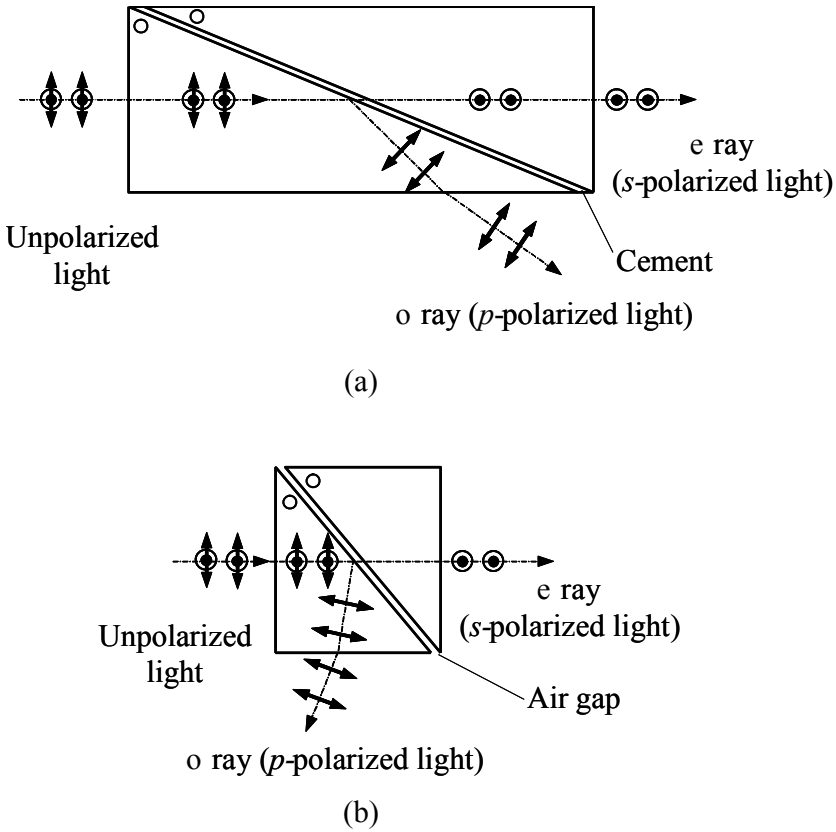


Figure 6.14 Birefringent polarizers: (a) Glan-Thompson prism and (b) Glan-Foucault prism.

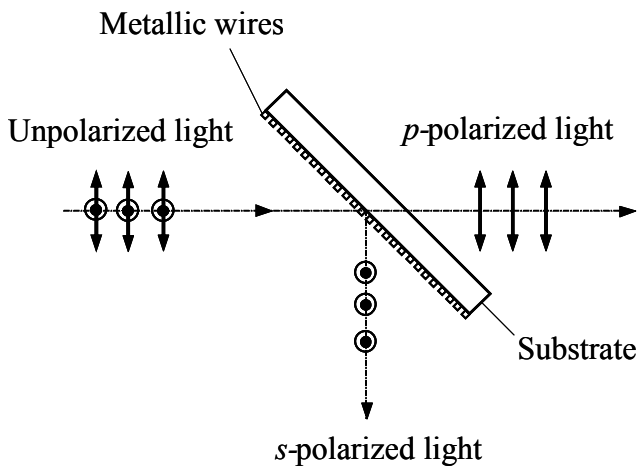


Figure 6.15 Schematic diagram of a wire-grid polarizer at 45 deg. This polarizer can be designed for normal or 45-deg incidence. (The wire grid is greatly exaggerated for illustration purposes.)

The electric field of the incident electromagnetic radiation interacts with the wires, causing electrons to oscillate along the wires but not across the wire grid. Since electrons re-radiate energy as they oscillate, they absorb energy only from the E -field parallel to the wires and thus create a reflected wavefront for only one polarization. Because the E -field perpendicular to the wires cannot interact with the electrons, this polarization is transmitted undisturbed through the polarizer. The transmitted wave has an electric field purely perpendicular to the wires and is, therefore, linearly polarized.⁵¹ The direction of polarization is defined by the wire direction, not by the plane of incidence. Thus, the skew rays have the same polarization direction as those in the plane of incidence. In Fig. 6.15, the transmitted beam is p -polarized, and the reflected beam is s -polarized. The wire-grid polarization beamsplitter can also be used to transmit s -polarization, but the contrast is lower. Therefore, the wire-grid polarization beamsplitter is usually used to transmit p -polarization.

Wire-grid polarizers have a high transmission and contrast ratio. The overall performance is determined by the period, shape, height, and width of the wire grid. With advanced microlithographic technology, wire-grid polarizers can be tuned for the UV, visible, and near-IR regions. They can be designed either for normal incidence or for 45-deg incidence. At normal incidence, the contrast ratio is better than 500 in the visible spectrum. The performance of a typical wire-grid polarization beamsplitter is shown in Fig. 6.16. The performance changes very little with the angle of incidence, resulting in uniform system performance over a large aperture. It can be used at acceptance angles of 20 deg without depolarization of the skew rays. As shown in Figs. 6.16(a)–6.16(d), the contrast ratio of the reflected s -polarization is lower than that of the transmitted p -polarization component.

The transmitted wavefront error of a wire-grid polarizer is small and suitable for most imaging applications. However, astigmatism and coma introduced by the tilted polarization beamsplitter are a concern for the transmitted beam in some optical system configurations.

Because it is made from metal wire and glass, a wire-grid polarizer can withstand temperatures as high as 200°C and power densities greater than 50 kW/cm². One limitation of the wire-grid polarizer is its surface quality. Its scratch-dig is around 80–50, which is relatively poor because of the manufacturing method.

6.4.1.5 Circular polarizers

A circular polarizer converts unpolarized light or linearly polarized light into left or right circularly polarized light. There are two types of circular polarizers. The commonly used one combines a linear polarizer with a quarter-wave retarder. The transmission axis of the linear polarizer is at a 45- or 135-deg angle to the fast axis of the quarter-wave retarder. A circular polarizer can be used for broadband spectrum applications only when the retarder is achromatic. The other type of circular polarizer utilizes the Bragg reflection of a cholesteric liquid crystal (CLC) film.⁵²

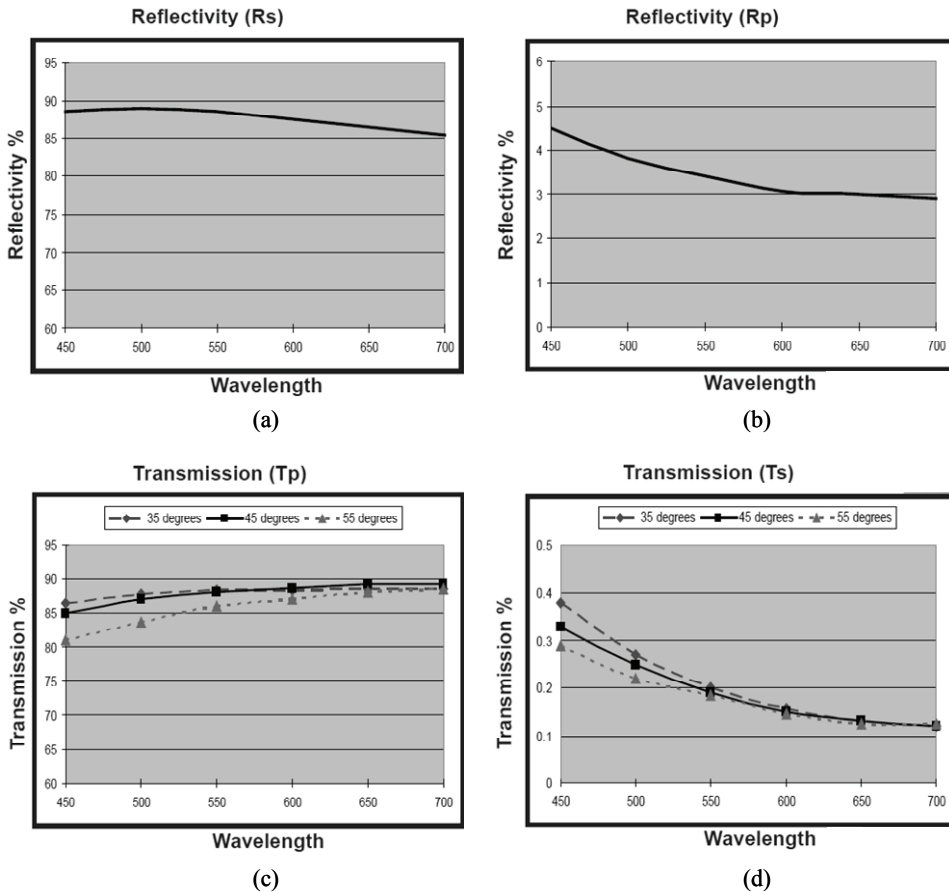


Figure 6.16 Performance of a wire-grid polarization beamsplitter. (a) Reflectivity of s -polarization light at 45 deg, (b) reflectivity of p -polarization light at 45 deg, (c) transmission of p -polarization light at different angles of incidence, and (d) transmission of s -polarization light at different angles of incidence. Courtesy of MOXTEK Inc.

6.4.1.6 Special polarizers

Conventional polarizers typically allow the light of a single polarization state to pass through, and the orientation of the transmitted polarization state is the same regardless of the location of incidence on the polarizer. However, in some applications, such as polarimeters, it is desirable to have a polarizer that passes light with different polarization orientations at different locations.

Patterned polarizers have been available only recently, with the advancement of manufacturing technologies. They have different transmitted polarization states and different spectral characteristics at different regions of the polarizer. Figure 6.17(a) shows the subregion of a patterned polarizer from MOXTEK Inc. Each pixel can have a different polarization orientation, for example, 0, 45, 90,

and 135 deg. These micropatterned polarizer arrays can be aligned with focal plane arrays through direct contact or relay lenses. The shape of the patterns can be square pixels, circles, or any other arbitrary shape or pattern. These shapes and patterns can be as small as a few microns or as large as several millimeters. For visible wavelengths, the pixel size can be as small as 8 μm .

Figure 6.17(b) shows another type of special polarizer that consists of a linear polarizer and a radial polarization converter. The polarization converter converts a linearly polarized beam into a beam that has a continuous radial or azimuthal polarization distribution. The polarization converter is a liquid crystal element whose entrance and exit plates are linearly and circularly rubbed, respectively. Azimuthally polarized light is achieved for incident light parallel to the axis (rubbing direction) of the polarization converter, and radially polarized light is achieved for incident light perpendicular to the axis.⁵³

6.4.2 Retarders

A retarder, also called a wave plate, is a polarization element that produces a specific phase difference between two orthogonal components of incident polarized light. It resolves a beam of polarized light into two orthogonal components, retards the phase of one component relative to the other, and then recombines them into a single beam with new polarization characteristics. Ideally, the retarder does not polarize unpolarized light, and it does not change the DOP of a light beam, but it usually changes the polarization state of incident polarized

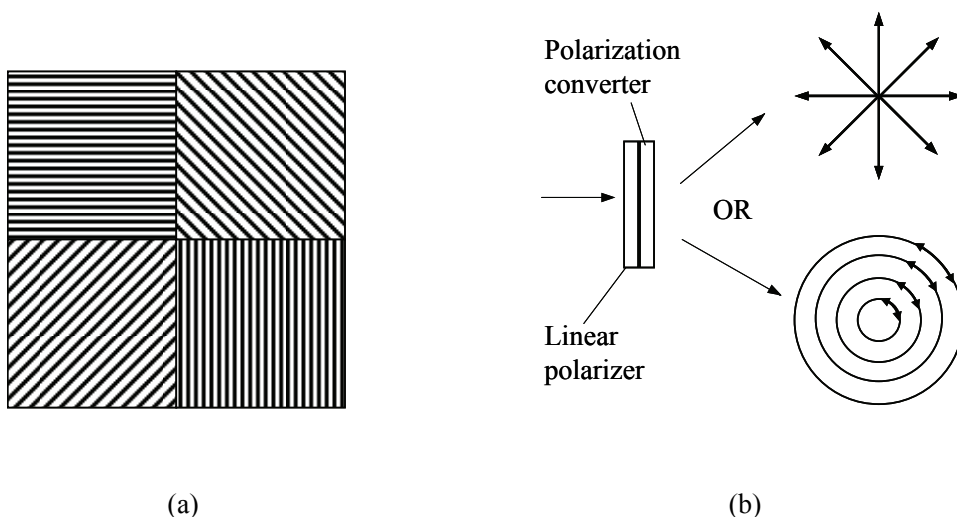


Figure 6.17 Special polarizers: (a) patterned polarizer and (b) polarization converter to convert the linearly polarized light to radially polarized light or azimuthally polarized light.

light. Retarders can be developed for any pair of orthogonal polarization states: linear, elliptical, and circular.

Most retarders are made from birefringent, uniaxial materials, such as crystals and birefringent polymers. Birefringent, uniaxial materials have two different refractive indices: the extraordinary index n_e and the ordinary index n_o . The difference between the two indices defines the material birefringence. For a positive uniaxial material ($n_e > n_o$), light polarized parallel to the ordinary axis travels at a higher speed and has a smaller optical path length than light parallel to the extraordinary axis. Therefore, the ordinary axis is referred to as the fast axis, while the extraordinary axis is referred to as the slow axis.

Retardance of a birefringent material is given by

$$\delta = \frac{2\pi}{\lambda} d(n_e - n_o), \quad (6.19)$$

where λ is wavelength of the incident light and d is the thickness of the birefringent element. This equation indicates that retardance is strongly dependent on both incident wavelength and retarder thickness.

A retarder alters the polarization of light in a manner that depends on the retardance and the angle between the retarder's fast axis and the input plane of polarization. It transmits two polarized lights without changing the polarization state except for the optical path lengths. These are the eigenpolarizations that define the axis of the retarder; all other incident polarization states are modified by the retarder.

6.4.2.1 Birefringent retarders

The most common retarders are the quarter-wave plate and half-wave plate. A quarter-wave plate has 90-deg retardance, introducing a quarter-wavelength delay between two orthogonal linear polarization states. It changes the linearly polarized light at ± 45 deg relative to the fast axis to be right or left circularly polarized. The two linear eigenpolarizations have unchanged polarization states except for the quarter-wave phase difference. All other linear states emerge as elliptically polarized. A half-wave plate has 180-deg retardance, introducing a half-wavelength delay between two orthogonal linear polarization states. It rotates linearly polarized light at an angle of θ to the fast axis by 2θ . Half-wave retarders convert right circularly polarized light into left circularly polarized light and vice versa. Table 6.5 lists output polarization states of horizontal fast-axis retarders for input light with different polarization states.

A full-wave retarder, which has one wavelength retardance, is valuable for eliminating unwanted polarization changes in an optical system. Some optical components, such as a metal-coated mirror, change the polarization state by introducing unwanted phase shifts. For example, linearly polarized light becomes elliptically polarized after reflecting off a metal-coated mirror. The induced

ellipticity can be accurately corrected by using a full-wave retarder and tilting it around either the fast or slow axis to change its retardance slightly.

When the net retardance for a retarder is only a fraction of a wavelength, it is known as a zero-order retarder. A multiple-order retarder is a retarder whose net retardance is an integral number m of the wavelength plus the desired fractional retardance $(1/n)\lambda$: $(m + 1/n)\lambda$. Multiple-order retarders may be less expensive than zero-order retarders, but they are sensitive to temperature and incidence angle.

When only one material is used in a zero-order retarder, it is called a true zero-order retarder. The zero-order retarder created by combining two multiple-order retarders with certain retardance differences is called a compound zero-order retarder. In a compound zero-order quartz retarder, the fast axis of

Table 6.5 Output polarization states of horizontal fast-axis retarders for input light with different polarization states.

Incident polarization state	Horizontal fast axis $\lambda/4$ retarder	Horizontal fast axis $\lambda/2$ retarder

one quartz plate is aligned with the slow axis of another, cancelling the large retardance values and resulting in the desired fractional retardance. One advantage of compound zero-order quartz retarders is that thermal stability is improved as the temperature effects of the two retarders cancel.

Two types of birefringent materials, namely, crystals (both solid and liquid) and oriented polymers, have been used in retarders. Natural crystalline materials (calcite, mica, and quartz), typically with small birefringence, have traditionally been the birefringent materials for retarders. Quartz has a birefringence of ~ 0.0092 at 589.3 nm; the thickness of a true zero-order quartz quarter-wave plate at 589.3 nm is only 16 μm . It is very difficult to fabricate and mount this thin and fragile crystal. More commonly, crystals are used in multiple-order retarders, which have a whole number of waves plus the desired fractional retardance. One limitation of a retarder made from a crystal is that a large-aperture crystal retarder is difficult to fabricate and becomes expensive in a diameter larger than 2 in.

Polymer materials have a lower birefringence than quartz and can be made into true zero-order retarders with reasonable thickness. They are usually laminated between two precision polished optical windows to ensure good transmitted wavefront quality and to minimize beam deviation. AR coatings and index-matching optical cement are typically used to maximize transmission in the visible to near-IR region. A wave plate using a material with a lower birefringence can have both low angular and thermal sensitivities of a true zero-order retarder and a thickness that makes it mechanically robust. Polymer retarders offer excellent angular FOV because they are true zero-order retarders. A polymer retarder changes by less than 1% over a ± 10 -deg incidence angle.

True zero-order retarders are preferred for most applications requiring retardance stability for wavelength, temperature, and angle of incidence. Multiple-order retarders can be extremely sensitive to incident angles, wavelength, and temperature. The retardance of a multiple-order retarder is a much faster function of the angle of incidence than that of a true zero-order retarder. A compound zero-order quarter-wave plate has the same dependence as the multiple-order version. The retardance dependence on the angle of incidence for true zero-order and multiple-order wave plates can be obtained from the manufacturers.

Birefringence, like refractive index, also varies nonlinearly with wavelength and is higher at shorter wavelengths. As Eq. (6.19) shows, the retardance of a retarder in waves varies quickly with wavelength, mostly because the retardance is inversely proportional to the wavelength. Therefore, the retardance of a retarder is only approximately the amount of true retardance because it is designed for a narrow wavelength range. For a multiple-order wave plate, the effective wavelength range for the designed retardance is even smaller than for a true zero-order wave plate.

6.4.2.2 Achromatic retarders

For many applications, an achromatic retarder, which has small variations in retardance over a large wavelength range, is required. As seen in Eq. (6.19), the retardance is inversely proportional to the wavelength and is proportional to the birefringence ($n_e - n_o$). However, no material has the birefringence proportional to the wavelength such that an achromatic retarder could be made from a single material.

The achromatic retarder can be made from two or more plates with different wavelength-dependent birefringences and different thicknesses. The ratio of the thicknesses determines the position of the peak retardance, and the thickness of one of the plates changes the overall retardance value. When two crystal materials such as quartz and MgF_2 are used, they are called a bi-crystalline achromatic retarder. The achromatic performance is the combined result of different wavelength dependences of birefringence of two crystals in a manner somewhat analogous to an achromatic doublet lens using two different materials. When three or more zero-order polymers are aligned with their optical axis at oblique angles to achieve achromatic performance, the retarder is called a Pancharatnam achromatic retarder. This type of retarder has an extremely wide spectral range and is insensitive to wavelength and temperature variations, given that they are made from zero-order retarders. The bi-crystalline achromatic retarder has a peak around the center of the achromatic wavelength range, and the Pancharatnam achromatic retarder has a relative constant retardance over the working spectrum.

The second type of achromatic retarder is based on the total internal reflection in a prism, as shown in Fig. 6.18. It is called the Fresnel rhomb. When the light is totally internally reflected inside the prism, a specific phase shift occurs between the s and p components of the light. This retardance depends on the internal angle of incidence and the refractive index of the prism. It also varies slowly with wavelength. The Fresnel rhomb has the limitations of being bulky, expensive, and limited in clear aperture and angular acceptance. The beam is displaced in a quarter-wave Fresnel rhomb [Fig. 6.18(a)] but not in a half-wave retarder [Fig. 6.18(b)], which consists of two quarter-wave prisms.

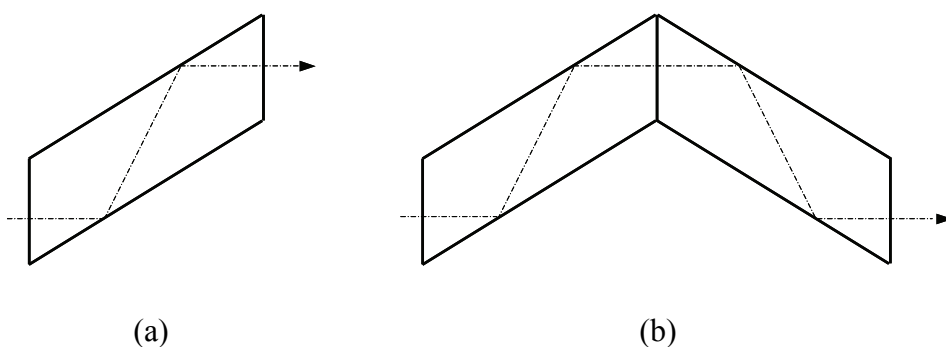


Figure 6.18 Fresnel rhomb: (a) quarter-wave and (b) half-wave.

6.4.2.3 Variable retarders

Some applications, such as highly sensitive polarimeters, require a retarder whose retardance can be varied continuously. Variable retarders can be constructed of movable elements, variable birefringent materials, Pockels cells, and piezoelastic modulators (PEMs).

One well-known variable retarder based on movable elements is the Babinet compensator, which consists of two wedges of birefringent material, such as quartz. One wedge is fixed, and the other is movable. The optical axes of both wedges are parallel to the outer faces of the wedge pair, but they are perpendicular to each other. The retardance of the Babinet compensator is determined by the birefringence and thickness difference of the two wedges at the beam location.

Liquid crystal variable retarders are made from liquid crystals whose birefringence can be changed by varying an applied voltage. The long axis of the liquid crystal molecules defines the slow index. With no voltage present, the molecules lie parallel to the windows, and maximum retardance is obtained. When a voltage is applied across the liquid crystal layer, the molecules rotate toward the direction of the applied electric field, reducing the effective birefringence and thus the retardance as well. Liquid crystal retarders can offer outstanding performance over large incidence angles. They require only low voltages at moderate drive powers. The disadvantages of a liquid crystal variable retarder include quick degradation under UV irradiation and slow modulation frequency.

A PEM is based on stress-induced birefringence. The retardance is tuned by compressing an optical glass, such as BK7. PEMs require considerable mechanical power to modulate.^{54,55} The advantage of the PEM is that it is stable in operation and shows no degradation at high intensity levels and UV irradiation. It has a large spatial and angular aperture. However, it is sinusoidally modulated with high frequency, which requires a high-speed detector.

6.4.3 Selection of polarization elements

Choosing the right polarization elements is one of the key steps in designing a polarization imaging system. Each polarization element has its own niche in polarization imaging with applications for which it is best suited. By understanding the requirements of an imaging system and the key specifications of the polarization elements (transmission, contrast ratio, transmitted wavefront quality, surface quality, beam deviation, laser damage threshold, wavelength sensitivity, angular sensitivity, working temperature, and mechanical specifications), one can locate the elements that best fit the specific application.

We also need to understand that polarization elements are not ideal because of fabrication errors (surface roughness, bulk scattering, random strains, and thin-film microstructure). A polarization element may depolarize and couple polarized light into unpolarized light. Multiple reflections inside the polarization elements may result in undesired polarization effects. For example, the multireflected

beams from a birefringent plate have different retardance. Other defects include chromatic variation, spatial nonuniformity, elliptical polarization, and nonideal diattenuation.

6.4.3.1 Polarizers

Table 6.6 summarizes the major specifications of the commonly used linear polarizers discussed above. The table is compiled from product catalogs; therefore, the data in the table may not be applicable to the same type of polarizer made by different manufacturers. The table presents the total wavelength range of all polarizers of the same type. Of all of the properties, the transmission and the contrast ratio are two key elements of quality.

The application and the key requirements of polarization and image quality, such as the wavelength range, contrast ratio, and wavefront quality, must be considered in order to choose the right polarization elements for a specific application. Generally, not all of the requirements are equally important. Certain trade-offs may have to be made to select a polarization element. When the required specifications for the desired application are identified, choosing a polarizer becomes easier because each type of polarizer has its strengths and weaknesses. Some guidelines are as follows:

Table 6.6 Summary of commonly used linear polarizers.

	Sheet polarizer	Laminated polymer	Dichroic glass	Crystal birefringent	Thin-film polarizer	Wire-grid polarizer
Wavelength range (nm)	325–1800	325–1800	325–1800	320–2300	440–1600	400–2000
Transmission	45–90%	45–90%	39–98%	85–95%	>95%	>85%
Maximum Contrast ratio	1000:1	1000:1	10,000:1	10,000:1	>500:1	>1000:1
Acceptance angle (deg)	±20	±20	±20	±5	±2	±20
Wavefront quality (/in)	Few	< $\lambda/2$	< $\lambda/2$	< λ	< $\lambda/5$	< $\lambda/4$
Beam deviation (arc min)	Large	2	5	3	3	1
Damage threshold (W/cm²)	1	1	1000	25 30	500	50,000
Temperature range (°C)	50	50	400	90	100	200
Size (mm)	Large	Small	Medium	Small	Small	Medium
Cost	Low	Medium	Medium	High	Medium	Medium

- One of the first parameters to consider is wavelength range. This can easily limit the types of polarizers to be considered. For example, for UV light with wavelengths between 270 and 400 nm, only a few polarizers, such as wire-grid and crystal polarizers, are available.
- Sheet polarizers are usually available in the visible spectrum and are rare for the near-UV and IR regions.
- Sheet polarizers or wire-grid polarizers are good candidates when a large polarizer is needed.
- Wire-grid polarizers have an extended spectral range and can be used in the UV, visible, and IR regions at the same time.
- For large angles of incidence, sheet polarizers, wire-grid polarizers, and dichroic glass polarizers have better performance.
- For very high contrast ratios, crystal optical components might be used, but they are thick, and the contrast ratio is extremely sensitive to the alignment of the component.
- For a high-transmission wavefront requirement, a beamsplitting polarization cube and a wire-grid polarizer are the optimal choices.
- In reflection mode, polarizers can only be chosen from wire-grid and beamsplitting polarizers (thin film and crystal).
- At a location with a high thermal effect, wire-grid and dichroic glass polarizers are the better choices.
- Generally, prism polarizers, such as a beamsplitting polarization cube and a crystal polarizer, require more space in the optical path.
- Thin-film polarizers are best suited for laser applications or for systems working in a narrow band of wavelengths because they function over only a range of wavelengths.
- Birefringent crystal polarizers are not well suited for imaging applications because they are limited in aperture size and can distort, scatter, and deviate the optical beam.

6.4.3.2 Retarders

When selecting a retarder, the key features under consideration include wavelength dependence, temperature sensitivity, acceptance angle, and aperture size. Retarders made from quartz cover a wide wavelength range, from 180–2700 nm. Their acceptance angle is relatively small, approximately 3 deg. The wavelength range (400–1800 nm) covered by the polymer retarders is less than that of quartz; the acceptance angle is much larger than that of the quartz retarder, about 10 deg. Given that it is based on total internal reflection, the Fresnel rhomb is achromatic by design, but its acceptance angle is small, only 2 deg.

One important consideration when choosing a retarder is the dependence of the retardance on the temperature, wavelength, and angle of incidence. The thicker the retarder, the more the retardance changes with temperature, wavelength, and angle of incidence. The retardance of a simple retarder made

from a single birefringent plate typically varies in an approximately linear fashion with wavelength and varies quadratically with the angle of incidence.

True zero-order retarders are preferred for most applications requiring retardance stability with wavelength, temperature, and angle of incidence. Multiple-order retarders can be extremely sensitive to incident angle, wavelength, and temperature. They have better light-power handling and environmental durability. Multiple-order quartz retarders are preferred for the applications using high-power lasers.

Liquid crystal retarders offer real-time, continuous control of retardance without moving parts. Babinet compensators change the optical path length while changing the phase. The PEM is not sensitive to high intensity levels and UV irradiation, but it is modulated with high frequency.

The retardance of a retarder can have some deviation as a result of fabrication errors. Retarders usually have some diattenuation because of the differences in absorption coefficients as well as different transmission and reflection coefficients at the interfaces.

6.5 Polarization Effects in Optical Systems

When light propagates through an optical system, the transmission, reflection, and optical path length are polarization dependent. In the meantime, the polarized light can be scrambled into partially polarized light. Therefore, it is necessary to understand the polarization effects in the optical system and analyze system performance.

In polarization imaging systems, polarization elements control and manipulate the polarization state. However, all optical components in the system change the polarization state to some extent because of refraction, reflection, coating, and birefringence. Polarization elements are usually not ideal; their properties can vary with angle of incidence and wavelength.⁵⁶

At nonzero angles of incidence, the s and p components of the field have different transmission and reflection characteristics in the refractive and reflective surfaces. The polarization state of a ray propagating through an optical system may change mainly because of birefringence in the lens and the coatings on the lens surfaces. Birefringence includes the intrinsic birefringence of the optical material and the stress-induced birefringence in the optical element. An AR coating on the lens surface introduces polarization-dependent transmission, and the reflection coating of the mirror introduces polarization-dependent reflection. The polarization changes may also result from interfaces of the anisotropic media, microstructures, and light propagation in anisotropic materials.

In order to assess the effect of polarization changes on the performance of an optical system, the concept of polarization aberrations (diattenuation and retardance) is introduced to characterize differences in the transmitted (or reflected) amplitudes and phases between the s and p components of the light as it propagates through the optical system.⁵⁷ Diattenuation and retardance are the properties of a ray path, not of a given polarization state.

Diattenuation is the intensity difference of two orthogonal polarization states. It causes polarization-dependent wavefronts and introduces astigmatism into on-axis beams. Retardance is the phase difference of two orthogonal polarization states. It causes polarization-dependent apodization of the pupil. Retardance occurs on coated transmissive or reflective surfaces, on TIR, and on propagation through birefringent materials. An element with pure retardance has no effect on the transmitted power and does not change the DOP of the incident light, but it can transform one polarization state into another.

6.5.1 Polarization effects of refraction/reflection

When light propagates through a lens surface, its transmission varies with the direction of polarization. At nonzero angles of incidence, the s and p components of a field have different transmission characteristics. The intensity reflection and transmission coefficients are functions of the angle of incidence and the refractive indices of the two media:⁵⁸

$$T_p(i_1) = \frac{n_2 \cos i_2}{n_1 \cos i_1} |t_p(i_1)|^2, \quad (6.20a)$$

$$T_s(i_1) = \frac{n_2 \cos i_2}{n_1 \cos i_1} |t_s(i_1)|^2, \quad (6.20b)$$

$$R_p(i_1) = |r_p(i_1)|^2, \quad (6.20c)$$

$$R_s(i_1) = |r_s(i_1)|^2. \quad (6.20d)$$

Amplitude transmission and reflection coefficients r_s , r_p , t_s , and t_p are governed by Fresnel equations

$$r_s(i_1) = \frac{-\sin(i_1 - i_2)}{\sin(i_1 + i_2)} = \frac{n_1 \cos i_1 - n_2 \cos i_2}{n_1 \cos i_1 + n_2 \cos i_2}, \quad (6.21a)$$

$$r_p(i_1) = \frac{\tan(i_1 - i_2)}{\tan(i_1 + i_2)} = \frac{n_2 \cos i_1 - n_1 \cos i_2}{n_2 \cos i_1 + n_1 \cos i_2}, \quad (6.21b)$$

$$t_s(i_1) = \frac{2 \cos i_1 \sin i_2}{\sin(i_1 + i_2)} = \frac{2n_1 \cos i_1}{n_1 \cos i_1 + n_2 \cos i_2}, \quad (6.21c)$$

$$t_p(i_1) = \frac{2 \cos i_1 \sin i_2}{\sin(i_1 + i_2) \cos(i_1 + i_2)} = \frac{2n_1 \cos i_1}{n_2 \cos i_1 + n_1 \cos i_2}, \quad (6.21d)$$

where, as shown in Fig. 6.19(a), n_1 and n_2 are the refractive indices of media 1 and media 2, respectively, and i_1 and i_2 are the incident and refractive angles, respectively.

On the basis of Snell's law, $n_1 \sin i_1 = n_2 \sin i_2$ when $n_1 > n_2$, $i_1 < i_2$, and vice versa. Therefore, r_s is either negative or positive, but it can never be 0 unless $n_1 = n_2$. However, r_p , t_s , and t_p can be 0 if the angle of incidence i_1 is at a special angle. Coefficient r_p is 0 when the angle of incidence i_1 is

$$i_B = \arctan\left(\frac{n_2}{n_1}\right), \tag{6.22}$$

where i_B is called Brewster's angle. At Brewster's angle, $i_1 + i_2 = 90$ deg.

If $n_1 > n_2$, $i_2 = 90$ deg when i_1 is

$$i_c = \arcsin\left(\frac{n_2}{n_1}\right), \tag{6.23}$$

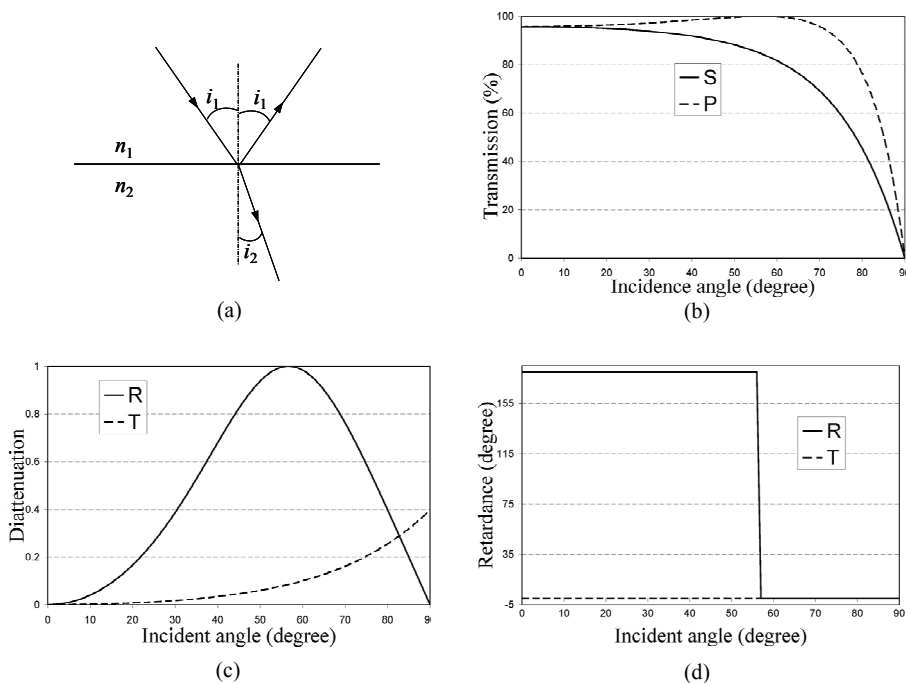


Figure 6.19 Polarization effect of refraction/reflection: (a) refraction and reflection at the interface, (b) transmission of the s and p components of incident light, (c) diattenuation of the transmitted and reflected light, and (d) retardance of the transmitted and reflected light.

where i_c is called the critical angle. When the angle of incidence is equal to and larger than this angle, TIR occurs. Both t_s and t_p are 0, which means all light—both s and p components—is reflected and there is no transmitted light.

When the light propagates through the air/BK7 interface, the intensity transmission coefficients for the s and p components are plotted against the angle of incidence in Fig. 6.19(b). At normal incidence $i_i = 0$, the s and p components are equally transmitted because the two components are not distinguishable. With increasing incident angle, the transmission of the s component drops, while the transmission of the p component increases to 1 at Brewster's angle, which is 56.595 deg at the air/BK7 interface, then drops to 0 at grazing incidence. Therefore, the reflected light is fully polarized at Brewster's angle. Another observation from Fig. 6.19(b) is that the p component always transmits with higher efficiency than the s component.

Diattenuation D , which measures the difference between the transmission or reflectance of the s and p components, is defined below and plotted in Fig. 6.19(c):

$$D = \frac{|r_s|^2 - |r_p|^2}{|r_s|^2 + |r_p|^2}. \quad (6.24)$$

The diattenuation of the transmitted light increases quadratically with the increasing incident angle, and the diattenuation of the reflected light increases faster than that of the transmitted light and reaches its peak at Brewster's angle and then drops to 0 at grazing incidence. The diattenuation also increases approximately linearly as the refractive index of the plate, as shown in Fig. 6.20. This means that the plate with the higher-refractive-index glass introduces more polarization aberration.

The retardance δ , which is the difference in phase, is defined as

$$\delta = \arg[r_s] - \arg[r_p]. \quad (6.25)$$

Figure 6.19(d) is the retardance plot as a function of the incident angle. The retardance is always 0 for the transmitted light, but the retardance for reflected light drops from 180 to 0 deg abruptly at Brewster's angle. Equation (6.21a) shows that the sign of r_s is always negative, which means that there is an abrupt change of phase of 180 deg in the process of reflection.

When $n_1 > n_2$ and the incident angle is larger than the critical angle, the light will be totally internally reflected. Because 100% of the light is reflected, there is no diattenuation. Retardance, however, is significant. For an air/BK7 interface, the retardance is plotted in Fig. 6.21. The retardance is 180 deg when the angle of incidence is smaller than Brewster's angle and is 0 between Brewster's angle and the critical angle. When the angle of incidence increases from the critical

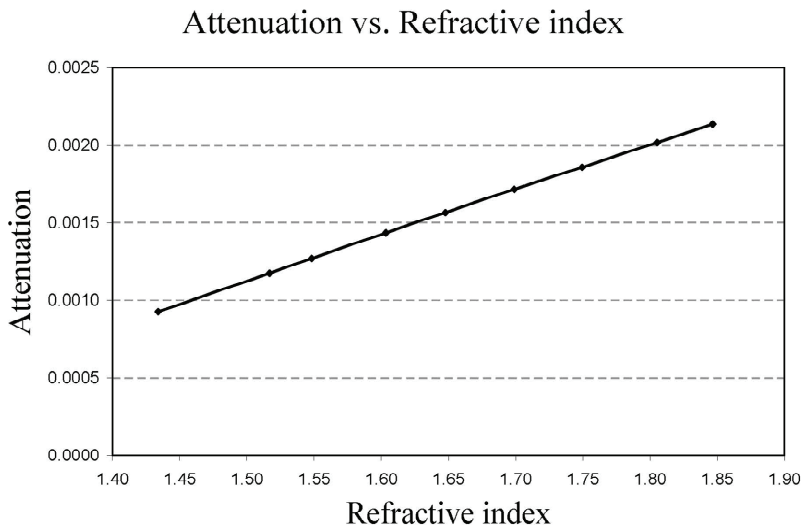


Figure 6.20 Diattenuation as a function of the refractive index at an air–glass interface.

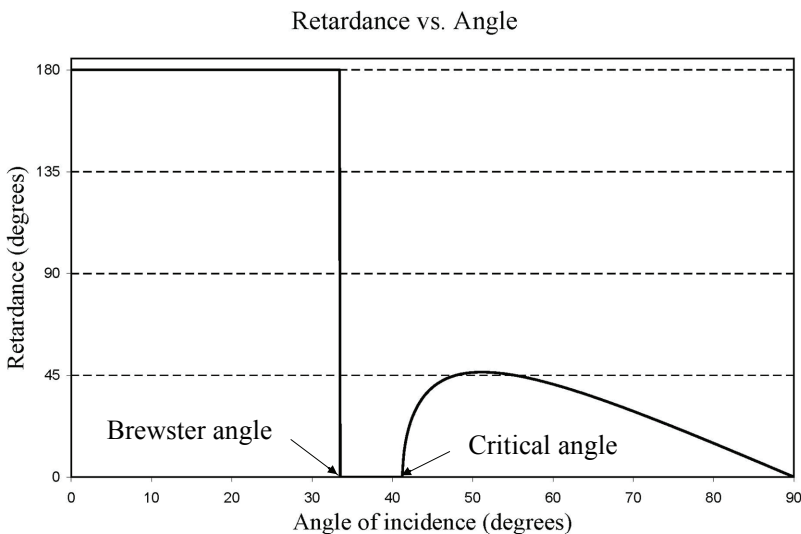


Figure 6.21 Retardance of reflected light at the air–BK7 interface.

angle, the retardance increases quickly from 0 deg to the peak of 46.38 deg and then decreases slowly to 0 deg as the grazing incidence is approached. When the angle of incidence is larger than the critical angle, the retardance δ is calculated theoretically from⁵⁸

$$\tan \frac{\delta}{2} = \frac{\cos i_1 \sqrt{\sin^2 i_1 - \left(\frac{n_2}{n_1}\right)^2}}{\sin^2 i_1}. \quad (6.26)$$

For a spherical lens surface, when the incident beam is collimated, the angle of incidence is the derivative of the surface sag:

$$i = \arctan \left(\frac{dz}{dy} \right) \approx \frac{y}{R}. \quad (6.27)$$

The angle of incidence increases approximately linearly in the radial direction, as shown in Fig. 6.22(a). For the noncollimated light in Fig. 6.22(b), the angle of incidence changes more rapidly. For the light emerging from the on-axis point, the angle of incidence is rotationally symmetrical around the optical axis. For the light from the off-axis point, the distribution of the incident angle is more complex.

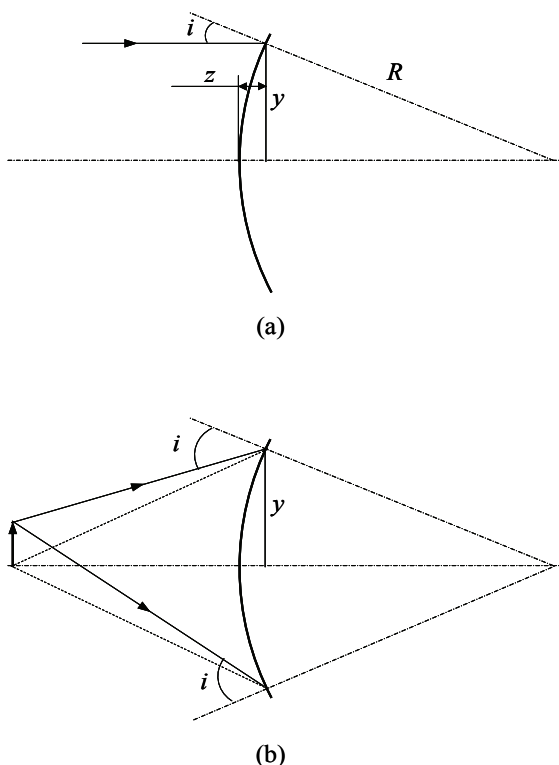


Figure 6.22 Refraction of light at a curved surface: (a) collimated beam and (b) divergent beam.

Generally, for a nonflat lens surface and noncollimated light, the angle of incidence varies across the lens surface; therefore, the lens surface acts as a weak spatially varying partial polarizer. Large variations in the angle of incidence can be an indication that polarization effects should be considered.

Figure 6.23 shows the extinction coefficient η , which is defined as the intensity ratio of the initial polarization over the orthogonal polarization of the output beam, as a function of the shape factor K of the lenses made from BK7 ($n = 1.51$) and SF57 ($n = 1.84$).⁵⁹ The shape factor K is defined as

$$K = (n-1) \frac{f}{R_1}, \quad (6.28)$$

where n is the refractive index of the lens, f is the focal length of the lens, and R_1 is the radius of its first surface.

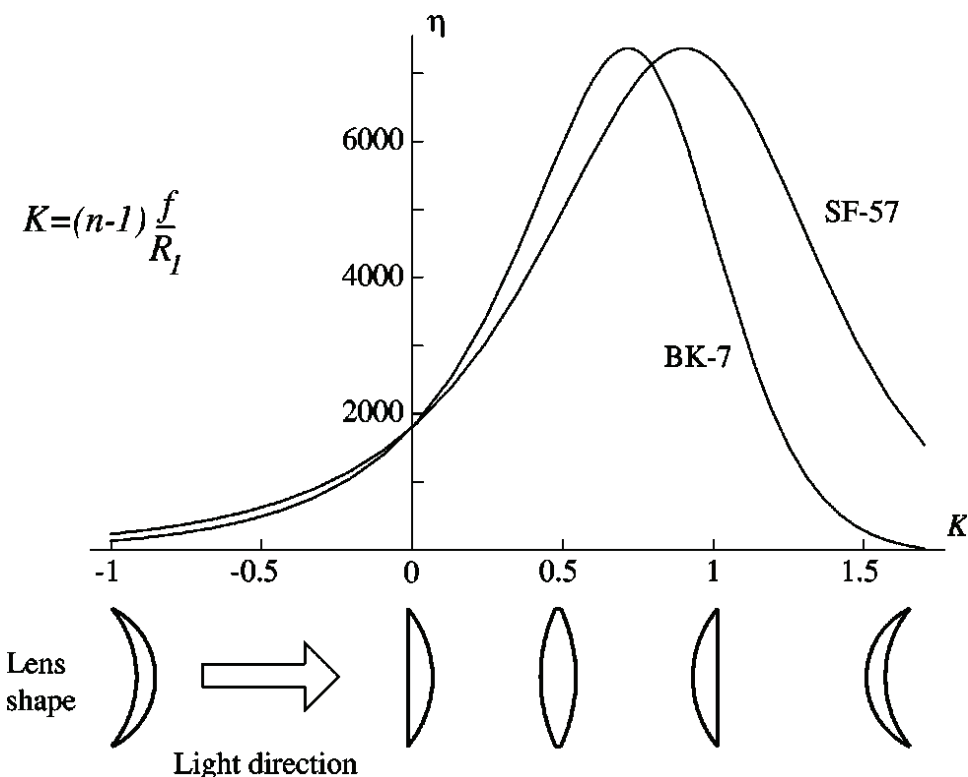


Figure 6.23 The extinction coefficient η of a ray with $NA = 0.3$ versus the shape factor K for single lenses made from BK7 ($n = 1.51$) and SF57 ($n = 1.84$). At the bottom of the figure, drawings of lenses that correspond to the shape factors are shown (light passes the lenses from left to right).⁵⁹

In Fig. 6.23, the extinction coefficient η corresponds to a circularly or linearly polarized ray with an azimuth of 45 deg. The input ray is parallel to the lens axis, and the output ray has an angle that is 17.46 deg from the optical axis. The uncoated lenses made from different optical glasses have the same maximal extinction when the lenses have about equal ray bending at both surfaces. Meniscus lenses produce maximal polarization aberrations.

Figure 6.24 shows the polarization pupil maps of uncoated low- and high-NA microscopic objective lenses. The incident light is vertically linearly polarized. Because there is no coating or birefringent material in the objective lens, all of the polarization effects are caused by refraction. As shown in Fig. 6.24, the polarization state is rotated at the corners, and the rotation increases from the center into the four quadrants. Also seen in Fig. 6.24, the rotation of the polarization state is smaller in the low-NA objective lens than in the high-NA objective lens because the incident angle in the corner is larger for the high-NA objective lens.

The polarization effect of refraction can be demonstrated by placing an uncoated singlet between two crossed polarizers, as shown in Fig. 6.25(a). The polarizer conditions the input collimated light so that only linearly polarized light passes through the lens. The analyzer passes only the light with the polarization perpendicular to the transmission axis of the polarizer, which is the depolarized light from the refraction on lens surfaces. Figure 6.25(b) shows the intensity distribution at the exit pupil of the lens. This pattern, known as the Maltese cross, shows the depolarized light caused by the uncoated singlet. It is the light farthest from the optical axis and at 45 deg with respect to the transmission axis of the polarizer that exhibits the most transmission between crossed polarizers.

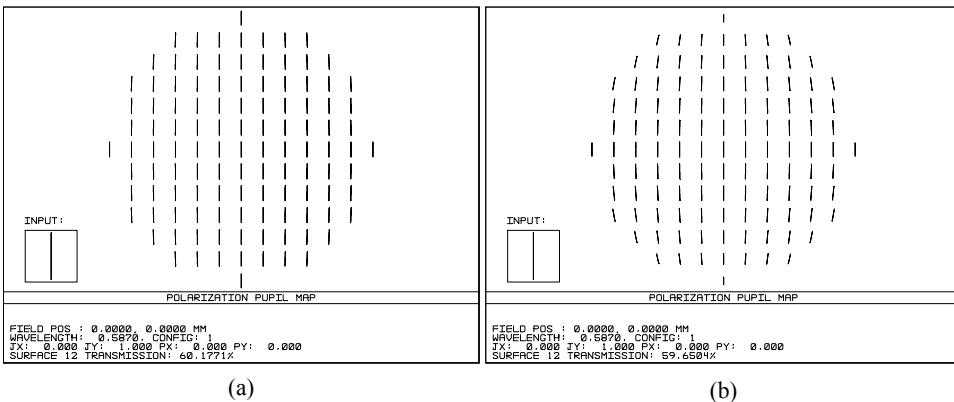


Figure 6.24 Polarization pupil maps of the (a) low- and (b) high-NA microscopic objective lens without coating.

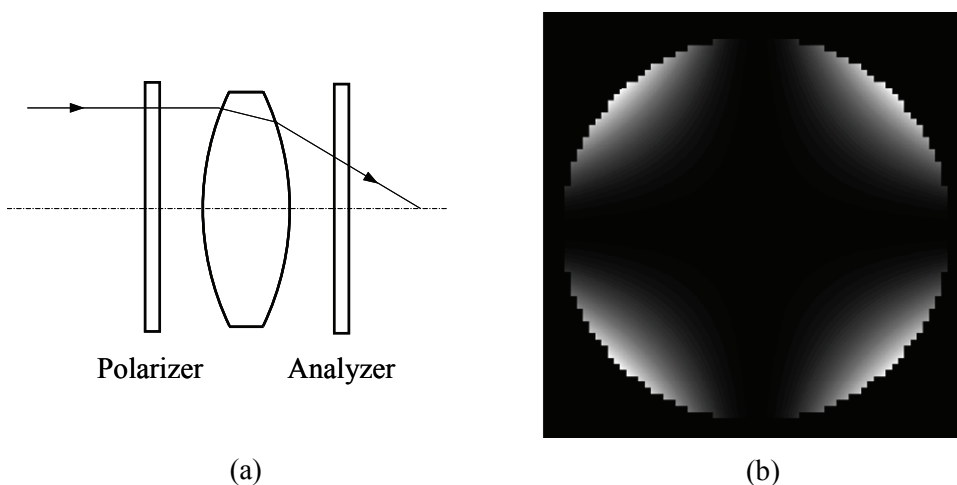


Figure 6.25 (a) Layout to demonstrate the polarization effect of an uncoated lens. An uncoated singlet is placed between the orthogonal polarizers. (b) The intensity distribution in the exit pupil of the singlet.

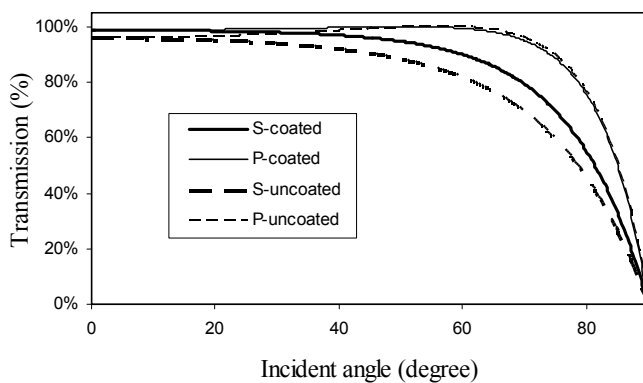
6.5.2 Polarization effects of coating

Almost all optical surfaces used in imaging systems have some coatings for various purposes: enhancement of the reflection/transmission, reduction of the reflection/transmission, separation of light with different spectra, and separation of light with different polarization states. Generally, all coatings have polarization effects to some extent. A coating on the lens surface may result in either more or less polarization effect compared to an uncoated surface. Typically, AR coatings tend to decrease the difference in transmission more than the difference in reflection at the interface.

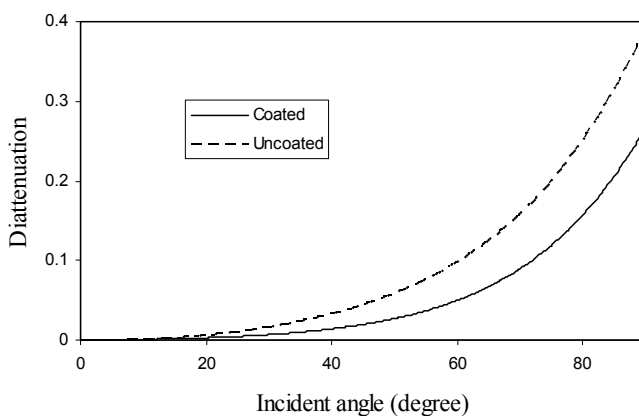
Most coatings in optical systems are designed to improve light throughput, either by reducing reflection on refractive surfaces or by enhancing the reflection on reflective surfaces, without considering phase retardance. Maximizing light throughput and minimizing phase shift may work against each other. Optimizing light throughput with minimum phase shift requires trade-offs.

A thin-film coating usually has an optical thickness of less than a few wavelengths, and its effect on the ray path can be accurately modeled given that each layer has parallel surfaces that follow the contour of the substrate. However, because the coating is so thin compared to the thickness of optical lenses, its effect on the ray path through the optical system is negligible—the coating primarily affects the amplitude and polarization of rays.

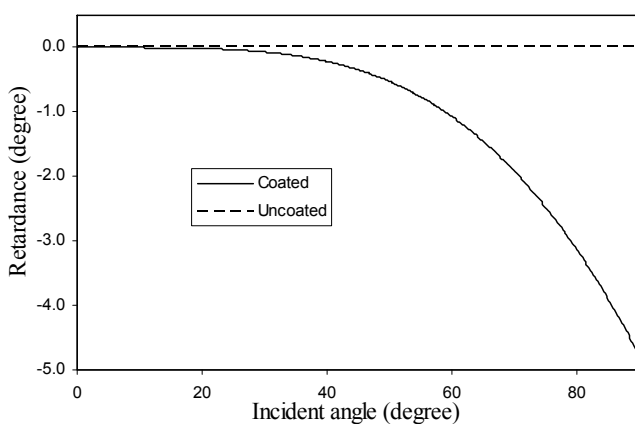
Figure 6.26 compares the transmission, diattenuation, and retardance of a flat surface with and without a single-layer MgF_2 AR coating. Figure 6.26(a) shows that the transmissions of both the s and p components are improved with an AR coating. The diattenuation in Fig. 6.26(b) also shows that the diattenuation is smaller for an AR coating. However, as shown in Fig. 6.26(c), AR coatings



(a)



(b)



(c)

Figure 6.26 Polarization effects of uncoated and coated surfaces: (a) transmission, (b) diattenuation, and (c) retardance.

introduce significant retardance, especially at large incident angles, compared to zero retardance when no coating is applied.

For systems with a high NA, the incident angle is large toward the edge of the clear aperture. Therefore, these systems are more sensitive to coating-induced phase retardance. Coating designs can minimize polarization effects because phase retardance is the sum of the phase retardance at each surface. The phase retardances of different coatings can have different signs. Thus, the total phase retardance can be minimized to a certain level by tuning the coatings on each lens surface.

Figure 6.27(a) shows the reflections of four different AR coatings with reflections less than 0.5% at normal incidence. The first three coatings, AR-1, AR-2, and AR-3, have six layers made from TiO_2 and MgF_2 . The fourth coating, AR-4, has 10 layers with the same coating materials. Figure 6.27(b) plots the retardance of each coating as a function of the incident angle up to 17 deg. It is clear that phase retardance is quite different for different AR coatings, even with the same coating material and same number of layers. Therefore, it is possible to optimize the coating for each lens surface so that the total phase retardance is minimized. As a simple example, for a flat BK7 plate in an $f/1.6$ imaging system, if AR coating AR-3 is applied to both surfaces, the total phase retardance is doubled. However, if one surface is coated with AR-3 and another with AR-4, the retardance is compensated for, and the total retardance is reduced significantly, as shown in Fig. 6.28. By using this principle, one can minimize the total phase retardance of an imaging system while optimizing the transmission.⁶⁰

Metallic coatings, such as aluminum and silver coatings, are commonly used on reflective surfaces to increase the reflectance. Figure 6.29 displays the reflectances and phases of the s and p components for an aluminum-coated reflective mirror. The reflectance of the s component continuously increases to 1 as the incident angle increases from 0–90 deg, while that of the p component decreases first and then bounces back to 1. The reflectance of the s component is always higher than that of the p component because the s component drives electrons parallel to the interface, and the p component has an electric field component in the interface which causes higher resistive losses. The average reflectance is nearly constant up to a 50-deg angle of incidence. At normal incidence, both s - and p -reflected light have a phase of 160 deg, which is determined by the refractive index of the aluminum. As the incident angle increases, the phase of the s -reflected light increases to 180 deg, while that of the p -reflected light drops to 0 deg. The phase change is 180 deg for an ideal reflection with no retardance because of sign convention.

Folding elements, such as prisms, folding mirrors, and beamsplitters, are typically deployed at a 45-deg angle of incidence to the optical axis. They may have large diattenuation and/or retardance. TIR prisms introduce only retardance, while metal-coated folding mirrors have both diattenuation and retardance. For any polarization imaging system that employs a folding mirror, there are potential challenges posed by the large incidence angles, which induce a variety

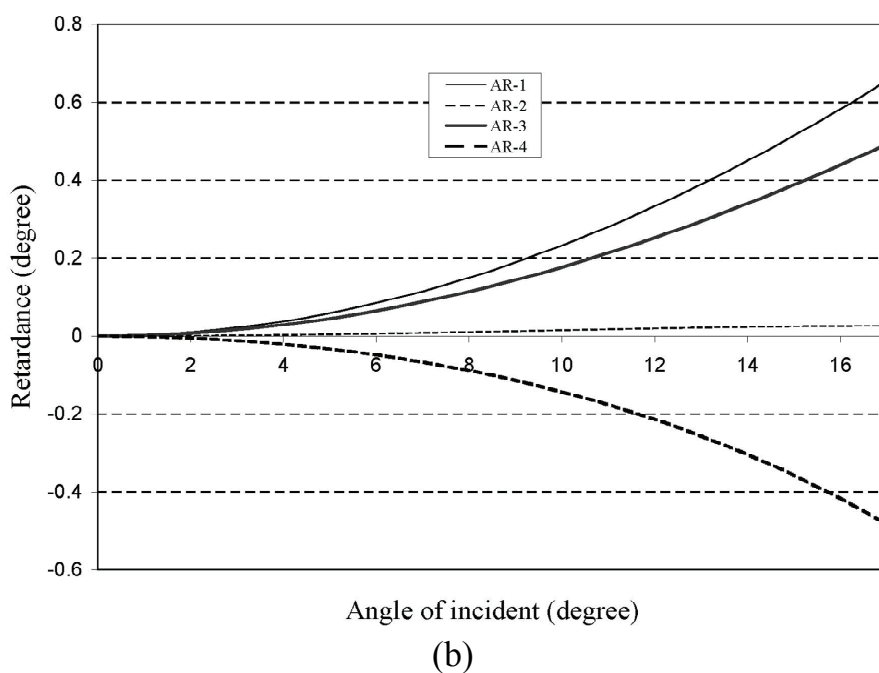
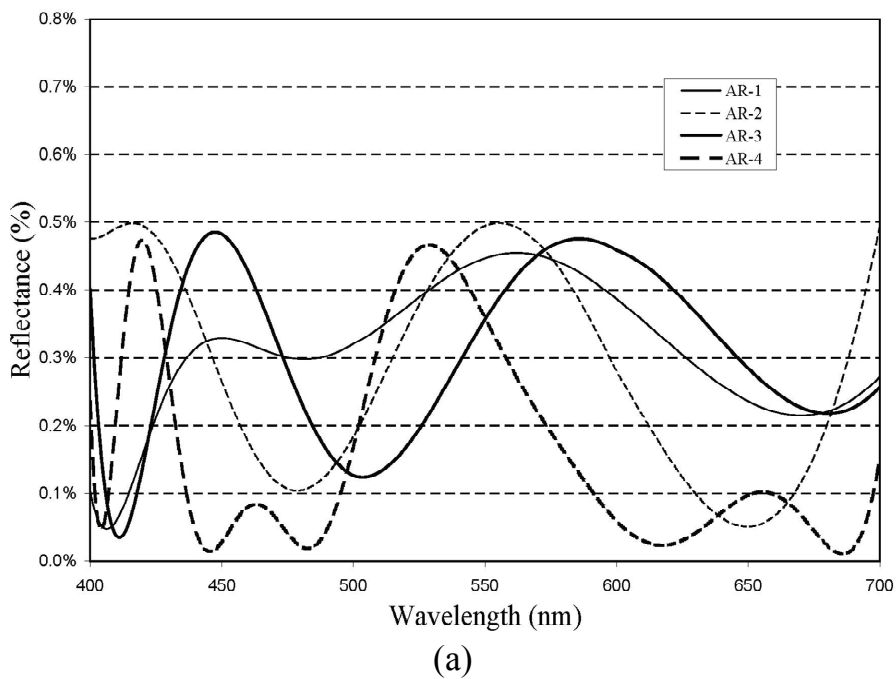


Figure 6.27 AR coatings with different phase retardances: (a) reflection and (b) retardance.

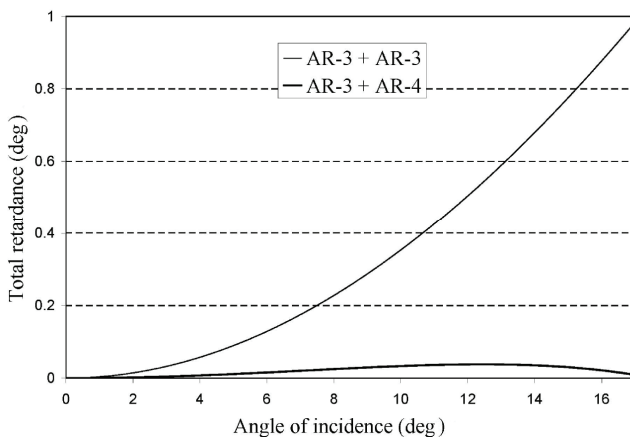
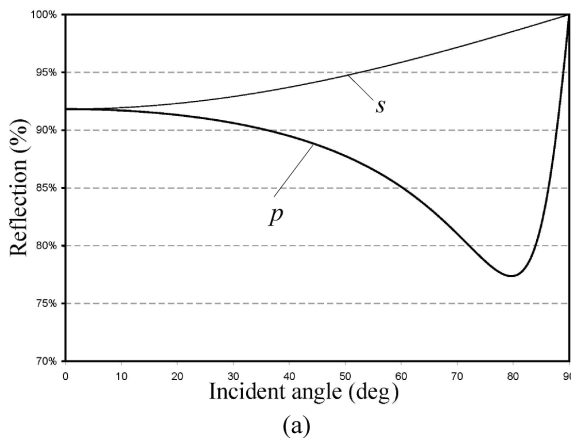
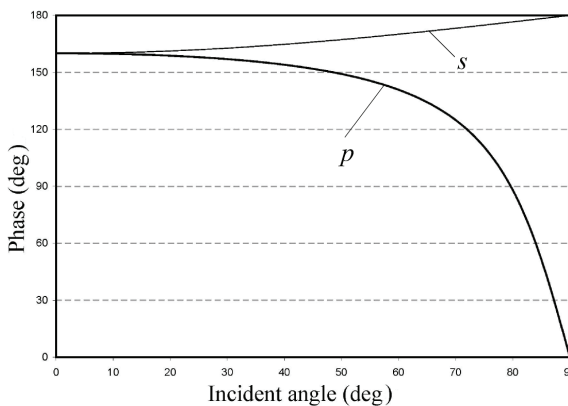


Figure 6.28 Total phase retardances of a flat BK7 plate with different AR coatings in an $f/1.6$ imaging system.



(a)



(b)

Figure 6.29 (a) Reflectance and (b) phase of the s and p components on an aluminum-coated reflective surface.

of negative effects: retardance, diattenuation, and apodization due to reflectance distribution across the pupil.

An aluminum-coated mirror is often used to fold the optical path of a beam 90 deg, as shown in Fig. 6.30(a), where the mirror is in an $f/2$ imaging system. The half convergence angle is 14.5 deg. Figs. 6.30(b)–6.30(d) show polarization pupil maps when the input light is vertically polarized, 45 deg linearly polarized, and horizontally polarized. The change of the polarization state is small when the input light is vertically or horizontally polarized. However, when the input light is 45 deg linearly polarized, the polarization state changes dramatically across the pupil. The light is almost circularly polarized at the top of the pupil and almost linearly polarized at the bottom of the pupil. Therefore, when a metal-coated mirror is used as the folding mirror in polarization imaging, the input light should be either vertically linearly polarized or horizontally linearly polarized. Any deviation will cause significant changes in the polarization state.

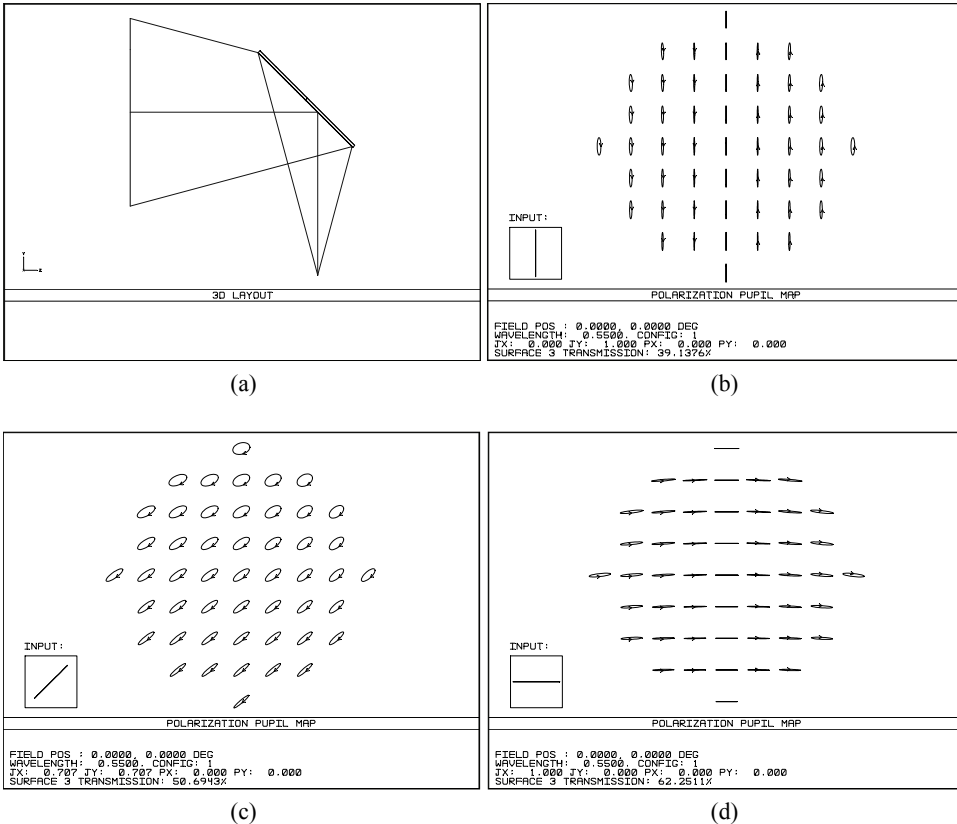


Figure 6.30 Polarization effect of an aluminum-coated folding mirror: (a) layout of a folding mirror, (b) polarization pupil map for a vertically linearly polarized beam, (c) polarization pupil map for a 45-deg linearly polarized beam, and (d) polarization pupil map for a horizontally linearly polarized beam.

As shown in Figs. 6.29 and 6.30, because of the increasing incident angle, the reflectance of the s component increases from the bottom toward the top of the folding mirror, while the reflection of the p component decreases. The polarization aberrations, both diattenuation and retardance, increase from the bottom to the top of the mirror. This phenomenon suggests that the crossed folding mirrors can compensate for the polarization effects induced by a single folding mirror because the s component on the first mirror becomes the p component on the second mirror and vice versa. Therefore, the diattenuation and retardance are exactly canceled for the central ray. For off-axis rays, the cancellation will not be perfect, but it may be good enough for some applications. Figure 6.31(a) is the configuration of crossed folding mirrors. Polarization pupil maps when the input light is vertically polarized, 45 deg linearly polarized, and horizontally polarized are given in Figs. 6.31(b)–6.31(d). Compared to Fig. 6.30,

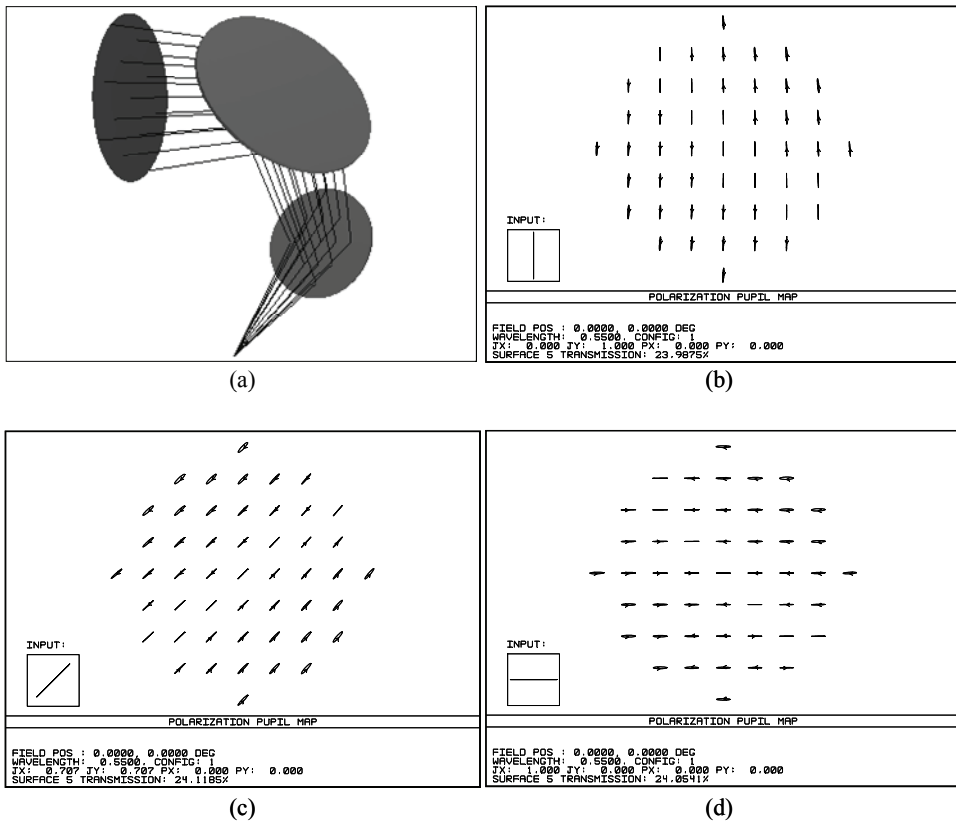


Figure 6.31 Polarization effect of aluminum-coated, crossed folding mirrors: (a) layout of the folding mirror, (b) polarization pupil map for a vertically linearly polarized beam, (c) polarization pupil map for a 45-deg linearly polarized beam, and (d) a polarization pupil map for a horizontally linearly polarized beam.

the polarization effect is much smaller, which means that the crossed folding mirrors effectively compensate for the polarization aberrations generated by the single folding mirror.⁶¹

An ideal thin-film beamsplitter splits an incident beam into two beams having close to 50% of the incident irradiance with some diattenuation. A polarization beamsplitter cube is commonly used in polarization imaging systems to separate or combine two polarization states. When two polarization beamsplitter cubes are used in a cascaded fashion, the FOV can be increased, and losses can be reduced by aligning the hypotenuses parallel to each other, as shown in Fig. 6.32. Losses are five times greater when the hypotenuses are aligned perpendicularly.⁶²

6.5.3 Birefringence

Birefringence is the difference in refractive indices of an optical material between two orthogonal polarization states. When the unpolarized light is incident upon a birefringent material, it is split into two paths with orthogonal polarization states, and the phase of each path accumulates at a different rate. This split leads to different optical path lengths and different focal lengths for the light with different polarization states, thereby degrading the imaging performance. The properties of birefringent materials can be modeled in commercially available optical design software.

In general, the influence of birefringence is regarded as a kind of polarization aberration caused by a change in the polarization state of the light being transmitted through an optical system. When the transmitted wavefront has a nonuniform polarization state across the exit pupil, the imaging properties are affected.

Birefringence can be caused by the crystalline structure of the material and the stress either inherent in or applied to the material. Intrinsic birefringence, also called natural birefringence, is a deviation of refractive index that cannot be reduced by the growth process. Calcite is one of the most important materials with intrinsic birefringence; it is used in a variety of polarizers, such as the Glan-Thompson prism.

Most optical elements are isotropic and have no natural birefringence. However, because optical elements are usually held using mechanical mounting methods, the interface between the optical element and the mechanical element

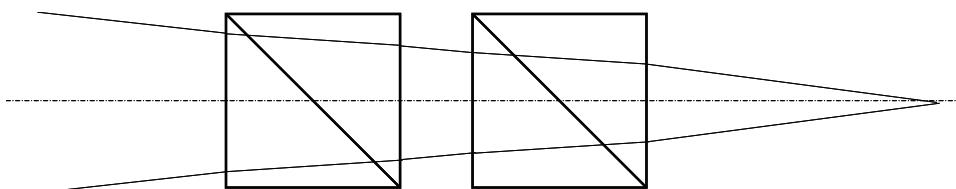


Figure 6.32 Optimal configuration of two polarization beamsplitter cubes used in a cascaded fashion.

may produce stress. Stress can introduce anisotropic material birefringence and result in a change in the optical path. The amount of stress-induced birefringence is determined by the stress-optical coefficient of the optical element and the mechanical stress exerted on it.

The amount and distribution of permanent inherent stress of an optical glass depends on the annealing conditions, glass type, size, and geometry. Stress inside glass is mainly due to the annealing process and variations in chemical composition within a melt. Chemical inhomogeneity can lead to local variations of the thermal expansion coefficient and, therefore, to permanent mechanical stress.

Stress birefringence Δn can be expressed as the optical path difference (OPD) between two orthogonal incident plane waves transmitting the sample with a length d at the measurement position,⁶³

$$\text{OPD} = \Delta n d = K \sigma d, \quad (6.29)$$

where σ (N/mm² or MPa) is the mechanical stress (positive for tensile stress) and K (mm²/N¹) is the stress-optical coefficient, which can be positive or negative depending on the material, temperature, and wavelength range.

The stress-optical coefficients K of the Schott glasses are measured at a wavelength of $\lambda = 0.5893 \mu\text{m}$ and an ambient temperature of 21°C. Figure 6.33 shows the stress-optical coefficients of selected Schott glasses. For most glass types, K is nearly constant in the visible range. High-lead-content glasses, like SF glasses, show a more pronounced dependence of the stress-optical coefficient on the wavelength. For most glasses, the stress-optical coefficient is between 0 and 3. Low-index glasses, such as N-FK51A, N-PK52A, and N-PK51, and high-index glasses, such as SF6 and SF57, have low stress-optical coefficients. The stress-optical coefficients for the lead-containing flint glasses (indicated with diamonds) are always smaller than their eco-versions in which lead is substituted by titanium or niobium. For example, the K values for SF11 and N-SF11 are 1.33 and 2.94, respectively.

The mechanical stress in optical glasses can be reduced with fine annealing by ideally crossing the strain as slowly as possible. However, for most glasses, the refractive index increases with decreasing annealing rate.

Birefringence is an inherent property of plastic optical materials and can be categorized into orientational birefringence and stress birefringence. The orientational birefringence is given by⁶⁴

$$\Delta n = f \Delta n_o, \quad (6.30)$$

where f is the orientational coefficient and Δn_o is the intrinsic birefringence.

Stress optical birefringence is given by

$$\Delta n_s = C \sigma, \quad (6.31)$$

where C is the photoelastic coefficient and σ is the stress.

Table 6.7 lists intrinsic birefringences and photoelastic coefficients of several polymers. Stress-optical coefficients of optical plastic are higher than for optical glasses.⁶⁴

In molded plastic elements, birefringence effects are process driven. Materials with higher flow rates are less susceptible to birefringence effects. When molding components, the harder a component is to fill, the more likely it is that birefringence will occur because more injection pressure or more holding pressure is required. Molding processes can be optimized to reduce birefringence at a possible cost, for example, rapid thermal response molding.

Except for the birefringence generated from mechanical loads during standard operation, birefringence can also be generated in optical components by a change in temperature. Because of the mismatch of coefficients in thermal expansion of cemented elements and the mismatch between the optical element and the mounting material, uniform temperature changes can produce mechanical stress in optical components.

The stress, whether mechanically induced, thermally induced, or residual, leads to a difference in refractive indices for light beams with different polarization directions. To understand the influence of stresses on optical performance, one can computationally simulate the effects of stressed optics by first calculating the mechanical stress and then calculating the optical performance. Unlike the birefringence in birefringent materials with homogeneous properties, which can be modeled by specifying the different refractive indices along the orthogonal axes in most commercially available optical design software, stress birefringence is more complex because mechanical stress typically varies in three dimensions within the optical element. The size, shape, and orientation of the index ellipsoid are different at each point. One practical method is to model the effects of a spatially varying stress field using interferogram files, which include the orientation of the crystal axis and the magnitude of birefringence. Stress birefringence interferogram files can be

Table 6.7 Intrinsic birefringence and photoelastic coefficients of selected plastic materials.

Optical plastics	Refractive index at 587nm	Intrinsic birefringence Δn_o	Photoelastic coefficient $C (\times 10^{-13} \text{ cm}^2/\text{dyn})$
PMMA	1.4918	-0.0043	-6
PC	1.5855	0.106	72
Polystyrene	1.5905	-0.1	-55
Zeonex COP	1.5261	N/A	5

generated using stress data obtained from finite element analysis. This topic is beyond the scope of this chapter, though more information can be found in Refs. 65–67.

An appropriate lens mount for an optical component can be designed to reduce stress-induced birefringence and distribute the residual stress evenly. The major considerations in designing lens mounts are the lens size, shape, and inherent rigidity. Other factors to consider include thermal effects and vibration, as well as long-term stability and cost. To provide a guideline for developing polarization imaging systems, ISO 101100-2 gives examples of permissible stress birefringence values for typical applications, listed in Table 6.8.⁶⁸ Generally, there is no requirement on the birefringence of the optical material used in the illumination path. The polarizer can always be placed as the last element in the illumination path to condition the illumination light. However, if the light from the light source is polarized or partially polarized, for example, as in a laser beam, it is preferable to use optical materials with low birefringence to reduce the light loss. For optical systems used in polarization imaging or for interferometric measurement, the requirement for the birefringence of the optical materials is tight. The permissible OPD through a system is less than 2 nm/cm.

6.6 Optical Design and System Analysis

6.6.1 General considerations

Generally, optical designs consist of lens and coating designs. Lens designs optimize the wavefront performance and the image quality without considering

Table 6.8 Permissible stress birefringence values for typical applications.⁶⁸

Permissible optical path difference per cm glass path	Typical applications
< 2 nm/cm	Polarization instruments Interference instruments
5 nm/cm	Precision optics Astronomical optics
10 nm/cm	Photographic optics Microscope optics
20 nm/cm	Magnifying glasses
Without requirement	Illumination optics

the coating effect. Coating designs maximize only transmittance or reflection within the wavelength range and the angle of incidence without consideration of diattenuation and retardance. In many situations, even for polarization imaging systems, this decoupled process, where the lens and the coating are designed independently, works well. However, for some polarization-sensitive imaging systems, it is necessary to analyze the polarization effects from the optical coatings when designing an optical system. Fortunately, most ray-tracing software now integrates thin-film optics and crystal optics calculations along with optical design calculations.

During the optical design phase, we need to determine the forms of polarization effects (diattenuation or retardance or both) and identify surfaces with large polarization effects. We also need to understand how the polarization aberrations add or cancel between surfaces. Low-polarization design requires that the optical system designer considers surface geometries, optical materials, coatings, and end-to-end polarization aberration analysis of the entire optical system. Measurements of coatings or dielectric samples during the design phase allow us to better simulate system performance.

The stress-induced birefringence and the coating are the major contributors to polarization aberrations. Stress birefringence usually occurs when the lenses are cemented together and accumulate with securely fitting frames. The optical glasses used in polarization imaging systems should have a low stress-optical coefficient to minimize the stress-induced birefringence. AR coatings also must be optimized for the most favorable polarization performance.

A typical design cycle of a polarization imaging system includes the following steps:⁶⁹

- (1) Design a preliminary optical system to meet the first- and third-order requirements other than polarization, such as the FOV, illumination uniformity, total track, NA, and MTF.
- (2) Select the appropriate polarization elements and optimize the illumination and imaging systems to meet the polarization-related requirements.
- (3) Identify the polarization-sensitive regions, which include surfaces with large angles of incidence, surfaces with small beam footprints, reflective surfaces, and folding mirrors.
- (4) Minimize these polarization-sensitive regions if it is not possible to avoid them.
- (5) Specify or design coatings to minimize the polarization effect within angles of incidence and wavelengths.
- (6) Evaluate the system performance through polarization ray tracing.
- (7) Tune the radius of the lenses as necessary to reduce the angle of incidence and redesign the coating to reduce the polarization aberrations of the entire system through phase compensation.
- (8) Assemble the system and measure the system performance. It may be necessary to simulate the system performance using measured data and comparing with the experimental results.

During the design stage, several rules of thumb, or design tricks, can be applied:⁶⁹

- Select the appropriate lens-mounting methods to minimize stress birefringence in the areas subject to thermal gradients.
- Note that plastic elements, especially injection-molded plastic elements, may have significant built-in stress birefringence.
- Use glasses with low stress-optical coefficients in locations with large stress.
- Keep the angle of incidence on the optical surface low to minimize the polarization effect.
- When folding is necessary, fold the system twice so that the s polarization on the first mirror aligns with the p polarization on the second mirror.
- Keep the optics out of the polarization-critical regions.
- Minimize the number of lens elements.

Figure 6.34 shows an example of a microscope lens where polarization effects are important. The first lens (L1) is more sensitive to stress-induced birefringence because all rays converge onto it and the beam size is smaller; therefore, it is necessary to use glasses that are resistant to stress birefringence. The ray angle in the front surface of the fifth lens (L5) is larger; the AR coating is critical on this surface. In particular, near the Brewster's angle of the lens surface, strong polarization effects occur, which have to be neutralized by a proper coating design.⁷⁰

6.6.2 Polarization ray tracing

To evaluate and optimize the performance of a polarization imaging system, it is necessary to perform a polarization ray trace. Polarization ray tracing presents a model of an optical system as sequences of media separated by optical

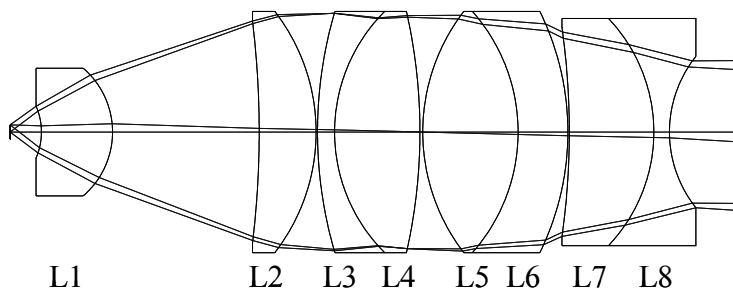


Figure 6.34 Microscope objective lens where polarization effects are of importance.

interfaces that are coated or uncoated air–glass interfaces. In addition to the ray coordinate and direction characterized in conventional geometrical ray tracing, additional attention has to be paid to the changes in the light amplitude, phase, and polarization state as well as to the effects of any polarization operators attached to the optical surface.

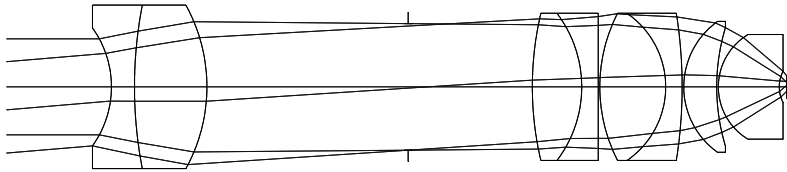
Detailed specifications of interface characteristics at each surface (for example, coating properties), polarization elements (idealized polarizers and retarders), and birefringent data of crystals are required in order to perform polarization ray tracing. However, a full polarization ray trace is not always possible because of the lack of information required to model the specific component.

Polarization ray tracing allows for the propagation of polarization states through the optical system and allows for the determination of the polarization properties of the ray paths. Using polarization ray tracing, one can calculate the magnitude of polarization aberrations, identify the sources of polarization aberrations, and tolerance coating. In short, it provides a detailed understanding of the polarization properties of the optical system.^{56,71–77}

The ray-tracing process is performed the same as usual, and the polarization properties of the ray are calculated as well. At each surface, the effects of the interface on the polarization state of the ray are calculated. The output of polarization ray tracing typically includes, but is not limited to, the following:⁷⁸

- The transmitted/reflected amplitude and phase,
- The relative intensity of the ray,
- Diattenuation,
- Retardance,
- Jones matrices (surface by surface),
- Mueller matrices (surface by surface),
- The degree of polarization of the ray,
- The direction of the rotation of the field vector for the polarized component of the field,
- Fractions of light in the incident and orthogonal polarization state,
- The orientation of the major axis of the polarization ellipse, and
- The ratio of the intensity of the minor axis component to the major axis component for a polarized portion of the field associated with the ray.

Figure 6.35(a) shows an infinite conjugate plan achromatic microscope objective lens with an NA of 0.7 and magnification of 40. Figure 6.35(b) shows the typical result of a polarization ray trace of a ray from the edge of the field to the 45-deg corner of the pupil when no coating is applied to any of the surfaces. The degree of polarization, the ratio of the intensity of the minor axis component to the major axis component, the orientation of the major axis of the polarization ellipse, and the direction of the rotation of the field vector for each surface are calculated. The accumulated phase, transmission, diattenuation, and retardance



(a)

Position 1, Wavelength = 587.6 NM						
	X	Y	POL FR	I RATIO	POL ANG	POL ROT
OBJ	0.00000	-0.8E+12	0.00000	0.00000	0.00000	LINEAR
1	2.04085	0.69676	0.00000	0.00000	0.00000	LINEAR
...						
4	2.72887	1.88414	0.01828	0.00000	-67.62670	LINEAR
STO	2.66982	2.64730	0.01828	0.00000	-67.62670	LINEAR
6	2.63298	3.12346	0.01960	0.00000	-62.41923	LINEAR
...						
17	0.13840	0.49575	0.22462	0.00000	-41.85416	LINEAR
18	0.09996	0.45269	0.22462	0.00000	-41.85416	LINEAR
IMG	0.00001	0.34073	0.22462	0.00000	-41.85416	LINEAR
Phase with respect to the chief ray (Waves):						
Yphase: 0.102 Xphase: 0.000						
Yphase2: 0.000 Xphase2: 0.102						
Intensity transmission: 0.4916						
Polarization properties for the ray path:						
	Magnitude	Orientation	Ellipticity			
Diattenuation:	0.2246240	-41.6530	0.0000			
Retardance:	0.021082(deg)	0.0000	-1.0000			
Polarization coordinate system rotation angle:	0.2117 (deg.)					
Product of beam projection factors along ray:	1.38501					

(b)

Phase with respect to the chief ray (Waves):			
Yphase:	0.000	Xphase:	NA
Yphase2:	NA	Xphase2:	0.000
Intensity transmission: 0.5047			
Polarization properties for the ray path:			
	Magnitude	Orientation	Ellipticity
Diattenuation:	0.0144356	0.0000	0.0000
Retardance:	0.000000(deg)	0.0000	0.0000
Polarization coordinate system rotation angle:	0.0000 (deg.)		

(c)

Figure 6.35 (a) An infinite conjugate plan achromatic microscope objective lens, (b) a polarization ray trace of the ray from the edge of the field to the 45-degree corner of the pupil, and (c) a polarization ray trace of the chief ray from the edge of the field.

are also computed. This ray has 0.225 diattenuation and 0.021-deg retardance, which is due to the Fresnel reflection. Figure 6.35(c) shows the result of a polarization ray trace of the chief ray from the edge of the field. The diattenuation is only 0.0144, much smaller than that of the marginal ray, which is due to the smaller angle of incidence at each surface. The intensity transmissions for the marginal and chief rays are 0.4989 and 0.5047, respectively.

The transmission, diattenuation, and retardance for the marginal and chief rays at 587.6 nm are listed in Table 6.9 for lens surfaces of the objective lens in Fig. 6.35(a) coated with a single layer of MgF_2 , two-layer V-coat, three-layer W-coat, 4-layer AR coating, and a combination of V-coat and 4-layer AR coating. The coatings listed in the table are designed without optimizing their retardance. As a comparison, the result of polarization ray tracing for the uncoated objective lens is also listed in the table.

The diattenuation and retardance of the coatings in Table 6.9 are plotted in Fig. 6.36 as a function of the angle of incidence at 587.6 nm. The diattenuation for different coatings is similar, but the retardance is quite different. The retardance of a single layer of MgF_2 and a W-coat are small and positive up to a 30-deg angle of incidence, but the retardance of a 4-layer AR coating increases quickly as the angle of incidence increases, and the retardance of the V-coat is negative and decreases as the angle of incidence increases. This property of the V-coat can be used to balance the retardance from other AR coatings, as demonstrated in Table 6.9.

Generally, AR coatings increase the transmission, but they also introduce diattenuation and retardance. With the proper design, coatings can increase the overall transmission but not for all of the wavelengths and all of the rays. When a W-coat is applied to all lens surfaces, the transmission of a marginal ray at

Table 6.9 Transmission, diattenuation, and retardance of the microscope objective lens in Fig. 6.35(a) when the lens surfaces are coated with different AR coatings.

Coating	Transmission		Diattenuation		Retardance	
	Marginal ray	Chief ray	Marginal ray	Chief ray	Marginal ray	Chief ray
Uncoated	0.4989	0.5047	0.1998290	0.014435	0	0
MgF_2	0.8721	0.8898	0.0680507	0.002692	1.064316	0.00821
V-coat	0.8709	0.9122	0.0206724	0.000753	11.40300	0.560862
W-coat	0.4760	0.7035	0.2172898	0.004009	19.52768	1.027359
4-layer	0.9362	0.9445	0.0236207	0.001984	16.53096	1.400331
V-coat + 4-layer	0.9246	0.9519	0.0157318	0.001665	0.407209	0.858399

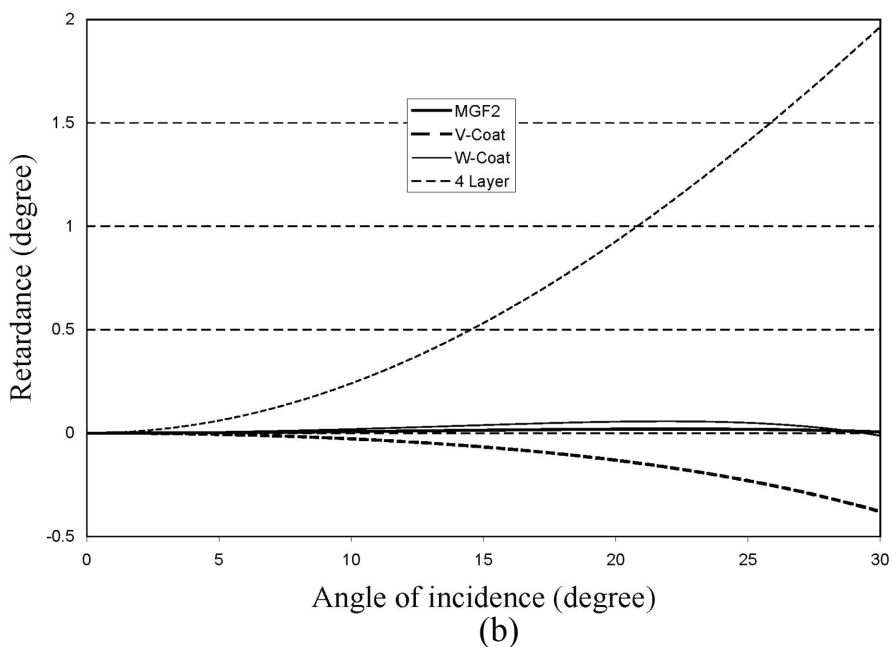
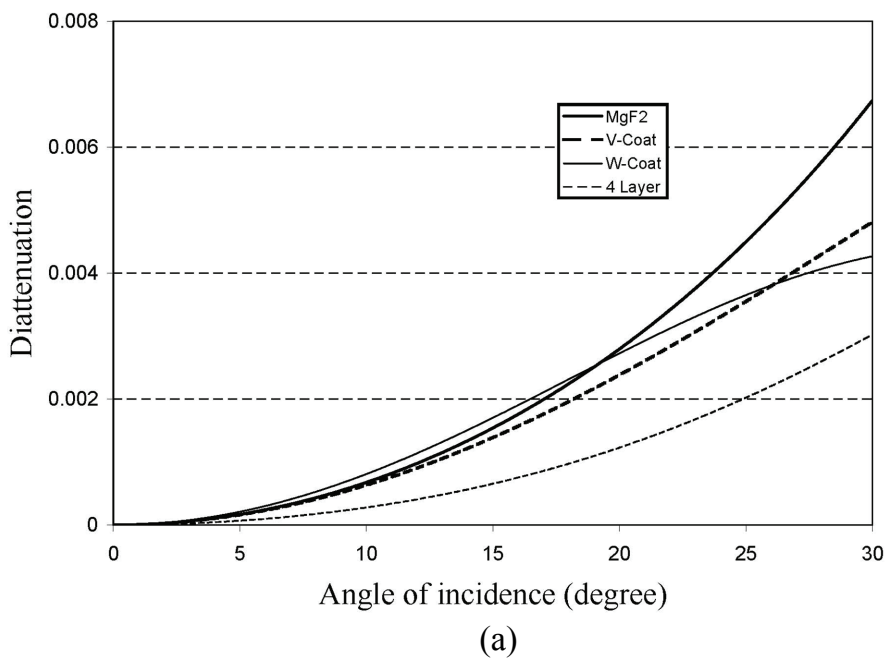


Figure 6.36 (a) Diattenuation and (b) retardance of four different coatings as a function of the angle of incidence at 587.6 nm.

587.6 nm is smaller than the transmission for an uncoated objective lens, and the W-coat introduces significant retardance. To enhance the transmission and minimize the diattenuation and retardance, lens surfaces can be coated with different AR coatings on the basis of the angle of incidence. As shown in the last row of Table 6.9, when lens surfaces 2–12 of the objective lens in Fig. 6.35(a) are coated with a V-coat coating and surfaces 13–17 are coated with a 4-layer AR coating, the transmission is higher than it is for the lens with one type of AR coating, and the diattenuation and retardance are smaller.

Instead of tracing a single ray, a polarization pupil map can give a comprehensive look at the polarization aberrations present in the system. Figure 6.37 shows the pupil maps of the polarization state, diattenuation, and retardance of an objective lens with a single-layer MgF_2 AR coating. The polarization effect

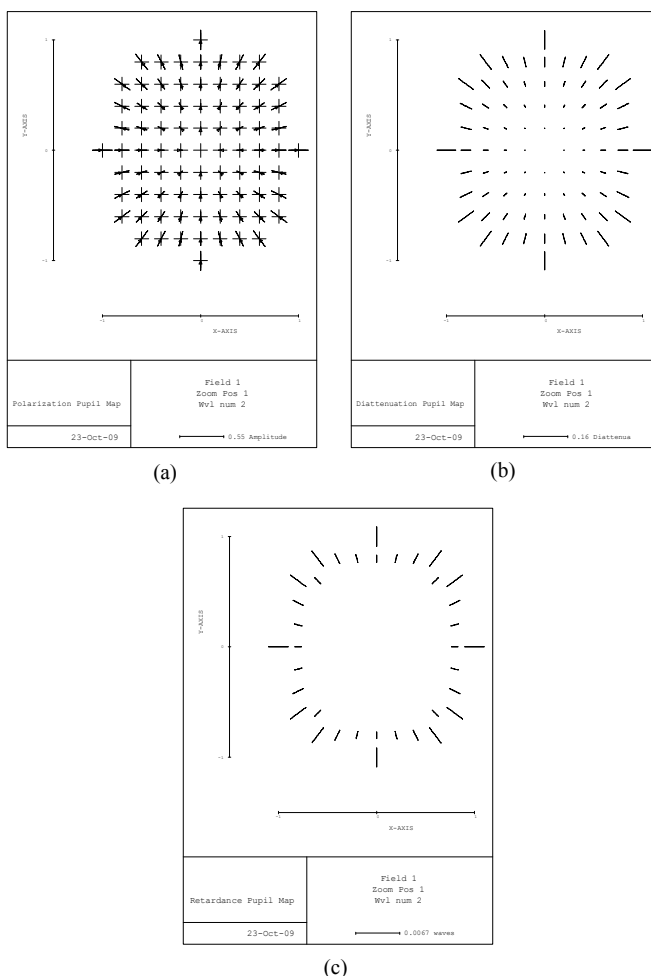


Figure 6.37 Pupil maps of the (a) polarization state, (b) diattenuation, and (c) retardance of an objective lens with a single-layer MgF_2 AR coating.

of the lens geometry and coating can also be demonstrated by the polarization intensity at the pupil. Figure 6.38 shows the polarization intensities for uncoated objective lenses and objective lenses with a single-layer MgF_2 AR coating. The overall polarization intensity distributions are similar, but the variation is different. Compared to the polarization intensity of an uncoated objective lens, single-layer MgF_2 AR coatings increase the intensity variation as a result of the polarization aberrations introduced by the AR coatings.

As another example, Fig. 6.39(a) shows a Schwarzschild microscope objective lens with an NA of 0.45 and a FOV of 0.35 mm. The two reflective mirrors are coated with a single layer of aluminum. Figure 6.39(b) shows the polarization ray trace for the ray from the center of the field to the 45-deg corner of the pupil. The first reflective mirror changes the linearly polarized light to be left elliptically polarized, and the second mirror changes it further to be right elliptically polarized. The differences between the phases of the s and p components introduced by the first and second reflective mirrors are 182.236 and 180.311 deg, respectively. The total retardance is 2.547, which is the sum of the phase difference between the two mirrors. This ray has an intensity transmission of 0.8433 and a diattenuation of 0.007.

Figure 6.40 demonstrates the polarization effect of the aluminum coating on the Schwarzschild microscope objective lens in Fig. 6.39(a). Figures 6.40(a) and 6.40(b) show the copolarized and cross-polarized components of a polarization intensity pupil map. The central white circle represents the central obscuration. The cross-polarized component is brighter at the ± 45 -deg corners because of the larger angle of incidence, while the copolarized component is brighter at the edge along the input polarization direction. Figures 6.40(c) and 6.40(d) show the copolarized and cross-polarized components of the PSF. There are four peaks in

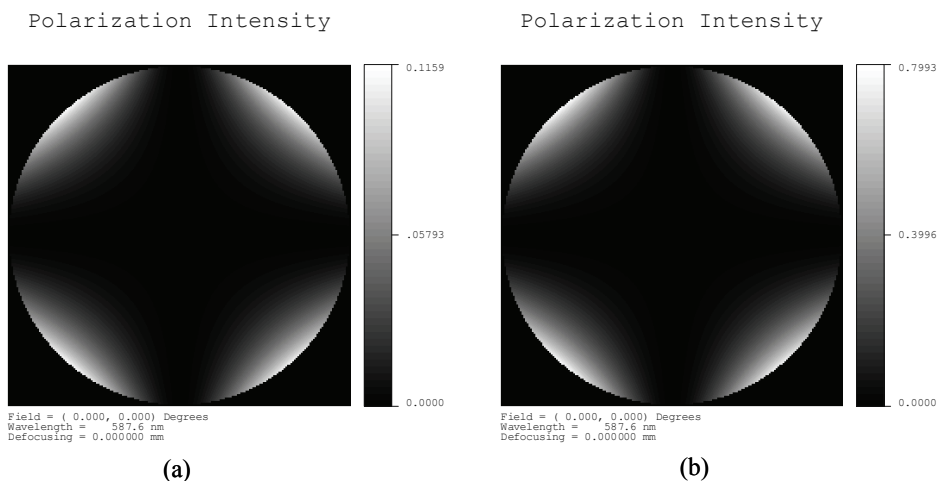
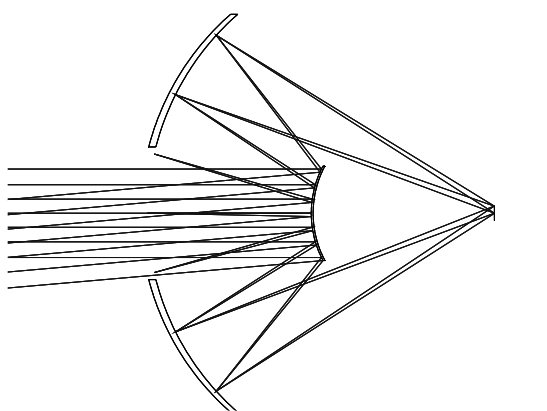


Figure 6.38 Polarization intensity of the objective lens without and with an AR coating: (a) uncoated and (b) with a single-layer MgF_2 AR coating.

the cross-polarized component at the ± 45 -deg corners, while the copolarization component has a peak at the center.

Polarization ray tracing can also calculate for Jones and Mueller matrices of each surface, which provide guidelines identifying the polarization-sensitive surfaces. Figure 6.41 shows the Jones and Mueller matrices of the Schwarzschild microscope objective lens in Fig. 6.39(a) with a single-layer aluminum coating.



(a)

Position 1, Wavelength = 587.6 NM								
	X	Y	POL FR	I RATIO	POL ANG	POL ROT	PHS	PHP
OBJ	0.000	0.000	1.000	0.000	0.000	LINEAR	0.000	0.000
1	2.227	2.227	1.000	0.000	0.000	LINEAR	0.000	0.000
2	2.227	2.227	1.000	0.000	-0.177	LEFT	165.548	-16.688
STO	9.145	9.145	1.000	0.000	0.202	RIGHT	164.624	-15.687
4	3.466	3.466	1.000	0.000	0.202	RIGHT	0.000	0.000
IMG	0.001	0.001	1.000	0.000	0.202	RIGHT	0.000	0.000
Phase with respect to the chief ray (Waves):								
	Yphase: -0.445		Xphase: 0.000					
	Yphase2: 0.000		Xphase2: -0.445					
	Intensity transmission: 0.8433							
Polarization properties for the ray path:								
	Magnitude		Orientation		Ellipticity			
Diattenuation:	0.0070509		45.0000		0.0000			
Retardance:	2.547076(deg)		45.0000		0.0000			
Polarization coordinate system rotation angle:					0.0000 (deg.)			

(b)

Figure 6.39 (a) Schwarzschild microscope objective lens and (b) polarization ray tracing for a ray from the center of the field to the 45-deg corner of the pupil.

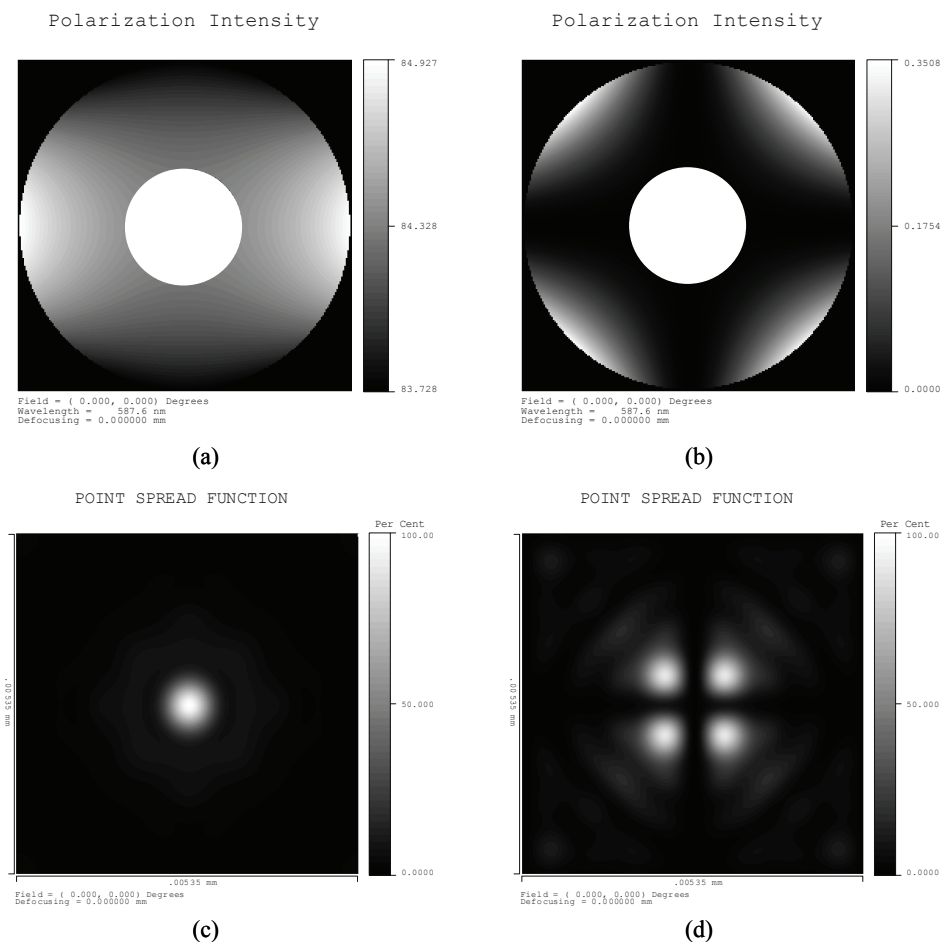


Figure 6.40 Polarization effect of aluminum coating on the Schwarzschild microscope objective lens in Fig. 6.39(a). (a) Copolarized and (b) cross-polarized components of a polarization intensity pupil map; (c) copolarized and (d) cross-polarized components of the point spread function.

	J11R	J11I	J12R	J12I
	J21R	J21I	J22R	J22I
OBJ	1.00000	0.00000	0.00000	0.00000
	0.00000	0.00000	1.00000	0.00000
1	1.00000	0.00000	-0.00000	-0.00000
	-0.00000	-0.00000	1.00000	0.00000
2	0.92334	-0.25660	0.00000	-0.00000
	-0.00000	0.00000	-0.92334	0.25660
STO	0.92334	-0.25660	0.00000	-0.00000
	-0.00000	0.00000	-0.92334	0.25660
4	0.00000	0.00000	1.00000	0.00000
	-1.00000	0.00000	0.00000	0.00000
IMG	1.00000	0.00000	0.00000	0.00000
	0.00000	0.00000	1.00000	0.00000

(a)

	M00	M01	M02	M03
	M10	M11	M12	M13
	M20	M21	M22	M23
	M30	M31	M32	M33
OBJ	1.00000	0.00000	0.00000	0.00000
	0.00000	1.00000	0.00000	0.00000
	0.00000	0.00000	1.00000	0.00000
	0.00000	0.00000	0.00000	1.00000
...				
2	0.91840	0.00000	0.00000	0.00000
	0.00000	0.91840	0.00000	0.00000
	-0.00000	0.00000	-0.91840	0.00000
	0.00000	0.00000	0.00000	-0.91840
STO	0.91840	0.00000	0.00000	0.00000
	0.00000	0.91840	0.00000	0.00000
	-0.00000	0.00000	-0.91840	0.00000
	0.00000	0.00000	0.00000	-0.91840
...				

(b)

Figure 6.41 (a) Jones matrix and (b) Mueller matrix of the Schwarzschild microscope objective lens in Fig. 6.39(a). The lens surface is coated with a single-layer aluminum coating. J_{xxR} and J_{xxI} are the real and imaginary parts of the Jones matrix, respectively.

References

1. W. S. Bickel, J. F. Davidson, D. R. Huffman, and R. Kilkson, "Application of polarization effects in light scattering: a new biophysical tool," *Proc. Natl. Acad. Sci.* **73**, 486–490 (1976).
2. R. R. Anderson, "Polarized light examination and photography of the skin," *Arch. Dermatol.* **127**, 1000–1005 (1991).
3. S. G. Demos and R. R. Alfano, "Temporal gating in highly scattering media by the degree of optical polarization," *Opt. Lett.* **21**, 161–163 (1996).
4. G. Jarry, E. Steimer, V. Damaschini, M. Epifanie, M. Jurczak, and R. Kaiser, "Coherence and polarization of light propagating through scattering media and biological tissues," *Appl. Opt.* **37**, 7357–7367 (1998).
5. S. G. Demos, W. B. Wang, and R. R. Alfano, "Imaging objects hidden in scattering media with fluorescence polarization preservation of contrast agents," *Appl. Opt.* **37**, 792–797 (1998).
6. J. A. Muccini, N. Kollias, S. B. Phillips, R. R. Anderson, A. J. Sober, M. J. Stiller, and L. A. Drake, "Polarized-light photography in the evaluation of photoaging," *J. Am. Acad. Dermatol.* **33**, 765–769 (1995).
7. S. L. Jacques, J. R. Roman, and K. Lee, "Imaging superficial tissues with polarized light," *Lasers Surg. Med.* **26**(2), 119–129 (2000).
8. S. G. Demos and R. R. Alfano, "Optical fingerprinting using polarisation contrast improvement," *Electron. Lett.* **32**, 2254–2255 (1997).
9. S. G. Demos and R. R. Alfano, "Optical polarization imaging," *Appl. Opt.* **36**, 150–155 (1997).
10. S. G. Demos, A. J. Papadopoulos, H. Savage, A. S. Heerdt, S. Schantz, and R. R. Alfano, "Polarization filter for biomedical tissue optical imaging," *Photochem. Photobiol.* **66**, 821–825 (1997).
11. V. Backman, R. Gurjar, K. Badizadegan, L. Itzkan, R. R. Dasari, L. T. Perelman, and M. S. Feld, "Polarized light scattering spectroscopy for quantitative measurement of epithelial cellular structures in situ," *IEEE J. Sel. Top. Quantum Electron.* **5**, 1019–1026 (1999).
12. L. T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J. M. Crawford, and M. S. Feld, "Observation of periodic fine structure in reflectance from biological tissue: a new technique for measuring nuclear size distribution," *Phys. Rev. Lett.* **80**, 627–630 (1998).
13. E. Collett, *Field Guide to Polarization*, SPIE Press, Bellingham, WA (2005) [doi: 10.1117/3.626141].
14. D. Goldstein, *Polarized Light*, Marcel Dekker, New York (2003).
15. J. S. Baba, J.-R. Chung, A. H. DeLaughter, B. D. Cameron, and G. L. Coté, "Development and calibration of an automated Mueller matrix polarization imaging system," *J. Biomed. Opt.* **7**(3), 341–349 (2002).

16. V. V. Tuchin, L. V. Wang, and D. A. Zimnyakov, *Optical Polarization in Biomedical Applications*, Springer, Berlin (2006).
17. S. G. Demos and R. R. Alfano, "Optical polarization imaging," *Appl. Opt.* **36**, 150–155 (1997).
18. F. C. MacKintosh, J. X. Zhu, D. Pine, and D. A. Weitz, "Polarization memory of multiply scattered light," *Phys. Rev. B* **40**, 9342–9345 (1989).
19. D. Bicot, C. Brosseau, A. S. Martinez, and J. M. Schmitt, "Depolarization of multiply scattered waves by spherical diffusers—Influence of the size parameter," *Phys. Rev. E* **49**, 1767–1770 (1994).
20. S. P. Morgan and M. E. Ridgway, "Polarization properties of light backscattered from a two layer scattering medium," *Opt. Express* **7**, 395–402 (2000).
21. G. Yao, "Differential optical polarization imaging in turbid media with different embedded objects," *Opt. Commun.* **241**, 255–261 (2004).
22. V. Sankaran, J. T. Walsh, and D. J. Maitland, "Comparative study of polarized light propagation in biologic tissues," *J. Biomed. Opt.* **7**, 300–306 (2002) [doi: 10.1117/1.1483318].
23. X. H. Ni and R. R. Alfano, "Time-resolved backscattering of circularly and linearly polarized light in a turbid medium," *Opt. Lett.* **29**, 2773–2775 (2004).
24. S. A. Kartazayeva, X. H. Ni, and R. R. Alfano, "Backscattering target detection in a turbid medium by use of circularly and linearly polarized light," *Opt. Lett.* **30**, 1168–1170 (2005).
25. R. Nothdurft and G. Yao, "Expression of target optical properties in subsurface polarization-gated imaging," *Opt. Express* **13**, 4185–4195 (2005).
26. S. L. Jacques, J. C. Ramella-Roman, and K. Lee, "Imaging skin pathology with polarized light," *J. Biomed. Opt.* **7**, 1–12 (2002) [doi:10.1117/1.1484498].
27. J. C. Ramella-Roman, K. Lee, and S. L. Jacques, "A hand-held polarized-light camera for the detection of skin cancer borders," *Proc. SPIE* **5192**, 54–61 (2003) [doi: 10.1117/12.534100].
28. W. Groner, J. W. Winkelman, A. G. Harris, C. Ince, G. J. Bouma, K. Messmer, and R. G. Nadeau, "Orthogonal polarization spectral imaging: a new method for study of the microcirculation," *Nat. Med.* **5**(10), 1209–1212 (1999).
29. M. Heger, J. F. Beek, K. Stenback, D. J. Faber, M. J. C. van Gemert, and C. Ince, "Darkfield orthogonal polarized spectral imaging for studying endovascular laser-tissue interactions in vivo—a preliminary study," *Opt. Express* **13**, 702–715 (2005).
30. P. T. Goedhart, M. Khalilzada, R. Bezemer, J. Merza, and C. Ince, "Sidestream dark field (SDF) imaging: a novel stroboscopic LED ring-based

- imaging modality for clinical assessment of the microcirculation,” *Opt. Express* **15**, 15,101–15,114 (2007).
31. M. P. Rowe, E. N. Pugh, Jr., J. S. Tyo, and N. Engheta, “Polarization-difference imaging: a biologically inspired technique for observation through scattering media,” *Opt. Lett.* **20**, 608–610 (1995).
 32. S. G. Demos, W. B. Wang, J. Ali, and R. R. Alfano, “New optical difference approaches for subsurface imaging of tissues,” in *Advances in Optical Imaging and Photon Migration, OSA Trends Opt. Photonics Ser.* **21**, 405–410 (1998).
 33. S. L. Jacques, J. C. Ramella-Roman, and K. Lee, “Imaging skin pathology with polarized light,” *J. Biomed. Opt.* **7**, 1–12 (2002) [doi:10.1117/1.1484498].
 34. S. G. Demos, H. B. Radousky, and R. R. Alfano, “Deep subsurface imaging in tissues using spectral and polarization filtering,” *Opt. Express* **7**, 23–28 (2000).
 35. K. Sokolov, R. Drezek, K. Gossage, and R. Richards-Kortum, “Reflectance spectroscopy with polarized light: is it sensitive to cellular and nuclear morphology,” *Opt. Express* **5**, 302–317 (1999).
 36. K. Sokolov, L. T. Nieman, A. Myakov, and A. Gillenwater, “Polarized reflectance spectroscopy for pre-cancer detection,” *Technol. Cancer Res. Treat.* **3**(1), 1–14 (2004).
 37. J. F. de Boer, S. M. Srinivas, J. S. Nelson, T. E. Milner, and M. G. Ducros, “Polarization-sensitive optical coherence tomography,” in *Handbook of Optical Coherence Tomography*, B. E. Bouma and G. J. Tearney, Eds., pp. 237–274, Marcel Dekker, New York (2002).
 38. Y. Yasuno, S. Makita, Y. Sutoh, M. Itoh, and T. Yatagai, “Birefringence imaging of human skin by polarization-sensitive spectral interferometric optical coherence tomography,” *Opt. Lett.* **27**, 1803–1805 (2002).
 39. G. Yao and L.V. Wang, “Two-dimensional depth-resolved Mueller matrix characterization of biological tissue by optical coherence tomography,” *Opt. Lett.* **24**, 537–539 (1999).
 40. J. S. Baba, J. R. Chung, A. H. DeLaughter, B. D. Cameron, and G. L. Coté, “Development and calibration of an automated Mueller matrix polarization imaging system,” *J. Biomed. Opt.* **7**, 341–349 (2002) [doi: 10.1117/1.1486248].
 41. M. H. Smith, P. Burke, A. Lompadó, E. Tanner, and L. W. Hillman, “Mueller matrix imaging polarimetry in dermatology,” *Proc. SPIE* **3911**, 210–216 (2000) [doi: 10.1117/12.384904].
 42. J. M. Bennett, “Polarization,” in *Handbook of Optics*, vol. 1, 2nd ed., M. Bass, Ed., pp. 5.1–5.31, McGraw-Hill, New York (1995).

43. J. A. Dobrowolski, "Optical properties of films and coatings," in *Handbook of Optics*, vol. 1, 2nd ed., M. Bass, Ed., pp. 42.3–42.130, McGraw-Hill, New York (1995).
44. J. M. Bennett, "Polarizers," in *Handbook of Optics*, vol. 2, 2nd ed., M. Bass, Ed., pp. 3.1–3.70, McGraw-Hill, New York (1995).
45. D. S. Kliger, J. W. Lewis, and C. E. Randall, *Polarized Light in Optics and Spectroscopy*, Academic Press, San Diego, CA (1990).
46. R. E. Fischer, B. Tadic-Galeb, and P. R. Yoder, "Polarization issues in optical design," Chapter 19 in *Optical System Design*, 2nd ed., pp. 387–420, SPIE Press, Bellingham, WA (2008).
47. J. L. Pezzaniti and R. A. Chipman, "Angular dependence of polarizing beam-splitter cubes," *Appl. Opt.* **33**, 1916–1929 (1994).
48. M. G. Robinson, J. Chen, and G. D. Sharp, *Polarization Engineering for LCD Projection*, John Wiley & Sons, Chichester, UK (2005).
49. S. Eckhardt, C. Bruzzone, D. Aastuen, and J. Ma, "3M PBS for high performance LCOS optical engine," *Proc. SPIE* **5002**, 106–110 (2003) [doi: 10.1117/12.485562].
50. T. Baur, "A new type of beam-splitting polarizer cube," *Proc. SPIE* **5158**, 135–141 (2003) [doi: 10.1117/12.510767].
51. D. Hansen, E. Gardner, R. Perkins, M. Lines, and A. Robbins, "The display applications and physics of the ProFlux wire-grid polarizer," *SID Symp. Dig. Tech.*, 730–733 (2002).
52. Z. Z. Zhuang, J. S. Patel, and Y. J. Kim, "Behavior of the cholesteric liquid-crystal Fabry-Perot cavity in the Bragg reflection band," *Phys. Rev. Lett.* **84**, 1168–1171 (2000).
53. M. Stalder and M. Schadt, "Linearly polarized light with axial symmetry generated by liquid-crystal polarization converters," *Opt. Lett.* **21**, 1948–1950 (1996).
54. B. Wang and J. List, "Basic optical properties of the photoelastic modulator, part I: useful aperture and acceptance angle," *Proc. SPIE* **5888**, 436–443 (2005) [doi: 10.1117/12.617904].
55. T. Oakberg, "Relative variation of stress-optic coefficient with wavelength in fused silica and calcium fluoride," *Proc. SPIE* **3754**, 226–234 (1999) [doi: 10.1117/12.366332].
56. R. A. Chipman, "Polarization in optical design," SPIE short course, SPIE, Bellingham, WA (2006).
57. R. A. Chipman, "Polarization Aberrations," Ph.D. Dissertation, Univ. of Arizona, Tucson (1987).
58. M. Born, and E. Wolf, *Principles of Optics*, 7th ed., Cambridge University Press, New York (1999).

59. M. Shribak, S. Inoué, and R. Oldenbourg, "Polarization aberrations caused by differential transmission and phase shift in high numerical aperture lenses: theory, measurement, and rectification," *Opt. Eng.* **41**(5), 943–954 (2002) [doi: 10.1117/1.1467669].
60. L. Sun and S. Edlou, "Low birefringence lens design for polarization sensitive optical systems," *Proc. SPIE* **6289**, 62890H (2006) [doi: 10.1117/12.679416].
61. D. G. Crandall and R. A. Chipman, "Polarization aberrations of crossed folding mirrors," *Proc. SPIE* **2537**, 83–93 (1995) [doi: 10.1117/12.216373].
62. J. Pezzaniti and R. Chipman, "Cascaded polarizing beam splitter cubes in imaging systems," *Opt. Eng.* **33**(5), 1543–1549 (1994) [doi: 10.1117/12.166529].
63. *Schott Technical Information TIE-27: Stress in optical glass*, Schott AG, Duryea, PA (2004).
64. K. Minami, "Optical plastics," in *Handbook of Plastic Optics*, S. Baumer, Ed., pp. 123–160, Wiley-VCH, Weinheim, Germany (2006).
65. K. Doyle, J. Hoffman, V. Genberg, and G. Michels, "Stress birefringence modeling for lens design and photonics," *Proc. SPIE* **4832**, 436–447 (2002) [doi: 10.1117/12.486447].
66. K. Doyle, V. Genberg, and G. Michels, *Integrated Optomechanical Analysis*, SPIE Press, Bellingham, WA (2002) [doi: 10.1117/3.460595].
67. V. L. Genberg, G. J. Michels, and K. B. Doyle, "Making FEA results useful in optical analysis," *Proc. SPIE* **4769**, 24–33 (2002) [doi: 10.1117/12.481187].
68. *ISO 10110-2:1996 Optics and optical instruments—Preparation of drawings for optical elements and systems—Part 2: Material imperfections—Stress birefringence*, ISO (1996).
69. R. E. Fischer, B. Tadic-Galeb, and P. R. Yoder, *Optical System Design*, 2nd ed., SPIE Press, Bellingham, WA (2008).
70. W. Singer, M. Totzeck, and H. Gross, *Handbook of Optical Systems*, vol. 2, *Physical Image Formation*, Wiley-VCH, Weinheim, Germany (2005).
71. J. P. McGuire and R. A. Chipman, "Diffraction image formation in optical systems with polarization aberration. I: Formulation and example," *J. Opt. Soc. Am. A* **7**, 1614–1626 (1990).
72. R. A. Chipman, "Polarization analysis of optical systems," *Opt. Eng.* **28**(2), 90–99 (1989) [doi: 10.1117/].
73. R. A. Chipman and L. J. Chipman, "Polarization aberration diagrams," *Opt. Eng.* **28**(2), 100–106 (1989).
74. J. D. Trolinger, R. A. Chipman, and D. K. Wilson, "Polarization ray tracing in birefringent media," *Opt. Eng.* **30**(4), 461–466 (1991) [doi: 10.1117/12.55815].

75. J. Shamir and R.A. Chipman, "Polarization aberration effects propagated in optical systems," *J. Mod. Opt.* **38**, 327–347 (1991).
76. S. C. McClain, L. W. Hillman, and R. A. Chipman, "Polarization ray tracing in anisotropic optically active media. I. Algorithms," *J. Opt. Soc. Am. A* **10**, 2371–2382 (1993).
77. S. C. McClain, L. W. Hillman, and R. A. Chipman, "Polarization ray tracing in anisotropic optically active media. II. Theory and physics," *J. Opt. Soc. Am. A* **10**, 2383–2393 (1993)
78. *Code V10.1 Reference Manual*, Optical Research Associate, Pasadena, CA (2009).

Chapter 7

Confocal Imaging

Since it was invented by M. Minsky in 1957,¹ confocal imaging has gradually become a widely used imaging tool in fundamental research and clinical applications. Confocal imaging is a technique for increasing the contrast of microscope images, particularly for thick specimens. It has several advantages over conventional optical microscopy, including controllable depth of field, better image quality, and the ability to collect optical sections of thick specimens.

The principle and basic components of confocal imaging will be discussed in Secs. 7.1 and 7.2. Sections 7.3 and 7.4 will focus on confocal scanning systems and optical design of the objective lenses. Fiber-optic confocal imaging systems will be discussed in Sec. 7.5.

7.1 Introduction

7.1.1 Principle of confocal imaging

The concept of confocal imaging is straightforward and is shown in Fig. 7.1. The illumination from the light source passes through the illumination pinhole P_1 and is focused on a small spot O on the object of interest. The objective lens collects the light returning back from the object and focuses the collected light onto the image plane, where a detection pinhole P_D is located. This pinhole is optically conjugated with the focused spot of the objective lens, which is also the illumination point. It blocks the majority of the light coming back from outside the focused point, such as the light from spots O_1 , O_2 , and O_3 (dashed line in Fig. 7.1) but passes the light that is returned from the focused point O (solid line). In the case of the reflectance confocal microscope where the illumination and detection light have the same wavelength, the beamsplitter directs part of the illumination light through the objective lens onto the object and then passes part of the reflected light that is collected by the objective lens onto the detector. Therefore, the beamsplitter introduces a light loss. For confocal fluorescence imaging, the beamsplitter is replaced by a dichroic mirror to separate the excitation wavelength and the fluorescence wavelength. Generally, dichroic mirrors are capable of reflecting and transmitting appropriate wavelengths at high

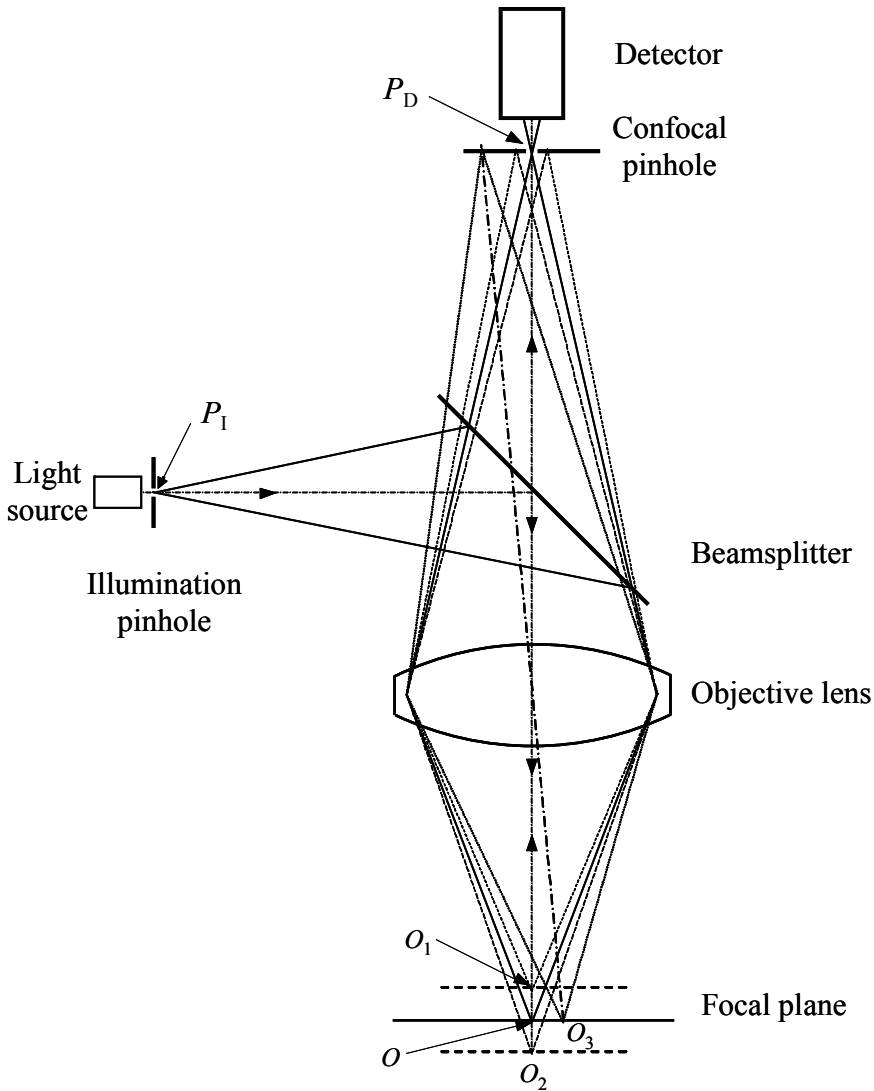


Figure 7.1 Schematic diagram of a confocal microscope.

efficiency (above 90%); additionally, they are often designed to reflect shorter wavelengths and to transmit longer wavelengths. Additional filters, such as the excitation filter in the illumination path and the emission filter in the detection path, are necessary to improve SNR.

Unlike a light microscope where the entire FOV can be imaged onto an image plane and can be viewed at any instant, a confocal microscope only images a single or multiple points onto the detector at a time. To obtain a continuous 2D image, a scanning system is required to scan the focused spot across the object. Details on scanning methods will be discussed in Sec. 7.3.

Confocal microscopy is usually performed either in an epi-illumination reflection mode or an epifluorescence mode. In addition to fluorescence and reflection imaging, the confocal technique can be combined with other imaging modes, such as dark-field and polarization imaging.

The image quality of a confocal imaging system is determined not only by the optics, as in a conventional microscope, but also by the size of the detection pinhole and the digitization of the detected signal. Another important factor affecting image quality is the noise from the light source, detector, electronics, and other components in the optical path.

7.1.2 Resolution

The detection pinhole is sufficiently small to only pass light from a focused spot; most of the light from outside the focused spot is blocked and will not reach the detector. Therefore, a confocal microscope can achieve better resolution than a conventional light microscope. As a result of out-of-focus light rejection, a confocal microscope is capable of imaging a thin section of the object; this capability is also referred to as optical sectioning. Optical sectioning, in turn, makes it possible to create 3D images of an object by reconstructing the optical sections captured at short and regular intervals along the optical axis.

The theoretical performance of a confocal microscope can be modeled using Fourier theory.^{2,3} The image of an ideal point object is always somewhat blurred because of the properties of an imaging system, which can be described in quantitative terms by the PSF. The PSF maps the intensity distribution in the image space. A confocal microscope generates point images in two ways: (1) by imaging a point light source into the object space and (2) by projecting a point of the object into the image space. The total PSF (PSF_{tot}) of a confocal microscope is composed of the PSFs of the illuminating beam path (PSF_{ill}) and the detection beam path (PSF_{det}):

$$\begin{aligned} \text{PSF}_{\text{tot}}(r) &= \text{PSF}_{\text{ill}}(r)\text{PSF}_{\text{det}}(r) \\ &\propto \text{somb}^2\left(\frac{2\text{NA}_{\text{ill}}}{\lambda_{\text{ill}}}\right)r \text{somb}^2\left(\frac{2\text{NA}_{\text{det}}}{\lambda_{\text{det}}}\right)r, \end{aligned} \quad (7.1)$$

where NA_{ill} and λ_{ill} are the NA and wavelength of the illumination system, NA_{det} and λ_{det} are the NA and wavelength of the detection system, and r is the radial position. PSF_{ill} corresponds to the light distribution of the focused illumination spot that scans the object. It is a function of the wavelength and the NA of the objective lens. It is also influenced by the diffraction of the objective pupil and the aberrations of all optical components in the illumination path. PSF_{ill} is further affected by tissue scattering and the spherical aberration introduced by the index mismatch when the focused illumination spot enters a thick and scattering object. PSF_{det} is influenced by all of the above factors as well as by the size of the

pinhole in front of the detector. Since the pinhole is never truly infinitely small, PSF_{det} is never smaller in dimension than PSF_{ill} .

For a reflectance confocal microscope with epi-illumination, the wavelengths of illumination and detection are the same, and the same objective lens is used for both illumination and detection; therefore,

$$\text{PSF}_{\text{tot}}(r) \propto \text{somb}^4\left(\frac{2\text{NA}}{\lambda}r\right). \quad (7.2)$$

Figure 7.2 shows the profile of the PSF_{tot} of an ideal confocal microscope objective lens. As a comparison, the PSF of an ideal conventional light microscope objective lens is also plotted as the dashed line. For both objective lenses, the NA is 0.5, and the wavelength is $0.488 \mu\text{m}$. The first zeros of the two PSFs are at the same location. However, the FWHM of the PSF of a confocal microscope is significantly smaller than that of a conventional system.

The FWHM of the PSF of a confocal microscope is determined by the wavelength and NA through the following equation:³

$$\text{FWHM} = \frac{0.37\lambda}{\text{NA}}. \quad (7.3)$$

When the wavelengths of the illumination and detection are different, λ is replaced by $\bar{\lambda}$, the mean wavelength of the illumination and detection wavelengths:

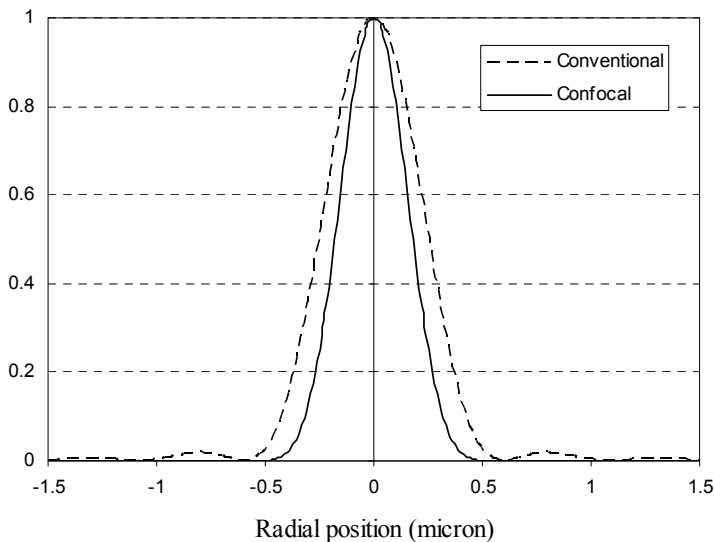


Figure 7.2 PSFs of ideal confocal and conventional microscope objective lenses. The NA of the objective lens is 0.5, and the wavelength is $0.488 \mu\text{m}$.

$$\bar{\lambda} = \sqrt{2} \frac{\lambda_{\text{ill}} \lambda_{\text{det}}}{\sqrt{\lambda_{\text{ill}}^2 + \lambda_{\text{det}}^2}}. \quad (7.4)$$

In a confocal fluorescence microscope, the emission wavelength is longer than the excitation wavelength; therefore, the FWHM of the related PSF is wider than that of a reflectance confocal microscope.

By Rayleigh's criterion, two points are resolvable when the dip between the two peaks is approximately 74% of the peak; the lateral resolution of a confocal microscope with an infinitely small pinhole is⁴

$$\Delta r_{\text{confocal}} = \frac{0.44\lambda}{\text{NA}}. \quad (7.5)$$

A comprehensive discussion of the theoretical axial PSF of a confocal microscope is provided by Wilson.² For an ideal point scan confocal microscope with an infinitely small pinhole,

$$\text{PSF}(z) \propto \text{sinc}^2\left(\frac{\text{NA}^2}{2\lambda_{\text{ill}}} z\right) \text{sinc}^2\left(\frac{\text{NA}^2}{2\lambda_{\text{det}}} z\right). \quad (7.6)$$

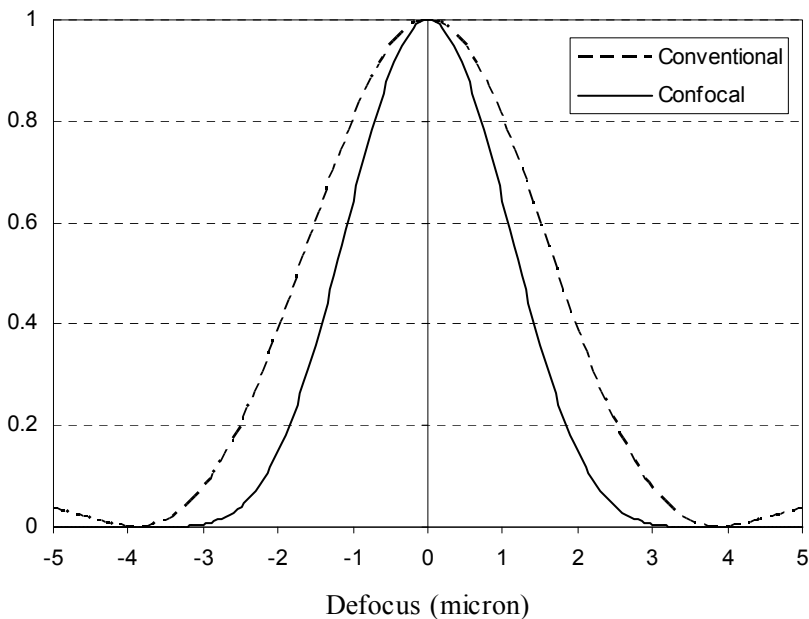


Figure 7.3 Axial PSFs of ideal confocal and conventional microscope objective lenses. The NA of the objective lens is 0.5, and the wavelength is $0.488 \mu\text{m}$.

The FWHM of the axial PSF is

$$\text{FWHM}_{\text{axial}} \approx \frac{1.28\bar{\lambda}}{\text{NA}^2}. \quad (7.7)$$

Figure 7.3 shows the axial PSFs for ideal reflectance confocal and conventional microscope objective lenses. The NA of the objective lens is 0.5, and the wavelength is 0.488 μm . The confocal microscope shows considerable improvement for point scanning through focus. By Rayleigh's criterion, the axial resolution of the confocal microscope with an infinitely small pinhole is⁴

$$\Delta z_{\text{confocal}} = \frac{1.5\bar{\lambda}}{\text{NA}^2}. \quad (7.8)$$

The axial and lateral resolutions in Eqs. (7.5) and (7.8) are derived for an infinitely small pinhole. When the pinhole is not infinitely small, the factors in Eqs. (7.5) and (7.8) increase.⁵ The axial resolution of a confocal microscope can be measured by moving a surface through the focal plane and plotting the measured signal as a function of z . $\text{FWHM}_{\text{axial}}$ relates to the axial resolution as⁴

$$\text{FWHM}_{\text{axial}} = 0.84\Delta z_{\text{confocal}}. \quad (7.9)$$

7.2 Basic Components and Requirements

Figure 7.4 is a schematic diagram of a confocal scanning microscope. Light from a coherent or noncoherent light source is collimated by the illumination optics before reaching the scan mirror. The scan mirror rotates on its pivot point to angularly scan the collimated beam. The telecentric relay lens images the scan mirror onto the entrance pupil of the objective lens to reduce the illumination falloff at the entrance pupil. The entrance pupil of the objective lens here is referred to as the image of the aperture stop on the image side. The rotation of the collimated beam in the plane of the scan mirror is translated to the beam rotation at the entrance pupil plane. The objective lens focuses the beam into a small spot on the object and converts the angular rotation of the scanning beam to a lateral movement of the focused spot. The backscattered light or excited fluorescence light is collected by the objective lens and then descanned by the scan mirror. After being descanned, the light from the object is directed to the pinhole lens by the beamsplitter, which can be a conventional amplitude beamsplitter, polarization beamsplitter, or dichroic beamsplitter, depending on the application. The light is then focused onto the detection pinhole and reaches the detector. It is helpful to have the detector at a conjugate plane of the entrance pupil of the objective lens so that the detector is filled for all pinhole sizes.

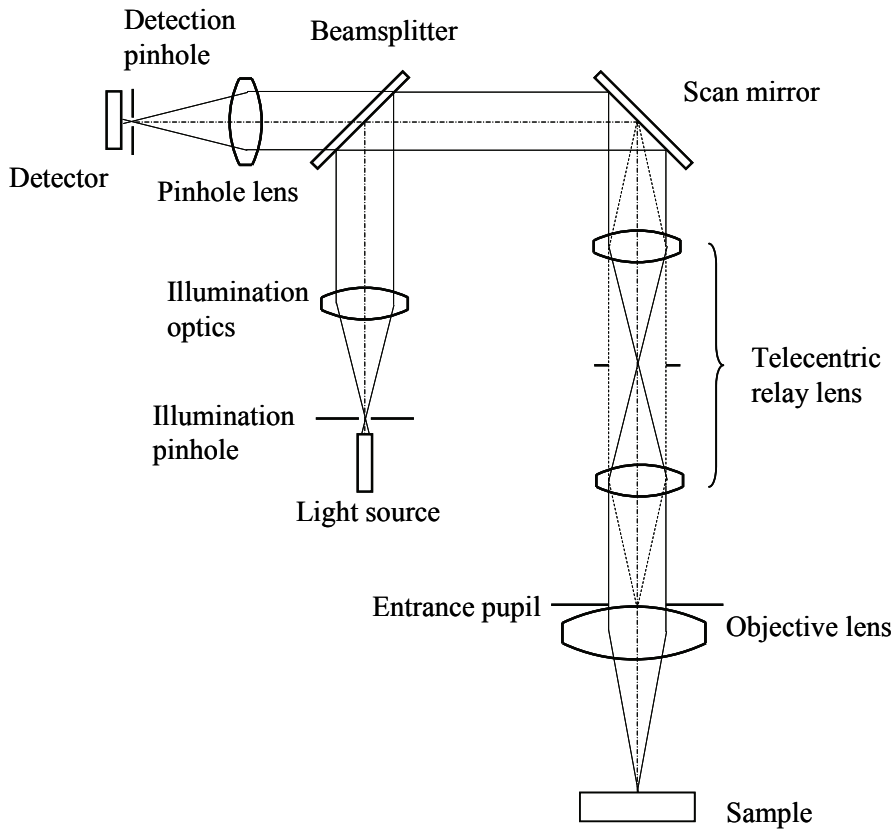


Figure 7.4 Schematic diagram of a confocal scanning microscope. The telecentric relay lens images the scan mirror onto the entrance pupil of the objective lens.

There are many variations of confocal imaging systems. However, the key components are the same: light source, beamsplitter, scanning device, objective lens, intermediate optics, and pinhole. The functions and fundamental requirements of each component will be discussed in the following sections.

7.2.1 Light source

Both coherent and noncoherent light sources are used in confocal microscopes.^{6,7} Lasers are coherent light sources and are ideal for confocal imaging because they have high brightness and low noise and can be focused to a very small spot. High-pressure, short-arc lamps and LEDs are two common noncoherent light sources used in confocal imaging systems.

Lasers are monochromatic light sources, which simplifies optical systems because no chromatic aberration correction is needed in the reflectance mode. Beam quality is a critical factor when selecting laser light sources. Ideally, a laser in TEM_{00} mode should be used because the beam can be collimated to fill the

objective lens uniformly without spatial filtering. If the laser beam contains more than one mode, it needs to be spatially filtered. Spatial filtering can be achieved by passing the focused laser beam through a small pinhole, which can also act as the illumination pinhole.

The most important specification of a laser for confocal imaging is its intensity stability because any variation will be interpreted as a change in reflected or fluorescence intensity of the object. On the other hand, wavelength stability is less important given that a small change in the source wavelength has little effect on the image quality.

Gas lasers, such as argon and helium-neon, have been commonly used in confocal imaging because of their long-term stability, long lifetime, low cost, and exceptional beam quality. One additional advantage of argon lasers is their ability to emit blue, green, yellow, and red wavelengths. The disadvantages of gas lasers include their high demand for water cooling, very small electrical-to-optical efficiency, noise generated by the power supply, bulkiness, and weight relative to other types of lasers.

Solid-state lasers have increasing application in confocal imaging because of their compactness, low heat dissipation, and long lifetime. Solid-state lasers offer a complete range of wavelengths spanning the UV to IR.

Diode lasers are compact and cost-effective light sources for confocal imaging, especially in the UV and deep blue wavelengths. Because of their elliptical wavefronts, beam-shaping elements are required to circularize the beam. Diode lasers can be modulated at high frequencies. They are compact and can be integrated into handheld devices.

Fiber lasers are a new type of laser used in confocal imaging. They convert light from an infrared pump diode directly into visible light. Fiber lasers are compact, have low power consumption, less heat dissipation, and fewer components than diode pump lasers. Fiber laser wavelengths are determined by the fibers and the resonator designs.

Noncoherent light sources, such as arc lamps and LEDs, are also used in confocal imaging. They eliminate the coherent interference of light reflected from different layers in an object, which can significantly alter image quality. Noncoherent sources are better suited for spinning-disk confocal microscopes than scanning confocal imaging. A potential disadvantage of using a broadband light source is that the chromatic aberration of the optical system needs to be minimized if the bandwidth of the bandpass filter is not narrow enough.

The most important characteristic of noncoherent light sources is radiance, which is determined by the source geometry, spatial distribution, angular emission profile, and the light output. Of the entire emitting cone, the conservation of étendue determines what fraction can be collected and delivered to the focused point.

Light from arc lamps, such as mercury or xenon arc lamps, can be easily focused onto an illumination pinhole compared to other types of broadband light sources, such as filament lamps. The high-pressure xenon short-arc lamp is often

used when a wide variety of excitation wavelengths is required in confocal fluorescence imaging.

LEDs are more efficient than other noncoherent light sources in converting electricity into light, ranging from UV to near-IR. The spectral bandwidths of light from LEDs are typically between 12 and 40 nm. LEDs are also more rugged and compact and can often outlive short-arc lamps. Light from LEDs may be modulated at high frequencies, for example, 10 MHz. The brightness may be regulated by the drive electrical current, eliminating the need for mechanical shutters and neutral-density filters.

LEDs have the potential for inexpensive, stable, efficient, and long-lived illumination in both wide-field and disk-scanning confocal microscopy. Like diode lasers, LEDs are ideal for handheld devices.

7.2.2 Illumination system

An illumination system consists of a light source and a light delivery system. The light delivery system further consists of one or more of the following subsystems: beam-shaping optics, relay optics, and scanning optics. Relay optics and scanning optics are usually in the detection path of the beam-scanning confocal imaging system as well and will be discussed in Sec. 7.3.

To achieve full resolution, the entrance pupil of the objective lens should be uniformly filled with the illumination light. A laser beam can be collimated readily by a single aspherical lens or several spherical lenses. The size of the collimated beam can be expanded to the size of the entrance pupil by a beam expander, which usually consists of two positive lenses or one negative lens and one positive lens. By adjusting the distance between the two lenses in the beam expander, the collimated laser beam can be turned to slightly diverging or converging and focused onto different focal planes along the optical axis of the scan lens. This helps confocal fluorescence imaging systems to colocalize the illumination and detection points, especially for objective lenses with chromatic aberrations.

For noncoherent light sources, such as an arc lamp, extra measures are required to illuminate the pupil efficiently because the emitter in noncoherent lamps radiates into a sphere or hemisphere, and the wavelengths from the noncoherent light sources, except LEDs, usually span a wide range of the spectrum. A collector, such as a reflector, is needed to redirect the light toward the entrance pupil of the objective lens. Unwanted wavelengths, IR and UV, must be filtered out using dichroic filters. The collected light is then delivered to the entrance pupil through the illumination system, for example, a Köhler illumination system.

A widely used and practical method of delivering the light from a remote light source, coherent or noncoherent, to the confocal imaging system is to couple the light into a single optical fiber. When the coherent light is coupled into a single-mode fiber, the output beam maintains its coherence and can be collimated and focused onto a diffraction-limited spot. When the noncoherent light is coupled into a long and curved step-index multimode fiber, the beam at

the output surface is spatially uniform. The light emerging from the fiber is then treated as the light source and is delivered to the entrance pupil of the objective lens by other optical elements.

7.2.3 Objective lens

The objective lens is the most critical component of the confocal imaging system in determining the contrast and resolution, the depth from which information can be obtained, and the lateral extent of the imaging field. Factors to consider when designing or selecting an objective lens include the NA, immersion medium, working distance, and aberration correction. The NA of the objective lens determines the light collection efficiency, resolution, and depth of field, and, therefore, the thickness of the optical section.

Microscopic objective lenses both with and without coverslip correction have been used in confocal microscopy. However, the observation plane of the confocal microscope is usually within a medium having a refractive index slightly larger than 1.33. This introduces some spherical aberration, which degrades contrast and resolution of the confocal image. Water is the closest match to the refractive index of biological tissue and is a popular embedding media—it helps to reduce spherical aberration even for very deep sections. In recent years, water-immersion objective lenses, with their long working distances and good UV transmission, have been developed for confocal microscopes.

The working distance is the distance between the focal plane and the closest mechanical element of the objective lens. It decreases significantly for an objective lens with a higher NA and a greater aberration correction. The working distance is critical when imaging thick objects or when manipulation of the object under an objective lens is required.

An objective lens in confocal imaging is usually designed to be telecentric so that the magnification does not change as it images deep into the tissue and the reflected light cone is perpendicular to the tissue surface. Because the objective lens in a confocal imaging system is used for both illumination and light detection, aberration in the objective lens must be kept to a minimum.

7.2.4 Beamsplitter

The majority of confocal microscopes use epi-illumination. A beamsplitter is required to separate the illumination and observation beams. The light loss in the illumination and detection paths is unavoidable when a beamsplitter is used to direct the illumination light to an object and deliver the detection light to a detector. A 50/50 beamsplitter is usually used to maximize the overall light efficiency, which is 25%. Half of the illumination light is lost, along with half of the detection light. For some applications, it may be advantageous to limit light exposure on the object and at the same time increase signal light collection from the object. In such cases, a 90/10 beamsplitter, which transmits 90% of the detection light and reflects 10% of the illumination light, can be used. The overall light efficiency is 9%, which is much less than the light efficiency of a

50/50 beamsplitter, but 90% of the light collected by the objective lens is directed to the detector, instead of 50%. The loss of the illumination light can be remedied by using a higher-power light source. Since the beamsplitter is in the imaging path, a transmitted wavefront error of $\lambda/4$ in a single pass is desired for most imaging applications.

In single-photon excitation fluorescence confocal imaging, the fluorescence always has a longer wavelength than the illumination. The beamsplitter in fluorescence imaging is usually a dichroic mirror that transmits one spectral band and reflects the other. Separating the excitation and emission spectra perfectly without loss in fluorescence is preferable. However, the two spectra are often too close to be separated completely by a dichroic mirror. Sometimes they overlap, making this separation even more difficult. If some light must be lost, it is better to keep as much fluorescence light as possible. A more powerful light source can be used to compensate for the illumination loss.

Instead of using beamsplitters, an alternative method to separate the illumination and detection beam paths is to spatially divide the pupil. For example, the circular pupil can be split into two halves, one for illumination and the other for detection. The pupil can be also divided annularly, with the central portion for illumination and the outer annulus for detection or vice versa, as shown in Fig. 7.5(a). Pupil division is a lossless means of separating the illumination and detection beams, and it does not require an additional component to separate the illumination and detection beams. The drawback is the loss in resolution or contrast because the full pupil is not available for illumination and imaging. This method is feasible for confocal systems with low resolution.

Another confocal configuration that does not use a beamsplitter is the dual-objective confocal system.⁸ As shown in Fig. 7.5(b), two separate low-NA objective lenses are used: one for illumination and the other for detection. The optical axes of the two objective lenses are oriented at an angle and intersect at the focal point. For the dual-objective system, the combined PSF is the product of two individual PSFs. The intersecting region, represented by the small black oval, is characterized by narrow transverse dimensions; therefore, subcellular resolution, both axial and lateral, can be achieved. This configuration has several advantages. Low-NA lenses are simpler and less sensitive to aberration. The working distance is longer than that of a single high-NA objective lens with the same achievable resolution.

7.2.5 Pinhole

The optical sectioning capability of a confocal microscope is derived from having a pinhole in front of the detector to reject light from outside the focused point. The detection pinhole in front of the detector is generally referred to as the pinhole of the confocal imaging system. A physical illumination pinhole is sometimes used to control the quality of the illumination beam and is the reference point for other optical components. When a fiber is used to deliver the illumination light, it acts as the illumination pinhole.

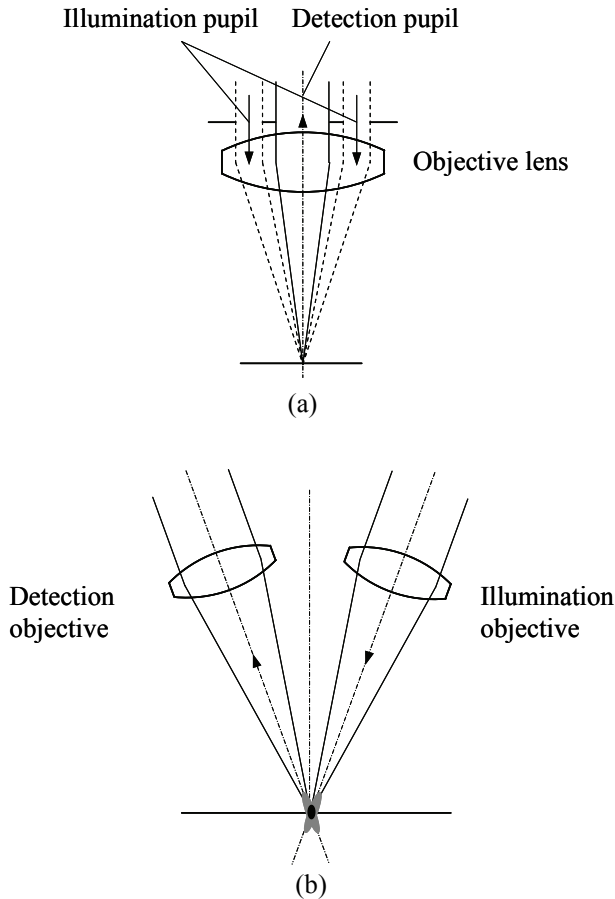


Figure 7.5 Schematic diagrams of confocal configurations without beamsplitters: (a) pupil division design and (b) dual-objective design.

The pinhole diameter plays a decisive role in resolution and depth discrimination. When the pinhole diameter is much larger than the Airy disk of the pinhole lens, the law of geometric optics determines the depth-discriminating properties. When the pinhole diameter is smaller than the Airy disk, the law of wave optics determines image formation.⁵ The smaller the pinhole is, the better the discrimination against the scattered light and the thinner the detected optical section. However, as the pinhole gets smaller, less light passes through it to reach the detector, leading to a reduced intensity of the detected signal and a lower SNR. In contrast, if the detection pinhole is larger, more light is detected from outside the focused point, and system performance is more stable. However, the larger the pinhole is, the lower the degree of confocality. When the pinhole aperture is fully open, the system is no longer a confocal system.

A compromise has to be made between the minimum thickness of the optical section and the maximum detection signal. It has been shown that optical sectioning does not improve considerably with a pinhole size below a limit that

approximates the radius of the first zero of the Airy disk. Thus, a good approximation is to make the pinhole about the size of the Airy disk. Sometimes, a circular iris, which can be opened and closed across a continuous range of diameters, is used to achieve the best compromise between confocal performance and signal strength without interruption of the imaging.

One of the issues in the early development of confocal microscopes was making and using a sufficiently small pinhole aperture.⁹ It is difficult to align and keep the small pinhole clean. The most effective solution is to enlarge the image significantly with the pinhole lens before casting it on the detection pinhole.

For multicolor confocal imaging systems, the detection light can be separated into each color, and each color has its own pinhole aperture and detector. This separation allows for an optimum optical section for each color and gives more flexibility when imaging multilabeled samples.

7.2.6 Pinhole lens

In general, a pinhole lens is placed in a location where the beam is already descanned and collimated. Therefore, the pinhole lens works only on axis as a focusing lens to focus the collimated beam onto the pinhole. Spherical and axial chromatic aberrations are the only two aberrations of concern. Uncorrected aberrations will spread the light distribution onto the pinhole plane, resulting in a reduced signal. Because the light passes through the pinhole lens only once, the imaging system is less sensitive to the aberrations in the pinhole lens than to those in the objective lens. A simple cemented doublet would meet the requirement of aberration correction when the detection spectrum is not too broad.

One requirement for the pinhole lens is that the focused spot should be reasonably large so that the pinhole aperture is not too small to make, align, and clean. Therefore, the focal length must be long enough. To reduce the total length of the detection path for a more compact system, a telephoto lens is usually designed. Figure 7.6 shows an example of a pinhole lens with a cemented positive lens and a negative singlet. The entrance pupil is 5 mm in diameter, the focal length is 125 mm, the NA is 0.02, and the total length from the lens to the focal plane is 50 mm. The PSF and MTF plots show that this lens has a diffraction-limited performance with an Airy radius of 18.2 μm . The axial chromatic aberration is 0.24 mm, smaller than the depth of focus, which is 0.6 mm, over the entire visible spectrum.

When a pinhole lens with a very low NA is used with a high-NA objective lens, the overall magnification is large, and the Fresnel number is small. Therefore, the intensity distribution is no longer symmetrical around the ideal image plane in the axial direction. The location of the highest intensity value on axis is shifted toward the pinhole lens instead of the ideal image plane.¹⁰ In confocal imaging systems, the depth discrimination is critical; therefore, this effect must be considered when building such a system.

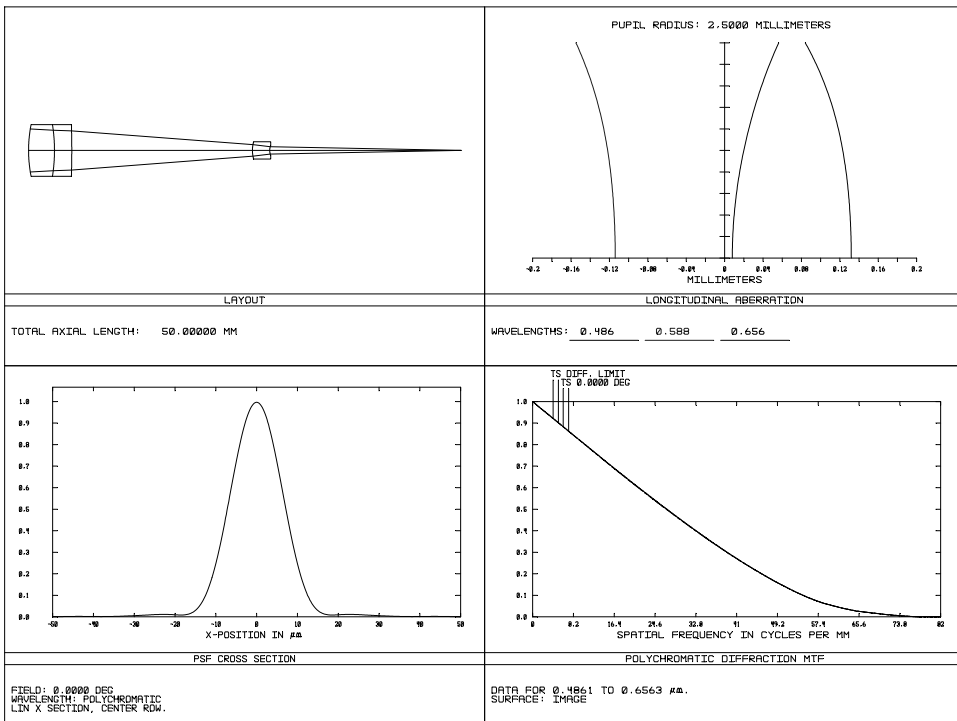


Figure 7.6 An example of a telephoto pinhole lens. The longitudinal aberration (top right), PSF (bottom left), and MTF (bottom right) are plotted to show the performance.

The primary performance requirement for a pinhole lens is that it should degrade the fundamental image information received from the objective lens as little as possible. On the other hand, with proper design, a pinhole lens can compensate for the remaining aberrations from the objective lens.

7.2.7 Scanner

As will be discussed in detail in Sec. 7.3, both a beam scan and stage scan have been used to scan the focused spot across an object to construct a 3D image. The beam scanning method is used in most commercial confocal imaging systems for biomedical applications. Several scanners have been employed in confocal imaging: galvanometer scanners, polygonal scanners, acousto-optical deflectors, and MEMS mirrors. Each of them has unique properties and is used in some applications.

7.2.7.1 Galvanometer scanner

A galvanometer scanner consists of three main components: the galvanometer, mirror, and servo driver that controls the system. The galvanometer itself has two major parts: the actuator that manipulates the mirror load and the integral position detector that provides mirror position information to a closed-loop

system. The mirror is an important component of the system, particularly for a high-speed scanner. Its parameters, including thickness, profile, cross section, and materials, influence the system's inertia as well as the stiffness and resonant frequency of the actuator and mirror assembly. The mirror surface must be large enough to hold the beam diameter over the entire scan range. For a two-axis scanning system without a relay lens, the second mirror is larger than the first, and the second mirror usually limits the speed of the scanning system.

A galvanometer scanner has two formats: a resonant scanner that vibrates at a fixed frequency and a linear galvanometer that operates at lower but variable frequencies. A resonant galvanometer usually operates at 4 kHz and above. A linear galvanometer can be operated at any drive frequency. Because of its operational adaptability, the linear galvanometer is often selected for experimental setups and applications of low-speed recurrent scanning or for precise random-access positioning. Major advantages of the resonant scanner are its simplicity, speed, compactness, long lifetime, and low cost. Its major disadvantages include its sinusoidal motion and its sensitivity to external, as well as self-induced, perturbations. The scanning characteristics of galvanometer mirrors are independent of wavelength; therefore, they are ideal for white-light or fluorescence imaging.

7.2.7.2 Polygonal scanner

A polygonal scanner incorporates a rotating optical element with three or more reflective facets, usually a metal mirror. There are several types of polygonal scanners; the two most frequently used are the prismatic scanner and the pyramidal scanner. The prismatic polygon scanner has a number of plane mirror facets that are parallel to, are equidistant from, and face away from a central rotational axis. The pyramidal polygon scanner, on the other hand, has a number of facets inclined at the same angle, typically 45 deg, to the rotational axis. Prismatic scanners are usually used with the input beam perpendicular to the rotation axis, whereas pyramidal scanners are used primarily with the input beam parallel to the rotational axis.

Simple and nondispersive, rotating polygons naturally scan the output beam to follow a simple sawtooth raster. However, rotating polygons have several limitations.¹¹ One of the issues is pyramidal error, which is the angular difference of each mirror surface with respect to the axis of rotation. The error causes the beam to fluctuate in the direction that is orthogonal to the scanning direction unless additional optical corrections are applied. Rotating polygons have been used in confocal imaging systems, but they are gradually being replaced by linear galvanometers.

7.2.7.3 Acousto-optical deflector

An acousto-optical deflector (AOD) is a device that deflects an optical beam into different directions by a surface acoustic grating. An AOD is based on the optical

effects of acoustic fields on a birefringent crystal. The scan angle depends on the wavelength of the light beam.

Compared to a galvanometer, an AOD provides a small deflection angle, a limited angle resolution, a noncircular aperture, and low light efficiency. It is capable of extremely high scan rates, up to 20,000 sweeps per second.

In a confocal imaging system, an AOD is usually used together with a galvanometer; the AOD performs a very rapid scanning on one axis, and the galvanometer controls the slower axis. Because their deflection angle is wavelength dependent, AODs are generally used in laser scanning systems, and they are not suitable for confocal fluorescence imaging systems.

7.2.7.4 MEMS scanner

A MEMS scanner is a device that has emerged in recent years to achieve the beam scan in confocal imaging.^{12,13} A MEMS scanner is a silicon device where the mirror is connected to small flexures that allow it to oscillate. One-dimensional and 2D MEMS mirrors have been developed and used in confocal imaging.

The MEMS scanner offers a number of advantages: small size, fast speed, high accuracy, and repeatability. However, the mirror size is limited and can only be used for small beams, possibly limiting the NA and, therefore, the resolution.

7.2.7.5 Rotating wedge scanner

The rotating wedge scanner consists of a pair of wedge prisms, which are usually identical, as shown in Fig. 7.7. The rotating wedge scanner has been implemented in some confocal imaging systems.¹⁴ Two prisms are placed with two tilted surfaces facing toward each other. Other configurations are possible, but they all have some limitations. Each prism produces a constant angular deviation of the propagation direction, and the azimuth of the deviated beam depends on the rotation position of the prism. By rotating two prisms independently, a ray normal to the prisms can be steered in any direction within a

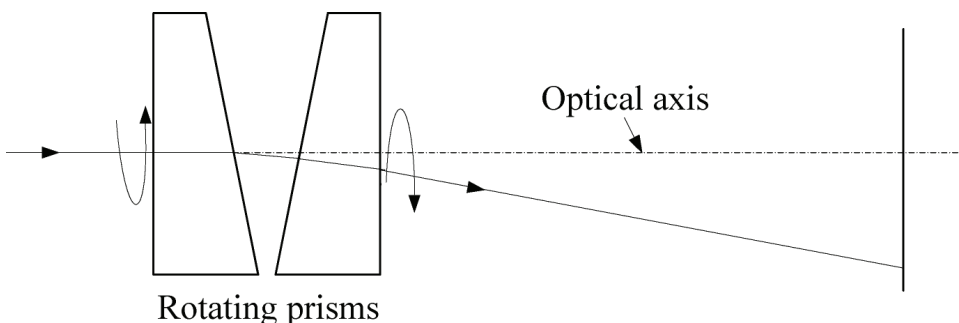


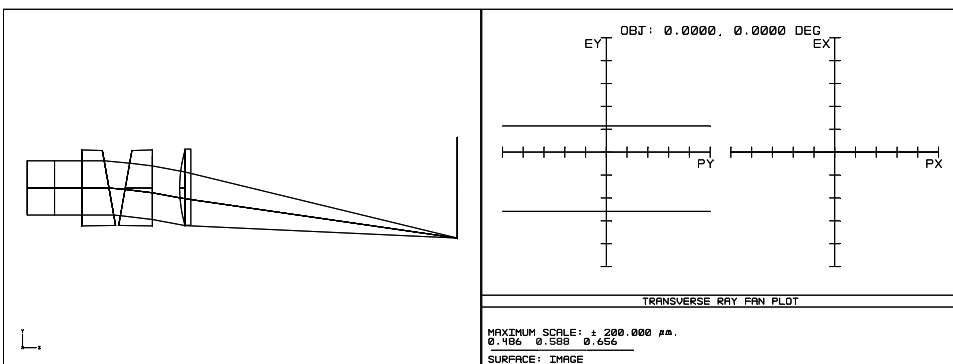
Figure 7.7 Diagram of a rotating wedge scanner. The two prisms rotate relative to each other.

cone around the optical axis. Any specific scanning pattern can be generated by synchronized rotation of the two prisms. When the wedge angle is increased to achieve a large scan angle, there is a distortion in the scan angle; this distortion depends only on the relative positions between the two prisms. The scan angle is either compressed or stretched in some directions.¹⁵

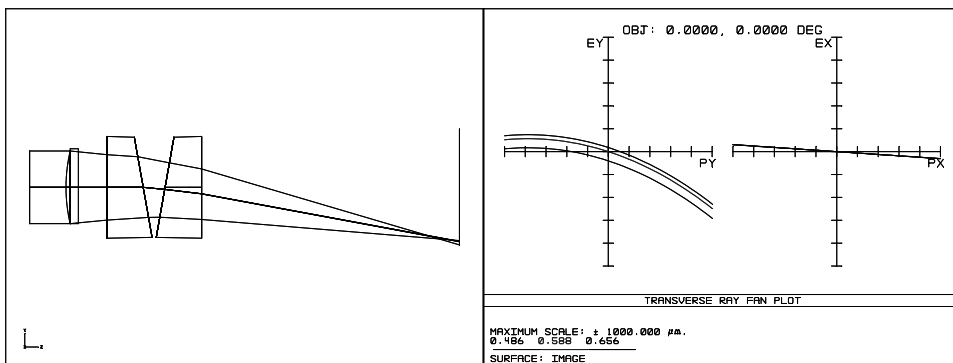
The rotating wedge scanner is optimally used in the collimated space; otherwise, it introduces all aberrations except field curvature.¹⁶ Figure 7.8(a) shows the ray diagram and ray fan plot of a pre-objective lens scanning configuration with a rotating wedge prism in front of the ideal lens. Since the rotating wedge scanner is in the collimated space, it introduces only lateral chromatic aberration. However, as shown in Fig. 7.8(b), when the wedge scanner is in the convergent space, which is the post-objective scanning configuration, it generates a significant amount of aberration.

7.3 Confocal Scanning Systems

The typical confocal imaging system is a point-by-point imaging device. To



(a)



(b)

Figure 7.8 Schematic diagrams and ray fan plots of scanning systems using rotating wedge scanners: (a) pre-objective scan and (b) post-objective scan.

obtain a single optical section, the focal plane must be raster scanned in both lateral directions. By the Nyquist criterion, the object should be sampled with at least twice the highest frequency of the object to avoid aliasing artifacts.

The stage scan, objective lens scan, and beam scan have been developed to take 2D and 3D confocal images. The beam path is fixed for the stage scan and the objective lens scan; the stage and objective lens are moved relative to each other. In the beam scan, the objective lens and the stage are stationary; the scan mirrors scan the beam angularly at the entrance pupil of the objective lens. To increase the imaging speed, line scan and multiple-point scan methods have also been developed that enable parallel image acquisition. Theoretically, the stage scanning and beam scanning methods can produce similar images, but the requirements on the objective lenses for optimum performance are significantly different.

7.3.1 Stage scan

Early confocal microscopes were based on the stage scan. Figure 7.9 shows a diagram of a stage-scanning confocal imaging system. The object is scanned with respect to the stationary focal point of the objective lens. The major advantage is that the objective lens always works on axis. It simplifies the aberration correction and provides an even optical response across the entire scanned field. All off-axis aberrations, namely coma, astigmatism, field curvature, distortion, and lateral chromatic aberration, are absent. Only spherical aberration and axial

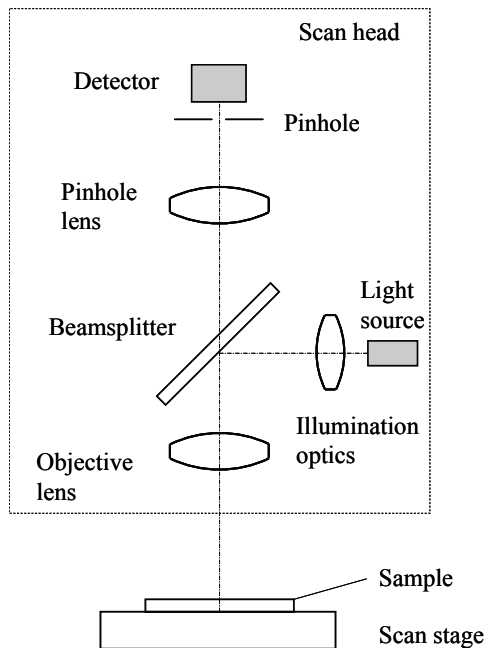
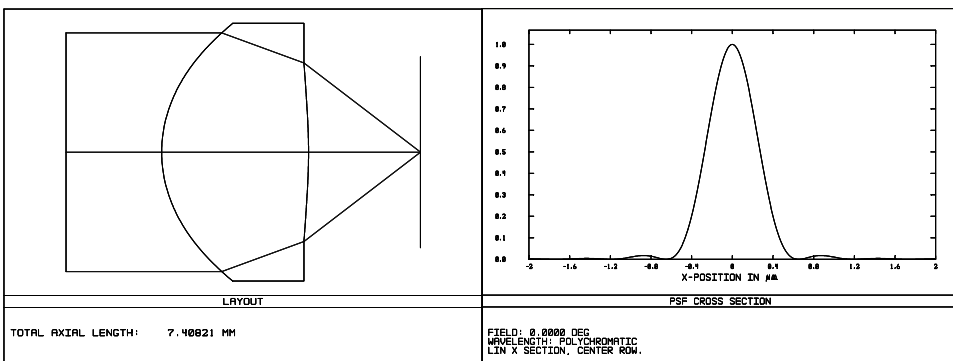


Figure 7.9 Diagram of a stage-scanning confocal imaging system.

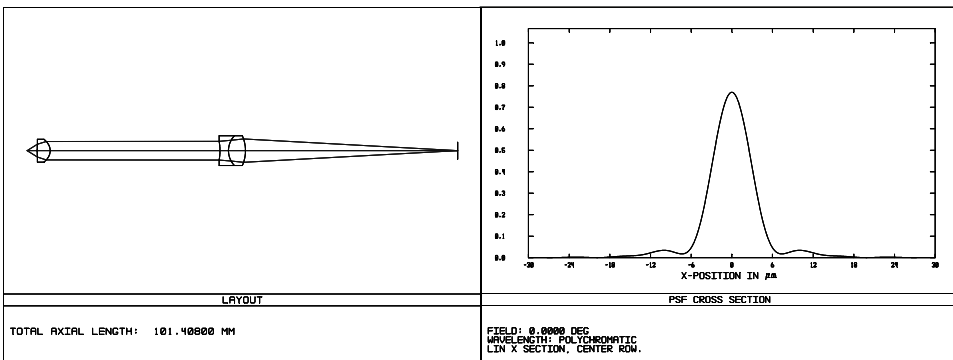
chromatic aberration remain to be corrected; therefore, the objective lens can be much simpler. Another advantage is that the FOV is only limited by the scan range of the stage, not by the imaging optics. The major issue of the stage scan is its low scanning speed. Other limitations include the stability of the scanning stage and the stability of the object, especially with living specimens.

Figure 7.10(a) shows an objective lens used in a confocal reader for biochip screening, which has an excitation wavelength of 650 nm and an emission peak of 670 nm. Since it is an on-axis scanning system, a single aspherical lens is adequate. The PSF plot shows that the FWHM of the focus spot is 0.6 μm . The spot size can be increased by reducing the beam size using a zoom beam expander or an aperture.

The emitted light with the peak of 670 nm is not collimated by the objective lens and is slightly divergent after leaving the lens. Since the emitted spectrum ranges from 0.64 to 0.70 μm and the objective lens does not correct the chromatic aberration, a doublet focusing lens is needed to correct the chromatic



(a)



(b)

Figure 7.10 Optical system for a stage scanning confocal imaging system. (a) The objective lens and its performance and (b) the detection path using a doublet focusing lens with a focal length of 50 mm.

aberration. Figure 7.10(b) shows the layout and the performance of a detection path when a doublet focusing lens with a focal length of 50 mm is used to focus the light onto the pinhole in front of the PMT. The radius of the Airy disk is $7.4\ \mu\text{m}$, so the optimal pinhole diameter can be $\sim 15\ \mu\text{m}$. As discussed in the previous section, a telephoto pinhole lens with a longer focal length can be used to enlarge the focal spot and keep the total track short.

A stage scan is a good choice for an experimental setup to prove concepts because it allows optical systems to be modified easily for different imaging configurations. It is also a good choice when the scan region is large, for example, in the confocal reader for a biochip discussed above.^{17,18} High accuracy, resolution, and vibration stability are required for scan stages. Typically, the stage resolution should be $0.1\ \mu\text{m}$ or better.

Objective lens scanning, also shown in Fig. 7.9, is equivalent to stage scanning. The objective lens is moved across the object, while the stage is stationary. This method has the same advantages as stage scanning, but the stability requirement on the object is less critical. Usually, the light source can be part of the scan head and moves together with the objective lens. Light can also be relayed to the objective lens through an optical fiber.

7.3.2 Beam scan

7.3.2.1 Introduction

The most straightforward form of beam scan is to raster scan the pinhole, as shown in Fig. 7.11(a), or to scan the fiber when a fiber is used. One benefit of beam scan is that the mechanical tolerance for the pinhole scanning system is less critical than that for the stage or objective lens scanning because the scan is demagnified by the objective lens. To maintain uniform illumination and light collection efficiency, a double telecentric objective lens is preferred. For a fiber scanning system, the requirements on the objective lens depend on the scanning methods, which will be discussed in Sec. 7.5. Another type of pinhole scanning system uses a large number of pinholes in a rotating disk called a Nipkow disk, instead of using a single pinhole.

More commonly, the beam scanning method uses one or two mirrors to scan the beam. Figure 7.11(b) illustrates post-objective scanning, where the scan mirror is located between the objective lens and the object. This configuration demands an objective lens that has a long focal length to accommodate for the scan mirror, limiting the working distance and the NA. The advantage of this scanning method is that the objective lens works only on axis. The major issue is that the focused spot is not on a flat surface as the scan mirror rotates. Another issue is that the chief ray angle varies across the field. As a result, the magnification is not constant for the points with the same lateral position but at different depths.

A pre-objective scan, where the scan mirror is in front of the objective lens, as shown in Fig. 7.11(c), is the most commonly used beam-scanning method.

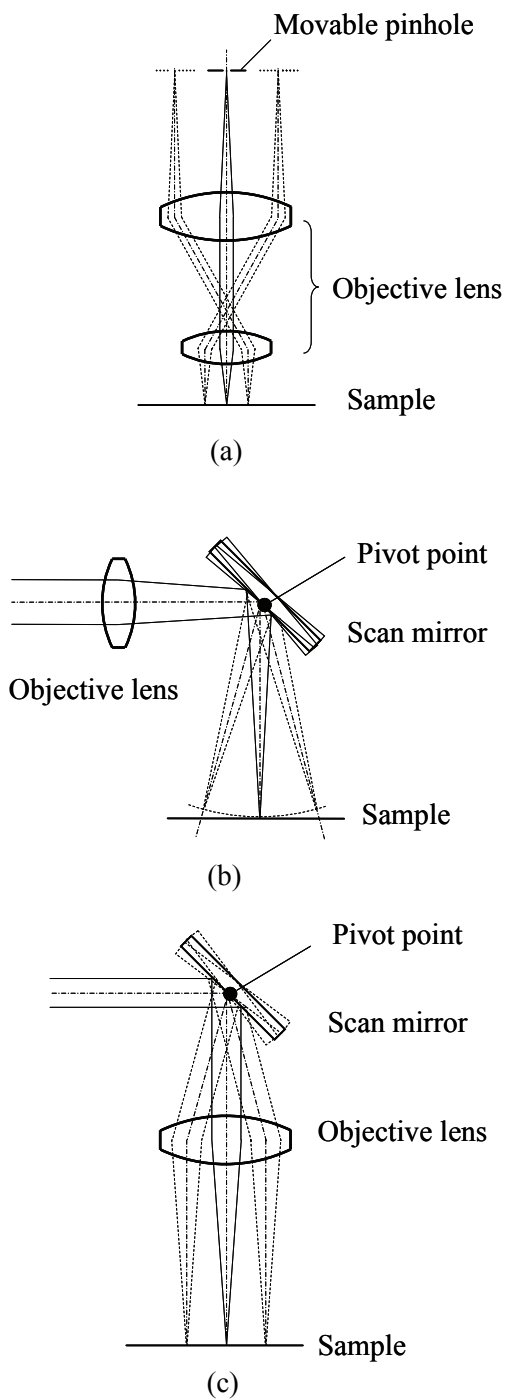


Figure 7.11 Configurations of beam-scanning systems: (a) pinhole scan, (b) post-objective scan, and (c) pre-objective scan.

Because there are no components, except for a possible folding mirror, between the objective lens and the object, the NA can be larger, and the working distance can be shorter.

In a pre-objective scanning configuration, the beam is first scanned into an angular field and then is usually focused onto a flat surface. Ideally, the scanning element should be located at the entrance pupil of the objective lens to achieve telecentricity in object space. The required clearance from the scanner to the objective lens is determined by the entrance pupil diameter, the input beam geometry, and the angle of the scanned field. The complexity of the objective lens depends on the optical requirements over a finite scanned field, including spot size, scan linearity, field curvature, and depth of focus.

Most confocal imaging systems for biological applications use a beam-scanning system to deflect the illumination beam through the focusing of the objective lens across the object in a raster pattern. This scanning method requires a fast line scan and a slow frame scan to obtain a 2D image. The frame scan offsets the fast scanning line to sequential positions from top to bottom of the frame.

7.3.2.2 Beam scan with a Nipkow disk

In 1884, P. Nipkow invented a technique to transfer a 2D image into a temporal electrical signal over a single cable using a scanning disk with a series of rectangularly perforated holes. In the late 1960s, M. Petráň and M. Hadravský pioneered this technique in confocal imaging by introducing a tandem-scanning reflected-light microscope (TSRLM) with a two-sided Nipkow disk, as shown in Fig. 7.12(a).¹⁹ The light illuminates one side of the Nipkow disk, and the image is observed through a conjugate set of pinholes on the opposite side of the disk. The disk is positioned in conjugate relation to a specimen and is rotated at a high speed to scan the object. The illumination and reflection lights are separated through a beamsplitter located between the disk and the objective lens. Unlike the conventional stage scanning and objective lens scanning systems, both the objective lens and the stage in the Nipkow disk scanning microscope are stationary; the object is scanned by rotating the Nipkow disk.

The major advantage of the Nipkow disk scanner is that the speed of the raster scan is not limited by the mass of the stage and the stability of the object; it is only limited by the speed and the efficiency of the scanning disk. Therefore, the object can be directly observed in real time through the eyepiece or the imaging detector. The advantage of the two-sided disk design is that light reflected from the top of the disk is isolated from the observation path and will not cause glare; the disadvantage is that adjusting and maintaining the alignment of the microscope is difficult.

To address this issue, a one-sided disk version was proposed as early as the 1960s.²⁰ Figure 7.12(b) shows a practical version of a one-sided Nipkow disk confocal microscope.²¹ The illumination and reflection light paths are separated by a beamsplitter in front of the Nipkow disk. By tilting the disk, the light reflected from the top of the disk is deflected away from the imaging path. The

reflected light from the disk can also be eliminated by using a polarization beamsplitter.

Both lasers and arc lamps have been used in Nipkow disk scanning microscopes. Arc lamps are preferred over lasers because coherent interference between the light reflected through adjacent pinholes in the disk can cause lines to appear in the image. Also, the speckle due to the coherent interference between different layers of the object can make the image difficult to interpret and quantify. Another feature of broadband illumination is that it generates color images directly. The Köhler illumination system is typically designed to uniformly illuminate the Nipkow disk.

The optimum size and spacing of the pinholes depend on the focal length of the tube lens and the pupil size of the objective lens. As a general rule, the diameter of the pinhole should be roughly equal to the FWHM of the Airy pattern projected onto the disk by the objective lens and tube lens, and the center-to-center spacing should be enough to minimize crosstalk of the light through neighboring pinholes, especially light from out-of-focus planes.³

After passing through the pinholes, the light fills the entrance pupil of the objective lens via the tube lens and the field lens. Generally, a doublet is optically sufficient for the field lens to achieve telecentricity. A tube lens is required for an infinity-corrected objective lens.

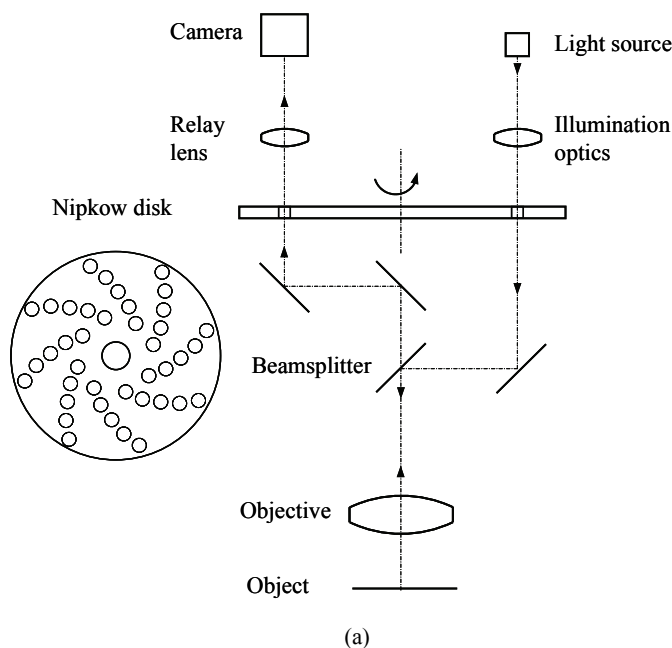
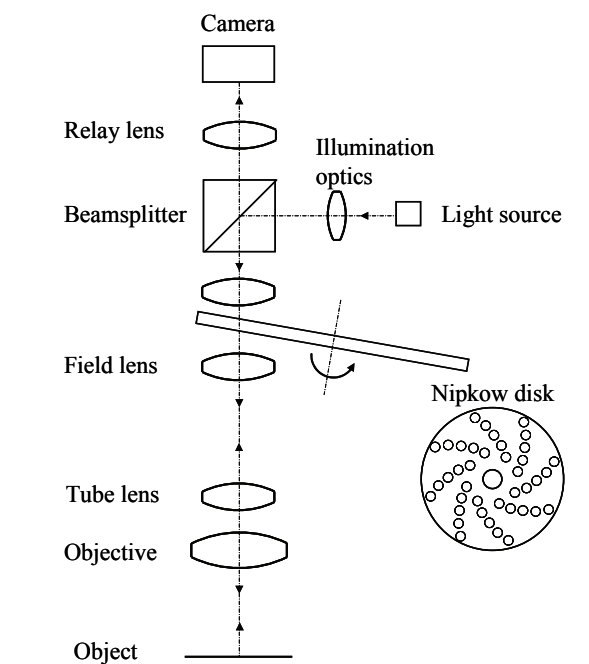
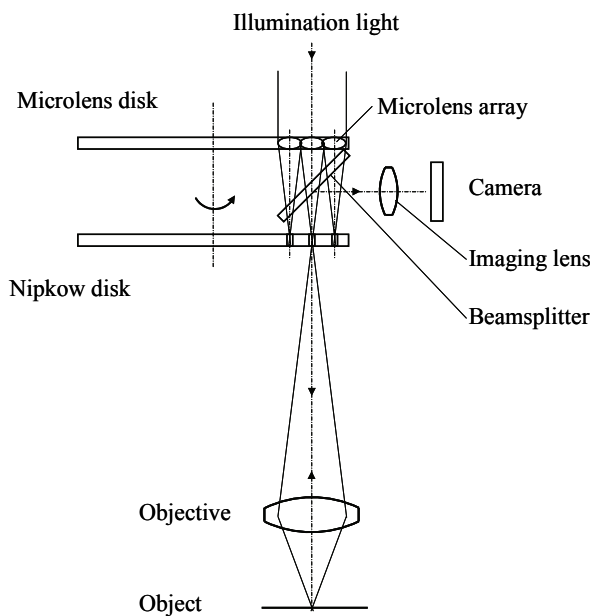


Figure 7.12 (a) Two-sided Nipkow disk confocal microscope: tandem-scanning reflected-light microscope (TSRLM). [Continued on next page.]



(b)



(c)

Figure 7.12 (continued) (b) One-sided Nipkow disk confocal microscope and (c) microlens Nipkow scanning confocal microscope.

The objective lens is one of the most critical elements in Nipkow disk confocal microscopes. The requirement of aberration correction is similar to that for the objective lenses used in other beam scanning systems. Ideally, the field lens, tube lens, and objective lens should be designed or selected so that the aberrations of the entire optical train are minimized. This can be achieved either by correcting the aberrations of each component individually or by optimizing the entire optical system as a whole.

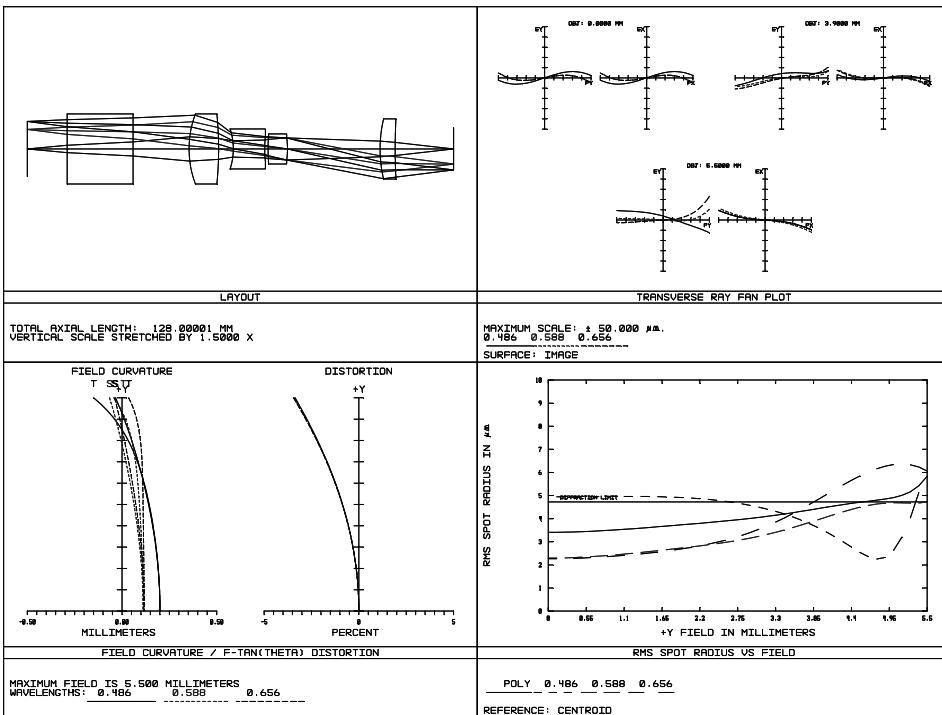
The light reflected off the object travels back through the objective lens, the tube lens, and field lens to the same pinhole in the Nipkow disk. After passing through the pinholes, the light is focused onto the image sensor through the relay lens and the beamsplitter.

One drawback of the Nipkow disk microscopes discussed above is the low illumination efficiency. The disk needs to be illuminated uniformly; however, only a small percentage of the light passes through the pinholes on the disk and is focused onto the object. The efficiency can be increased by increasing the number of pinholes in order to maintain high axial discrimination ability, but only to some extent. To address this issue, several methods have been proposed and implemented. For example, a kinoform-based Nipkow disk scanner using a computer-generated phase mask has been used to transfer almost all of the incident light into a predesigned Nipkow disk pattern.²² A more effective method is to use a microlens array, which focuses the light onto the pinholes, as shown in Fig. 7.12(c).²³ Each microlens focuses the light onto its corresponding pinhole, thus significantly increasing the fraction of the illumination light that is transmitted through the pinhole. Compared to the $\sim 2\%$ light efficiency for the configuration in Fig. 7.12(b), this method allows 40% of the illumination light to pass through the pinholes and fill the aperture of the objective lens.

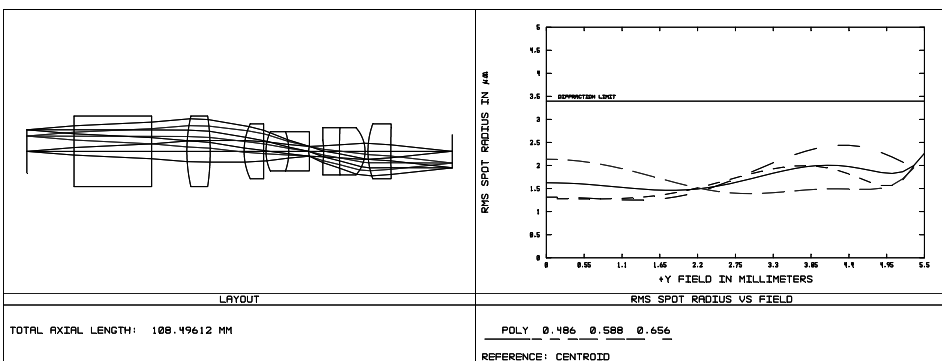
For all three configurations, the objective lens, together with the tube and field lenses, focuses the light onto the Nipkow disk telecentrically. One of the requirements of the relay lens, which focuses the light from the pinholes to the camera, is the telecentricity in the pinhole side so that all of the light passing through the pinholes can be imaged onto the sensor. Telecentricity on the camera side is not essential, but it is desired, to maintain the uniform quantum efficiency. Other requirements include a short total track and enough distance from the disk to the first surface of the relay lens to accommodate the beamsplitter.

Figure 7.13(a) shows one example of a relay lens used in a Nipkow confocal microscope in the visible range. It is a double telecentric design with an NA of 0.06 on the Nipkow disk side and a total track of 128 mm. It consists of three positive singlets and one negative singlet. The singlets close to the disk and camera are field lenses that achieve a telecentric condition. This relay lens has reasonable performance over the entire FOV; however, there is some remaining spherical aberration, coma, field curvature, and distortion. One limitation of this relay lens is that the NA is relatively low.

Figure 7.13(b) shows another double telecentric design with an NA of 0.08 and a shorter total track of 108.5 mm. It is a double Gauss relay lens with two



(a)



(b)

Figure 7.13 Design examples of relay lenses used in a Nipkow disk confocal microscope: (a) double telecentric relay lens with three positive and one negative singlets and (b) double Gauss relay lens with two doublets and three positive singlets.

doublets and three singlets. The odd aberrations, namely, coma, distortion, and lateral chromatic aberration, are better corrected as a result of the symmetrical layout. Axial chromatic aberration is also smaller with the contribution from the two doublets. Spherical aberration and field curvature are the two major remaining aberrations. This configuration has the potential for a higher NA.

7.3.2.3 Beam scan with a spatial light modulator

Besides the Nipkow disk, designed with many pinholes to obtain confocal images in high speed, a spatial light modulator is also capable of generating thousands of pinholes for confocal imaging.²⁴ Figure 7.14(a) shows a confocal configuration using a spatial light modulator, such as a digital micromirror device (DMD). A DMD consists of hundreds of thousands of switchable micromirrors that are controlled by underlying CMOS electronics, as shown in Fig. 7.14(b).²⁵ The mirrors are highly reflective and are used to modulate light. Each mirror is capable of switching a pixel of light by tilting between two states, +12 deg for the “on” state or -12 deg for the “off” state. When the mirror is not operating, it sits in a “parked” state at 0 deg.

DMDs make it possible to change observation conditions easily and avoid unwanted vibration caused by the rotating disk. Each mirror in a DMD serves as a pinhole for both illumination and detection in a predefined spatial and temporal pattern. The illumination pinhole is created by turning on one micromirror while the surrounding micromirrors remain in the off state. Thus, only the light reflected from the one micromirror will pass through the optical system. The objective lens images the on pixel onto the object. Returning light or fluorescence from the object at the focused point is then focused back onto the same micromirror. A CCD or CMOS camera is then used to acquire one integrated, continuous image as the DMD performs the scanning.

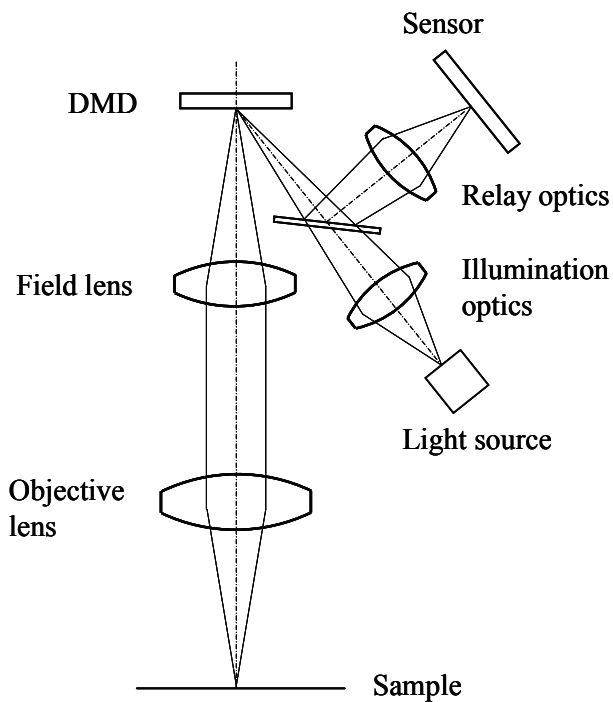
To capture the entire FOV, a mosaic is configured by shifting the on and off mirrors in a time-varying pattern that covers the entire FOV. It is also possible to scan the object randomly with multiply focused spots simultaneously.

7.3.2.4 Beam scan with scan mirrors

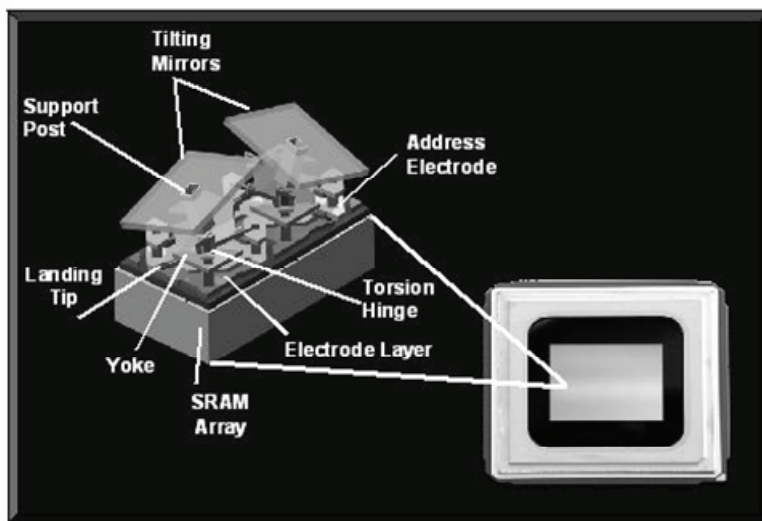
An efficient beam scanning system with scan mirrors should have the following characteristics: a good duty cycle, stationary pivot point, high scan speed, and the capability of scan rotation.²⁶ A good duty cycle requires that all nonproductive scanning be minimized. Typically, the focus spot is raster scanned across the object. The focused spot is scanned quickly in one direction (line scan) and slowly in the orthogonal direction (frame scan).

For all scanning systems, a stationary pivot point is required and should be at the entrance pupil or imaged at the entrance pupil plane of the objective lens. With a stationary pivot point at the entrance pupil plane or its conjugate plane, the beam at the entrance pupil remains constantly in position and purely rotates, achieving the smallest focused spot it can achieve and maintaining a constant illumination. When the beam scans at the pupil plane, the objective lens works in a telecentric mode and converts the change of the scan angle at the pupil plane to the movement of the focused spot in the object plane.

A high scanning speed is another essential requirement for a confocal scanning system, especially for *in vivo* applications. The inertia of the scanning components should be minimized to achieve a high scanning speed. Minimizing



(a)



(b)

Figure 7.14 (a) Schematic diagram of a confocal imaging system with a DMD as the scanner. (b) Schematic of two DMD mirror pixels in a typical DMD light modulator.²⁵

the nonproductive scan also reduces unnecessary photon damage that may be caused by a poor duty cycle.

The scanning direction should be rotatable to obtain the best possible image quality and time resolution with respect to the specimen shape or other characteristics. For example, the time resolution of the signal from muscle fiber is enhanced when the fast scan is parallel to the long axis of the fiber.

For beam scanning systems with scan mirrors, the light from the object must be descanned so that it does not move relative to the pinhole. This is automatically achieved by reflecting a second time from the scan mirrors when the beamsplitter is placed between the first scanner and the illumination pinhole.

A simple beam scanning configuration places the scanner at the entrance pupil plane of the objective lens, as shown in Fig. 7.11(c). This configuration requires that the entrance pupil of the objective lens be some distance outside of the lens to accommodate the scan mirror. This requirement poses some challenges to aberration correction, especially for a telecentric objective lens with a high NA.

When the entrance pupil is close to or inside the objective lens, the pivot point of the scan mirror can be relayed to the entrance pupil, as shown in Fig. 7.15. This configuration has no special restriction on the scanner. Therefore, a more robust, stable, and high-speed scanner can be used.

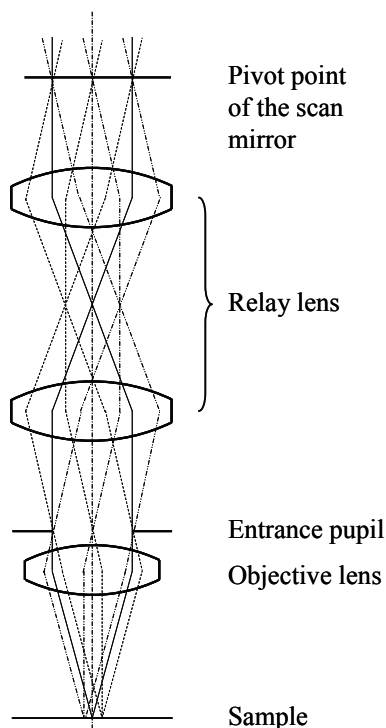


Figure 7.15 A beam scanning system with relay lens. The beam rotates around the pivot point, which is relayed to the entrance pupil of the objective lens.

To obtain 2D or 3D images, the two-axis scanner, such as the fast steering mirror and MEMS mirror, is ideal for scanning the focused illumination spot across the object. However, the scanning speed is typically slow for a two-axis scanner with reasonable mirror size. Another configuration has one small galvanometer head mounted on a second larger scanner whose rotation axis is perpendicular to the first galvanometer axis; the two rotating axes cross at the mirror center. The small galvanometer with low inertia provides a fast line scan, and the larger scanner scans more slowly for a frame scan in the orthogonal direction. The major disadvantage of this configuration is the huge inertia in one of the rotation axes; therefore, the scanning speed is slow.

A typical scanning system utilizes two high-speed scan mirrors driven by galvanometer motors or a combination of a galvanometer scanner and an acousto-optical deflector or a polygon scanner. The two scanners, mounted on mutually perpendicular axes, are either close coupled or placed in conjugated telecentric planes of a relay system and scan in orthogonal directions, producing a rectilinear raster scan.

Figure 7.16(a) shows the configuration of a close-coupled, two-mirror scanning system. The two scan mirrors are positioned as closely as possible, with the entrance pupil of the objective lens or its conjugate plane located at the midpoint between them. One scan mirror oscillates quickly and produces the line scan, and the other produces the frame scan perpendicular to it. This configuration is relatively simple and compact. However, it only achieves approximate telecentricity because the telecentric plane is between the two scan mirrors. The closer the two mirrors, the smaller the telecentric error. Figure 7.16(b) shows the telecentricity of the objective lens when the first mirror scans. The plot is calculated for an objective lens with a focal length of 10 mm. The scan angle is 3 deg, and the conjugated pupil plane is 5 mm from the first scan mirror. The chief ray angle is approximately linear with the FOV.

Another issue with close-coupled mirrors is the beam shift on the pupil plane. While the mirrors scan, the beam moves around the pupil plane instead of being stationary. Therefore, the intensity of the image varies across the field; the resolution and contrast are reduced because only part of the pupil is effectively used for illumination and detection while scanning. The pupil can still be fully filled by scanning a large-sized beam, with some light loss during the scanning process.

To address the issues in the close-coupled configuration, an improved version of a close-coupled scanner, called a paddle scanner, was developed.¹¹ The first scan mirror, used as the slow scanner, rotates around an axis that is remote from the point of reflection, resulting in a beam shift. The position of the beam is almost constant on the second and fast scanner. The remaining issue is that successive scan lines in the image frame have slightly different optical path lengths, limiting its usage in applications that require high image quality. The properties of a close-coupled scanner and paddle scanner are discussed in detail in Ref. 11.

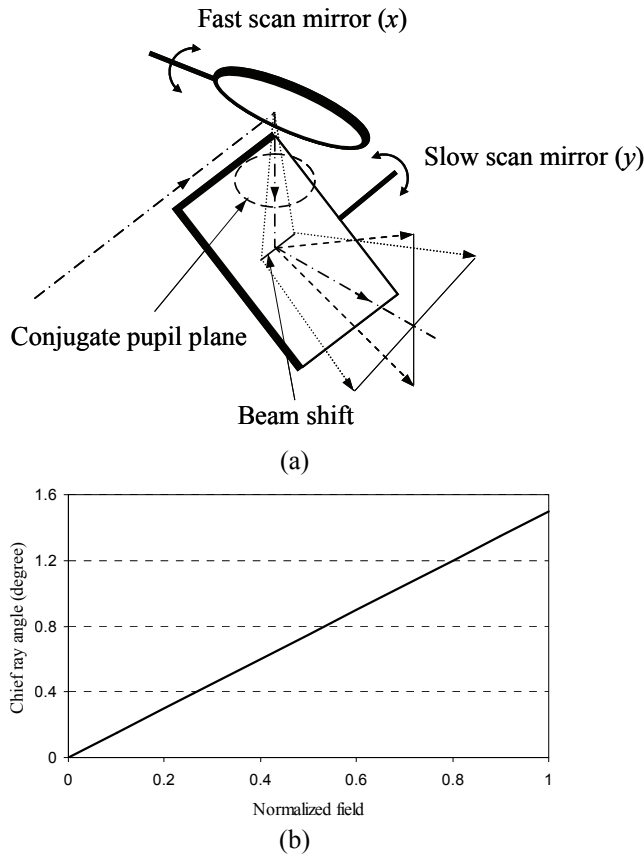


Figure 7.16 (a) Scanning system with two close-coupled mirrors. The rotation of the first scan mirror introduces the beam shift on the second mirror, resulting in a loss of light or an unfilled pupil. (b) The chief ray angle of the objective lens when the first mirror scans. The focal length of the objective lens is 10 mm. The scan angle is 3 deg, and the conjugate pupil plane is 5 mm from the first scan mirror.

The most commonly used configuration with two scan mirrors has the mirrors couple with the relay optics, as shown in Fig. 7.17. Similar to the close-coupled configuration, the two mirrors rotate on perpendicular axes to produce a raster scan. The key difference is that a relay system is employed to image one mirror onto the other so that both mirrors can be exactly at the conjugate planes of the entrance pupil of the objective lens. The beam position is always fixed at both scanning planes because of the conjugate condition; therefore, there is no beam movement at the entrance pupil during the scanning process. This configuration meets all of the requirements of a good scanning system, as previously discussed: good duty cycle, stationary pivot point, high speed, and rotatable scanning direction.

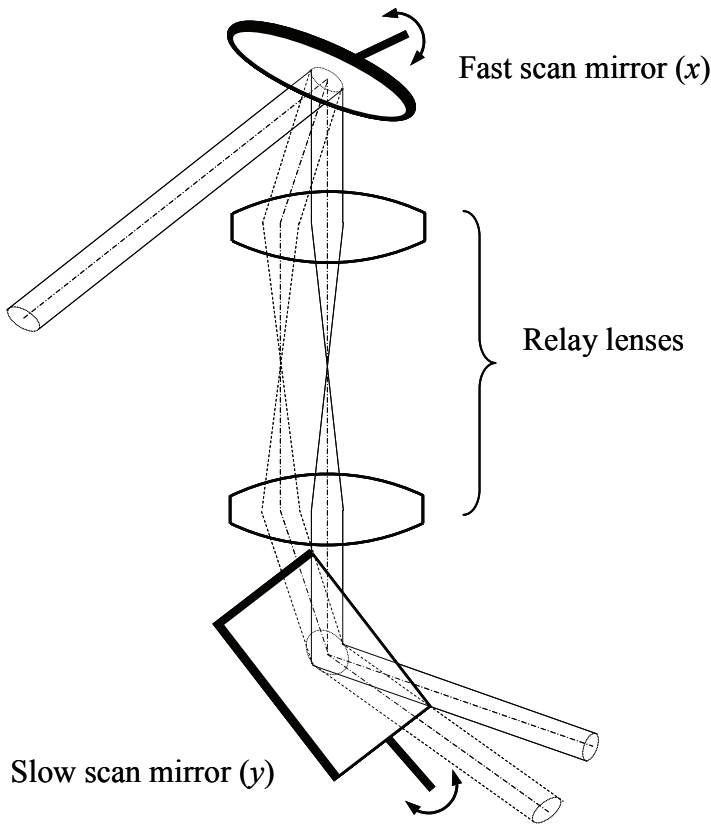


Figure 7.17 A scanning system with refractive relay lenses.

7.3.3 Relay systems

In beam scanning confocal imaging systems, another relay system is usually needed to image the scan mirror onto the entrance pupil of the objective lens in addition to the relay system that optically couples two scan mirrors, as shown in Fig. 7.15. The function of the relay system is twofold: it optically conjugates the scan mirrors and the entrance pupil of the objective lens, and it adjusts the beam size to match the size of the mirror or the diameter of the entrance pupil. The basic requirements of a relay system in confocal imaging are that the aberrations are well controlled and the relay lens is telecentric.

The most straightforward relay system is a $4f$ system consisting of two identical doublets assembled afocally. The distance between the two lenses is the sum of the two focal lengths. The first scan mirror is located at the front focal plane of the first lens, and the second scan mirror is at the rear focal plane of the second lens. The collimated beam with different scan angles is focused by the first lens telecentrically at the common focal plane of the two lenses, and the light is collimated again by the second lens and intersects at the rear focal plane

of the second lens, where the second scan mirror is located. Therefore, the beam scanning achieved by the first scan mirror in one direction is relayed to the second scan mirror, which scans the beam again in the orthogonal direction.

The relay lens does not have to be unitary; the focal lengths can be optimized to match the sizes of the scan mirrors or the entrance pupil. The advantage of unit magnification is that the odd aberrations, such as coma, distortion, and lateral chromatic aberration, are canceled automatically.

Figure 7.18 illustrates a relay system with two identical doublets. The focal length is 50 mm, and the beam size is 5 mm in diameter. To simplify the system analysis, an ideal scanning objective lens with a 5-mm entrance pupil diameter and a 5-mm focal length is attached to the relay system. As the performance plots show, it has a diffraction-limited performance over a ± 5 -deg scan angle. The odd aberrations—coma, distortion, and lateral chromatic aberration—are canceled out because of the symmetrical configuration.

Axial chromatic aberration can be an issue in refractive relay systems. As shown in Fig. 7.18, the focal points of 0.486 and 0.656 μm are separated by approximately 1.3 μm . This distance is acceptable for a confocal objective lens

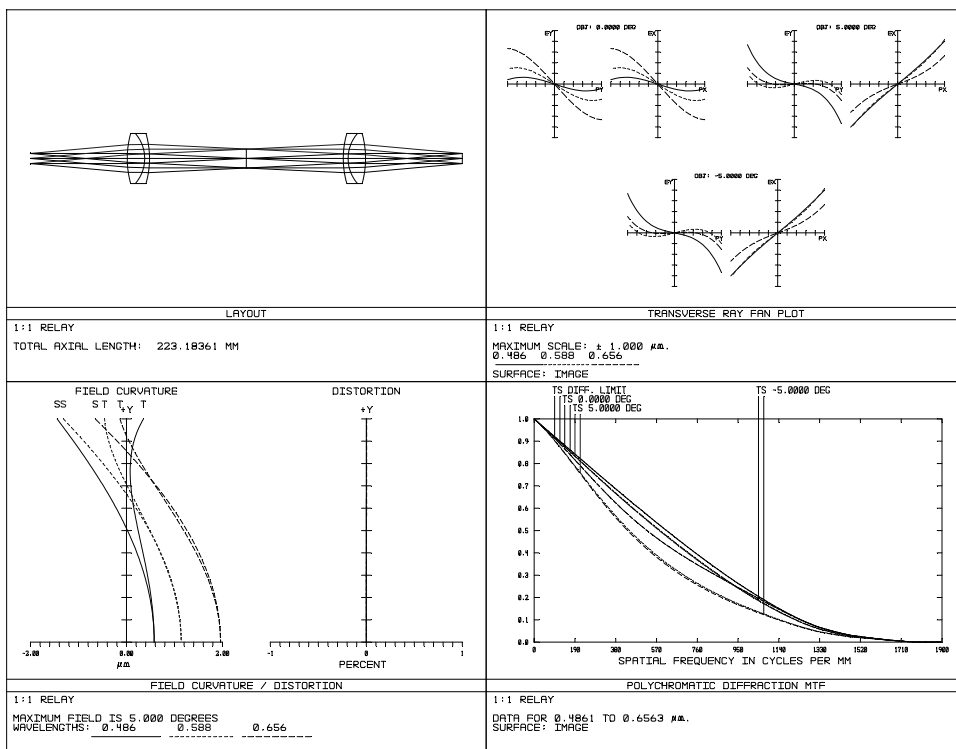


Figure 7.18 Telecentric refractive relay lens with two identical doublets and its performance (ray fan plot, field curvature/distortion, and MTF).

with a 0.5 NA because this lens has an axial resolution of 3 μm . When the spectrum extends to the UV and near-IR, longitudinal chromatic aberration will become a critical issue.

Astigmatism and field curvature of a relay lens depend quadratically on the scan angle or field size, limiting the off-axis performance. Figure 7.19(a) shows an RMS wavefront error as a function of the scan angle when the focal length of the doublet is 50 mm. The solid line is the RMS wavefront error of a diffraction-limited system. The system has diffraction-limited performance up to +5 deg. The RMS wavefront error grows quickly as the scan angle increases.

The performance of the refractive relay system strongly depends on the focal lengths of the two doublets. The larger the focal length is, the better the performance the relay system can achieve. Figure 7.19(b) plots the RMS wavefront errors of the relay systems with different focal lengths, but the same entrance pupil diameter and scanning angle. The thick straight solid line is the RMS wavefront error of a diffraction-limited system. As shown in the figure, when the focal length becomes shorter, the performance of the relay system drops significantly. The quality degradation is caused by off-axis aberrations, mainly, field curvature and astigmatism. The performance of the refractive relay system also strongly depends on the beam diameter since all aberrations are pupil dependent. With the same relay lens, the performance drops rapidly as the beam diameter increases.

To address the chromatic aberrations inherent in refractive relay lenses, reflective relay lenses have been developed.^{27,28} Given that the optical properties of a reflective surface only depend on the radius of the mirror, chromatic aberrations disappear.

Figure 7.20 is a relay configuration consisting of one concave mirror and one flat mirror. The flat mirror is located at the focal plane of the concave mirror. Two scan mirrors are located at the focal plane but are shifted laterally and

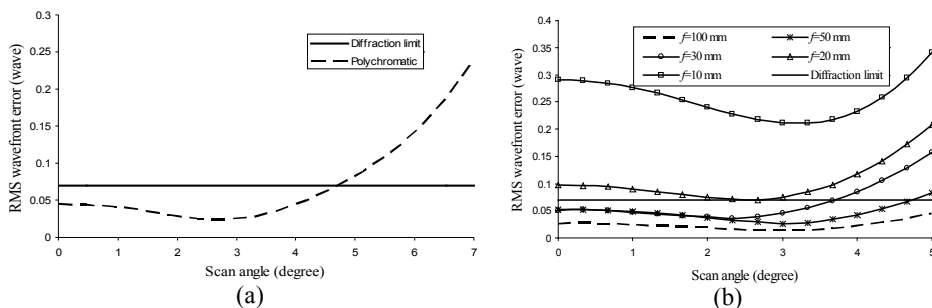


Figure 7.19 Performance of refractive relay lenses in Fig. 7.18(a). (a) The RMS wavefront error as a function of scan angle and (b) RMS wavefront error as a function of scan angle for different focal lengths of the relay lens. Here f is the focal length of the doublet.

symmetrically from the focal point. While scanning, the collimated beam is focused by the concave mirror onto the flat mirror at the focal plane, and then it is collimated by the same concave mirror and reaches the second scan mirror.

The first scan mirror can scan either horizontally or vertically, but usually, it is configured to scan horizontally for larger FOV and better image quality. When the first mirror scans vertically, the FOV is limited by the distance between the two scan mirrors, as shown in Fig. 7.21(a). With a 200-mm-radius and a 60-mm-diameter concave mirror, a 6.5-deg scan angle, and a 5-mm pupil diameter, the pupil height H , which is the distance between the first scan mirror and the optical axis of the spherical mirror, has to be larger than 15 mm so that no light will be blocked. The performance is limited by astigmatism, as clearly shown in Fig. 7.21(b); the tangential MTF is much smaller than the sagittal MTF.

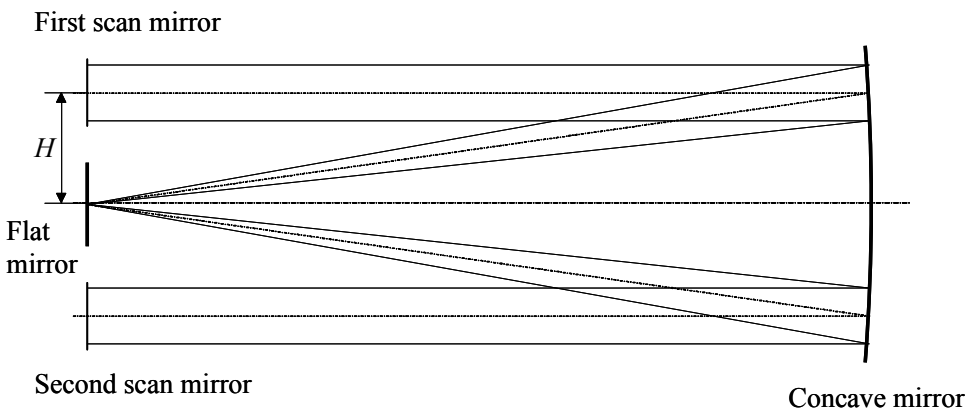


Figure 7.20 Relay system with one concave mirror and one flat mirror.

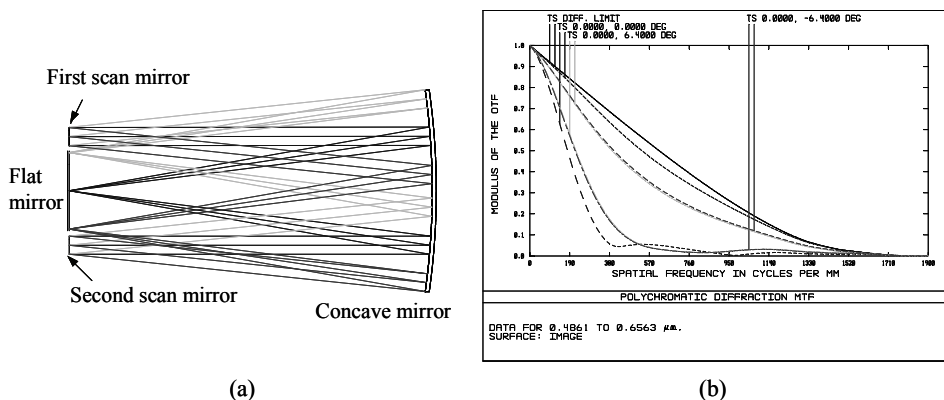


Figure 7.21 (a) Configuration of a reflective relay system with one concave mirror and one flat mirror. The first scan mirror scans the beam vertically. (b) The MTF of the system when the reflective relay system is used with a perfect objective lens. A perfect objective lens has a focal length of 5 mm and a pupil diameter of 5 mm.

The performance of the relay system in Fig. 7.21(a) strongly depends on pupil height H . Figure 7.22(a) shows the RMS wavefront error as a function of the scan angle for a different pupil height H . The RMS wavefront error increases quickly as the pupil height H increases. Like the refractive relay lens, the focal length of the reflector, which is half the radius of the concave mirror, has significant impact on the performance of the relay system. The RMS wavefront error at full field as a function of pupil height for the different radii of the spherical reflector is shown in Fig. 7.22(b). As the radius decreases, the RMS wavefront error increases rapidly.

When the first scan mirror scans the collimated beam horizontally, as shown in Fig. 7.23(a), the FOV is only limited by the field curvature and astigmatism as long as the diameter of the concave mirror is large enough. Another advantage of a horizontal scan is that the pupil height H can be much smaller than that in Fig. 7.21(a) for the same scan angle. Compared with the MTF in Fig. 7.21(b), the MTF in Fig. 7.23(b) is much better over the entire FOV. This relay system has a diffraction-limited performance up to 8 deg.

The concave mirror in Figs. 7.21 and 7.23 has a dual function: first, it focuses the collimated light onto the flat mirror, and then, it collimates the same light reflected from the flat mirror. This mirror can be replaced with two spherical mirrors, as shown in Fig. 7.24. In Fig. 7.24(a), the two mirrors are facing each other, and their optical axes overlap. Optically, it is equivalent to the configuration in Fig. 7.23(a). In Fig. 7.24(b), both spherical mirrors tilt slightly relative to the incoming beam; the two optical axes do not overlap but are shifted relative to each other.

The collimated beam is scanned across the first concave mirror by the first scan mirror and then is focused to a line at the focal plane of the concave mirror. The second concave mirror collimates the beam and directs the collimated beam to the second scan mirror. The effective scan is a pure rotation about a point on

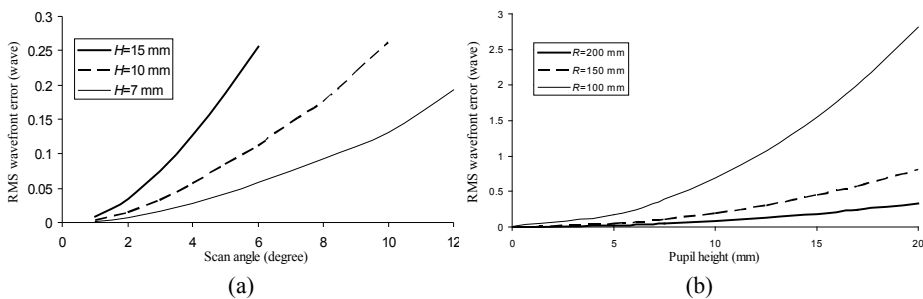


Figure 7.22 Dependence of the RMS wavefront error of the reflective relay system in Fig. 7.21(a) on the pupil height H and the radius of the spherical mirror. (a) RMS wavefront error as a function of the scan angle for different pupil height H and (b) RMS wavefront error at full field as a function of pupil height for spherical mirrors with different radius R .

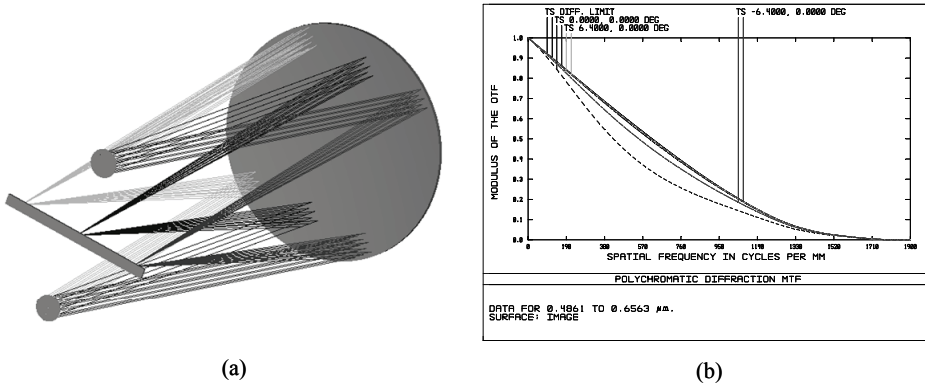


Figure 7.23 (a) Configuration of a reflective relay system with one concave mirror and one flat mirror. The first scan mirror scans the beam horizontally. (b) The MTF of the system when the reflective relay system is used with a perfect objective lens. A perfect objective lens has a focal length of 5 mm and a pupil diameter of 5 mm.

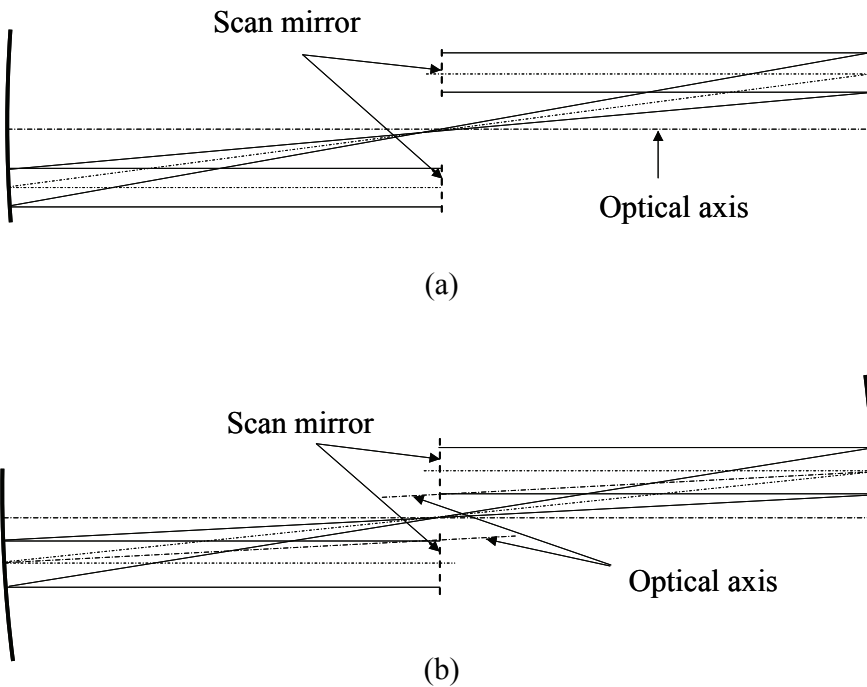


Figure 7.24 Configurations of a relay system with two concave mirrors. (a) Two mirrors share a common optical axis, and (b) the optical axes of two mirrors are parallel but shifted.

the second scan mirror without beam shift. One advantage compared to the configurations in Fig. 7.23 is that the beam size can be adjusted to match the size of the second scan mirror or the pupil diameter of the objective lens by using concave mirrors with different radii.

The difference in performance of the two configurations in Fig. 7.24 is clearly shown in Fig. 7.25. Both configurations have an RMS spot size smaller than the Airy radius up to a certain scan angle. Figure 7.25(a) indicates that the on-axis performance is slightly worse than the off-axis performance for the configuration in Fig. 7.24(a) because of spherical aberration and astigmatism. In contrast, the off-axis performance is not as good as the on-axis performance for the relay system in Fig. 7.24(b). Figures 7.25(c) and 7.25(d) show the ray fan plot and the field curvature/distortion plot. Because of the symmetrical arrangement, odd aberrations are compensated. As shown in Fig. 7.25(c), curvature and astigmatism are the major aberrations limiting system performance. There is also

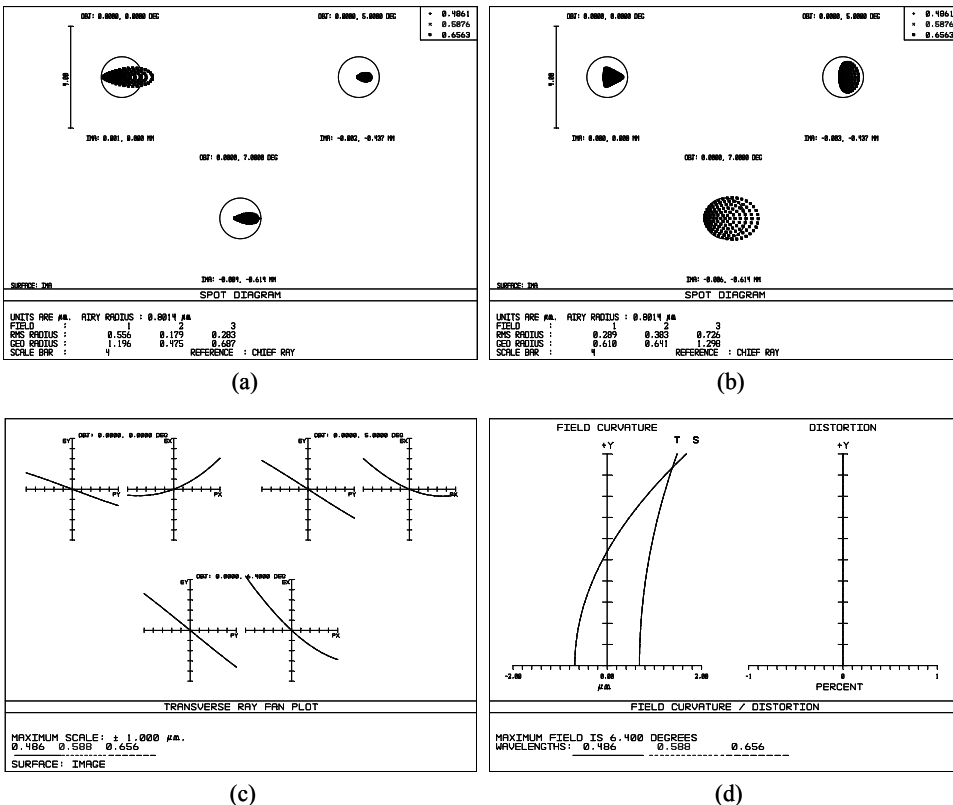


Figure 7.25 Performance of the reflective relay systems. (a) Spot diagram of the configuration in Fig. 7.24(a) and (b) the spot diagram, (c) ray fan, and (d) field curvature/distortion of the configuration in Fig. 7.24(b).

a small amount of spherical aberration, which can be further reduced by using an aspheric reflective mirror.

The Offner system, consisting of two concentric spherical mirrors, is another reflective relay system.²⁹ However, it is not commonly used for pupil relay since its optical performance does not meet the requirements for confocal imaging.

7.4 Optical Design of the Objective Lens

As discussed earlier, the objective lens in confocal imaging is the most critical component. It determines the image contrast, resolution, depth of field, and the lateral extent of the field. Except for a confocal imaging system with dual objective lenses (one for illumination and another for detection), the objective lens also serves as the illumination condenser. Therefore, in addition to the aberration correction at the detection wavelength, some additional requirements, such as aberration correction over a wide range of wavelengths and high transmission in the UV, are imposed onto the objective lens for confocal imaging.

7.4.1 Aberrations in the objective lens

All seven basic aberrations affect the performance of a confocal objective lens. Spherical and axial chromatic aberrations affect the entire field, while coma, astigmatism, field curvature, distortion, and lateral chromatic aberration are of significance to the beam scanning confocal system because they are field dependent. Aberrations in the objective lens cause a reduction of intensity at the pinhole and a reduction of depth discrimination. All seven aberrations must be minimized in order for a beam scanning confocal imaging system to work effectively.

Spherical aberration in the objective lens causes axial and peripheral rays to arrive at different points, blurring the focused point. It reduces the image intensity, contrast, and resolution. Generally, the objective lens is designed with the spherical aberration corrected for a specified distance relationship between the objective lens, object, and image planes. However, in practice, the objective lens may be used in conditions that deviate from the nominal condition, inducing spherical aberration. For a well-corrected dry objective lens, spherical aberration is the major limiting factor when it is used to image deep into a tissue whose refractive index is slightly larger than 1.33. Figure 7.26(a) shows the ray diagram in the object space for a diffraction-limited dry objective lens; all of the rays converge to the focal point. When this objective lens is used to image deep into a tissue, the refraction at the air/tissue interface introduces significant spherical aberration, as shown in Fig. 7.26(b).

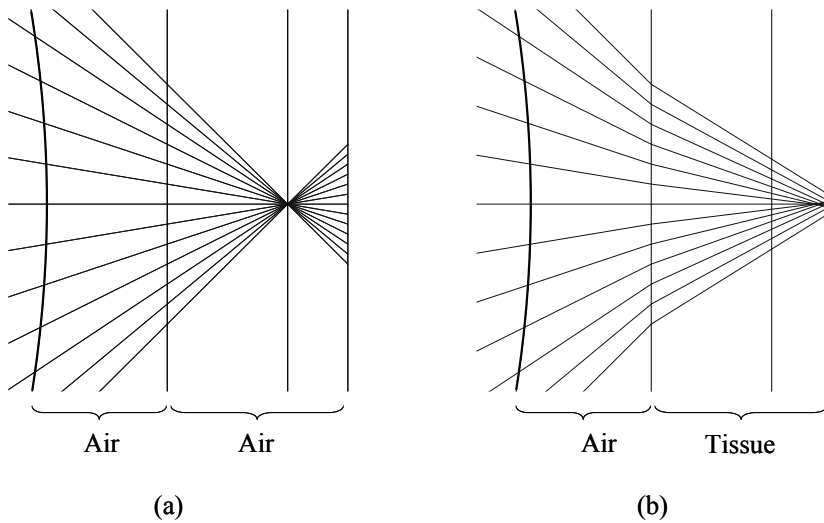


Figure 7.26 (a) Ray diagram in the object space of a diffraction-limited dry objective lens and (b) ray diagram when the dry objective lens is used to image deep into a tissue.

The objective lens should be designed for a specific medium in the object space. For example, a highly corrected water-immersion objective lens should be designed for aqueous specimens. However, there are always some variations in the cover-glass thickness and the refractive indices of immersion media and tissues.^{30–33} The refractive index of the tissue may also vary along the depth.³⁴ These small variations have profound influence on the amount of spherical aberration, even for a well-corrected water-immersion objective lens.³² To accommodate these changes, a movable optical element can be added to the objective lens to provide variable compensation for the spherical aberration induced by the refractive index variation in the immersion media, tissue, and cover glass, as well as the cover-glass thickness variation.

Coma manifests because of refraction differences by the light rays passing through the various lens zones as the incident angle increases. It is similar to spherical aberration, but it is encountered with off-axis light and is most severe when the optical system is out of alignment. With uncorrected coma, the objective lens produces a series of asymmetrical spots of increasing size, resulting in a comet-like focused spot.

Coma is often considered the most problematic aberration because of the asymmetry it produces in images; it reduces image intensity, sharpness, and contrast. Since coma is an off-axis aberration, it has significance in beam scanning confocal systems, which utilize off-axis beam paths, but is not an issue with confocal imaging systems working on axis.

Astigmatism is due to the different focal distances in the tangential and sagittal planes. The best focus is between the tangential and sagittal focal points. Astigmatism also reduces image intensity, sharpness, and contrast, with

increasing effects at greater distances from the optical axis. Astigmatism can result from design, poorly centered elements, and other misalignments of the lenses in the optical path.

Field curvature is a natural artifact of the lens with curved surfaces; it characterizes the natural tendency of optical systems to have curved image planes. An objective lens with field curvature images a flat specimen onto a curved surface. Therefore, the central and peripheral areas of the field cannot be simultaneously brought into sharp focus. Field curvature can be corrected optically with more-complex lens designs. For confocal biological applications where thick sections or live-cell preparations are imaged, the field curvature is relatively less significant because of the uncertainty of other factors about the exact tissue shape.

Distortion is contributed by the nonlinearity of magnification from the center to the edges of the image field. It does not affect the resolution but skews the true dimensions in the image. Similar to field curvature for confocal biological applications, the effect of a small amount of distortion is generally not significant. Small amounts of distortion can also be digitally corrected.

Chromatic aberration is caused by material dispersion, namely, the variation of the refractive indices of the optical materials with wavelength. Axial chromatic aberration is referred to as the difference of the focused points for different wavelengths along the optical axis. Lateral chromatic aberration is the difference in magnifications for different wavelengths caused by the variation of focal lengths with the wavelengths. The correction of axial chromatic aberration is usually accomplished in the objective lens design by a combination of multiple lens elements having different optical characteristics. Because the two types of chromatic aberration are related, objective lenses highly corrected for axial chromatic aberration generally have minimal lateral chromatic aberration as well.

In multicolor confocal imaging systems, the objective lens with axial chromatic aberration focuses the light with different wavelengths at different depths. Therefore, this aberration limits the system's ability to demonstrate colocalization in the axial direction. Similarly, in a confocal fluorescence imaging system, chromatic aberration results in the excitation of one volume, the imaging of a different volume, and the reduction of the fluorescence signal. When multiple fluorophores are used, the axial chromatic aberration reduces the possibility of establishing accurate positions of the fluorophores in the axial direction. For beam scanning confocal systems, the loss of signal is even greater as the imaged spot moves away from the optical axis.

In a confocal scanning optical system, lateral chromatic aberration can result in a loss of signal at the pinhole, particularly in the peripheral region, since the objective lens will focus the light at different locations for different wavelengths, either nearer to or farther from the optical axis than their true locations. Lateral chromatic aberration only affects off-axis image quality and must be corrected; it is not an issue when the confocal system works on axis.

Chromatic and spherical aberrations are the major aberrations degrading confocal performance, and they affect the entire image field. In general,

chromatic aberration is difficult to compensate for and must be corrected by the design of the optical systems. Spherical aberration is commonly caused by improper use of the objective lens or by the introduction of mismatched optical components or media in the optical path, and they can be minimized or compensated for.

The requirements for aberration correction depend on the confocal configurations and applications. For stage scan and objective scan, given that the objective lens only works on axis, only spherical and axial chromatic aberration need to be corrected. For single-wavelength reflectance confocal imaging, the aberrations need to be corrected only for the working wavelength. For a reflectance confocal imaging system with multiple wavelengths, the difference in the focal length for the whole spectrum must be within the depth of focus $\text{DOF} = (n\lambda)/\text{NA}^2$.

In confocal fluorescence imaging, the aberration correction is more demanding. Generally, the aberrations of the objective lens should be corrected for both excitation and emission wavelengths. If multiple fluorophores are used or when the difference between excitation and emission wavelengths is large, the aberrations should be minimized over a wide range of wavelengths. When UV excitation is used for visible emission, it is difficult to maintain the parfocality between the excitation point and the detection spot. Additional optical components may be employed in the illumination path to accommodate the wide range of wavelengths.

7.4.2 Design considerations of the objective lens

The design of the objective lens represents a compromise of the following parameters: infinite or finite conjugate, NA, telecentricity, immersion medium, working distance, transmission, and aberration correction. Ideally, a high NA is necessary to collect sufficient reflectance light or fluorescence light; a long working distance is preferred to allow maximum penetration depth and sample manipulation; a flat field is desired for accurate 3D reconstruction; a low axial chromatic aberration is needed for parfocality of the multiple wavelengths or fluorophores; low lateral chromatic aberration is required for precise registration of multiple fluorescence images; and high transmissions at the illumination and detection wavelengths are necessary to increase the overall efficiency.

As discussed in Chapter 4, in an infinity-corrected objective lens, light from a point in the focal plane becomes parallel after passing through the objective lens, and it does not converge until passing the tube lens. Other optical components, such as filters and beamsplitters, can be placed between the objective lens and the tube lens without shifting the location of the image point, both axially and laterally. Another advantage of an infinity-corrected objective lens is that focusing is possible by moving the objective lens, instead of the stage, without introducing aberration. On the other hand, for a finite conjugate objective lens, all of the components in the imaging path should be considered when optimizing the objective lens. Any deviation from the design condition will

introduce some aberration. Therefore, it is difficult to insert an optical element or change the distance from the object plane to the image plane.

Similar to other imaging systems, the signal intensity and resolving power in confocal imaging systems are proportional to the square of the NA of the objective lens. Therefore, a high NA is desired for most applications. However, a high NA means a complex optical system, small FOV, short working distance, and low light transmission.

Telecentricity is generally required for the objective lens used in confocal imaging—it ensures that every point on the tissue is illuminated at the same angle, and the light cone collected by the objective lens is the same for each point. Telecentricity also maintains the same magnification for points at the same lateral position but at different depths. Telecentricity is important to obtain accurate 3D information.

When designing an objective lens, one must pay attention to the media in the object space. The objective lens is designed with some predefined parameters, such as the working distance and the refractive index of the media in the object space, and it cannot accommodate different working conditions. Any deviation from the design parameters will degrade imaging performance, especially for an objective lens with a high NA.

Depending on the applications, some confocal imaging systems focus deep into the tissue, and other systems only capture the information on the tissue surface. For systems that image deep into tissues, highly corrected water-immersion objective lenses are required.

Transmission is determined by the material absorption and AR coatings. In some fluorescence imaging systems, such as immunofluorescence imaging, only a few photons are available for detection. Therefore, the transmission property is very important and can have a greater significance than some other specifications, such as field curvature. For high-performance objective lenses, such as the apochromatic lens, fluorite and other glasses with low transmission in the UV are usually needed to achieve aberration correction. Therefore, when UV light is used to excite fluorescence, semiapochromatic correction is the ideal compromise between the transmission and the aberration correction.

The working distance of the objective lens is another parameter to consider when designing confocal imaging systems. Generally, the working distance is shorter for an objective lens with a higher NA and a better aberration correction. To design an objective lens with a long working distance, the techniques discussed in Chapter 4 are effective.

Specific requirements in confocal imaging may lead to the development of optics that need to be less stringently corrected in one or more of the resolution-compromising aberrations in favor of achieving other design goals that are more critical in confocal performance. For example, the need for higher photon efficiency in living-cell imaging may provide an incentive to maximize NA and transmittance but sacrifice correction of certain aberrations.

For an objective lens used in a pre-objective scan, when its distortion is corrected in the usual sense, the position of the focused spot from the optical axis

is $f \tan\theta$, where f is the focal length and θ is the scan angle. For many applications, an exact linearity between the scan angle and the position of the focused spot is desired. When some barrel distortion is introduced into the lens, the image height will be a linear function of the scan angle $f\theta$. Lenses with this property are called F - θ lenses. An F - θ lens enables the system to scan the focused spot with constant linear velocity and to collect the data at a constant rate, thereby obtaining an equally spaced measurement.

Figure 7.27 plots the image height difference between an F - θ lens and a distortion-corrected lens as a function of scan angle. For the distortion-corrected lens, the error in the linearity is smaller than 1% when the scan angle is less than 10 deg. For most confocal scanning microscopes, the scan angle is small; therefore, the objective lenses do not necessarily have to be an F - θ type.

During the design process of optical systems for illumination and detection, flat optical plates are used to represent the filters, beamsplitters, polarizers, and other components with flat surfaces. Even without optical power, any component that is located between the objective lens and the detection aperture may affect the confocal image quality and must be of good optical quality. Most components should have transmitted wavefront distortion and surface-flatness specifications on the order of one wave per inch or better.

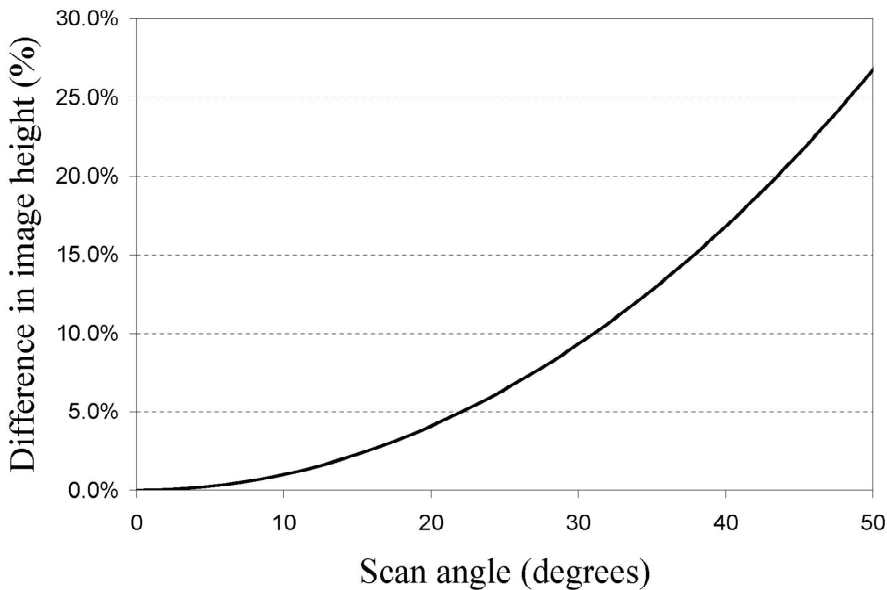


Figure 7.27 Image height difference between an F - θ lens ($f\theta$) and a distortion-corrected lens [$f \tan\theta$] as a function of scan angle.

7.4.4 Design examples

7.4.4.1 Objective lenses for microarray scanners

The microarray biochip is a new technology in biology and life sciences that can be used as a powerful analyzing tool for DNA and protein research. Microarray biochips are small substrates containing thousands of DNA sequences that represent the genetic codes of a variety of living organisms, including humans, plants, animals, and pathogens. They provide researchers with volumes of information in an efficient format. There are two main approaches to reading out the fluorescent signal of every array spot. One is composed of a laser light source, a scan lens, and a PMT, and the other is composed of a wide-field illumination system, an imaging system, and a CCD camera. The first technology, often using the confocal principle, scans the microarray point by point. It has the advantages of high detection sensitivity, high image resolution, and high fluorescence collection efficiency.

Figure 7.28 shows a telecentric objective lens used in a microarray scanner. The lens is optimized for a wavelength range from 640–700 nm, covering the peak excitation wavelength of 649 nm and the peak emission wavelength of 666 nm of Cy5. One doublet in the lens is used to reduce chromatic aberrations. The

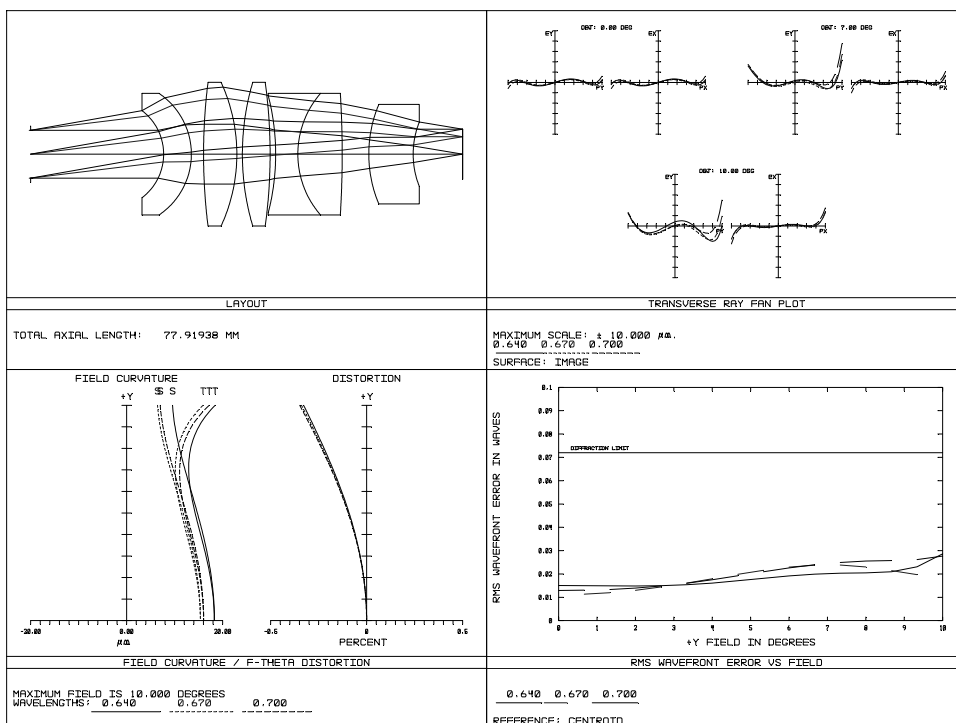


Figure 7.28 Objective lens for a microarray scanner. The lens has a 0.17 NA, 9-mm FOV, and 8-mm working distance.

NA of the objective lens is 0.17, and the FOV is 9 mm, corresponding to a ± 10 -deg scan angle. The clear working distance is 8 mm, and the clear distance from the pupil to the first lens is 20 mm, which is enough for a galvanometer scanner. This objective lens is diffraction limited over the entire FOV. The F - θ distortion is less than 0.5%. The lens has a 1- μm lateral chromatic aberration and an 8- μm axial chromatic aberration, which is smaller than the depth of focus.

7.4.4.2 Achromatic objective lens in the near-IR

Most microscopic objective lenses are designed for the visible spectrum and have a certain amount of axial and lateral chromatic aberrations in the near-IR spectrum. For some applications, near-IR is preferred because of low tissue absorption. Therefore, a custom objective lens is required.³⁵

Figure 7.29 shows a telecentric water-immersion objective lens with five elements, two of which are cemented doublets. It is designed for three wavelengths: 785, 810, and 850 nm. The specifications for the objective lens include water immersion, infinite conjugate, 0.75 NA, 0.70-mm FOV, 4.75-mm EFL, and a chief ray angle of less than 1 deg. The doublet away from the aperture stop helps flatten the field. The field curvature and lateral color plots show that

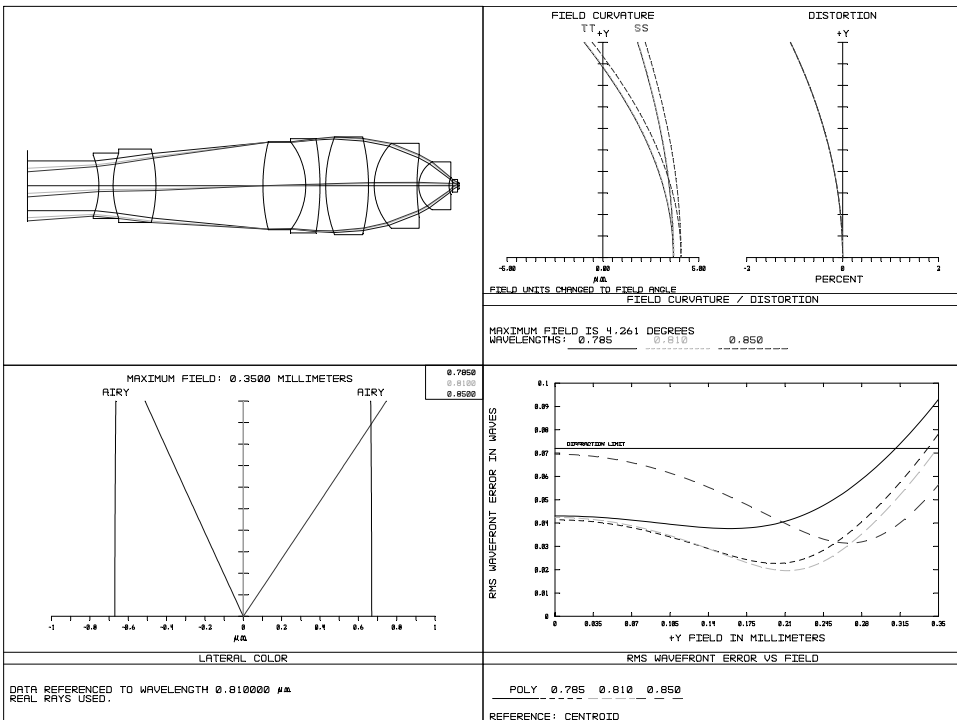


Figure 7.29 Telecentric water-immersion objective lens for confocal imaging in the near-IR.

the axial chromatic aberration is less than 1 μm , and the lateral chromatic aberration is 1.26 μm , which is slightly larger than 1 μm . The distortion is less than 1.2%. The RMS wavefront error is less than 0.07λ for the entire FOV. Basically, it is a diffraction-limited objective lens. The maximum chief ray angle is less than 0.6 deg. The diameter of the entrance pupil is 7.3 mm, and it is located 21 mm inside the objective lens from the first surface. A relay lens is required to image the scan mirror to the entrance pupil.

7.4.4.3 Objective lens for UV and visible light

A confocal imaging system using UV wavelengths has several advantages over systems using visible light only. For example, some dyes can only be excited by UV or near-UV light. Also, because of the shorter wavelength, a confocal imaging system can have better spatial resolution. However, extending the working wavelength to the UV presents a number of problems, such as material selection and coatings. Few optical glasses have good transmission in the UV, and those glasses usually have similar dispersion characteristics, resulting in difficulties in correcting chromatic aberrations. In addition, some cement materials absorb UV light strongly and may exhibit autofluorescence, which limits the use of cemented lenses.

Calcium fluoride, fused silica, and quartz are commonly used glasses in UV systems. Calcium fluoride is particularly helpful in controlling chromatic aberrations. Currently available glasses make it very difficult to correct the chromatic aberrations over the entire imaging plane. Therefore, diffraction-limited performance with high-NA objective lenses can often be obtained only for a limited spectral range.

Figure 7.30 shows a fairly complex objective lens designed for confocal imaging at 350 and 633 nm.³⁶ It is a water-immersion objective lens with an NA of 1.0 and a FOV of 75 μm (diagonally). To correct for chromatic aberration, other glasses besides CaF_2 are used, such as FPL51, PBL26, BSL7, FSL5, and ADF10. The entrance pupil is located inside the objective lens, 16.845 mm from the first lens surface. This objective lens has diffraction-limited performance at both 350 and 633 nm. The axial and lateral chromatic aberrations are 0.45 and 0.08 μm , respectively. Even though this objective lens is only optimized for 350 and 633 nm, it can be used for any individual wavelength between 350 and 633 nm with defocus.

7.5 Fiber-Optic Confocal Imaging Systems

To translate the imaging capabilities of a confocal microscope into more confined spaces for clinical applications, the basic requirements of a miniature confocal imaging probe include a small diameter, short rigid length, and flexible umbilical connection to the remainder of the device. Optical fibers are one of the key elements in flexible confocal imaging systems. Generally, fiber-optic confocal systems can be classified into two groups: fiber-bundle systems and single-fiber systems. The single fiber, or each individual fiber in a fiber bundle, is

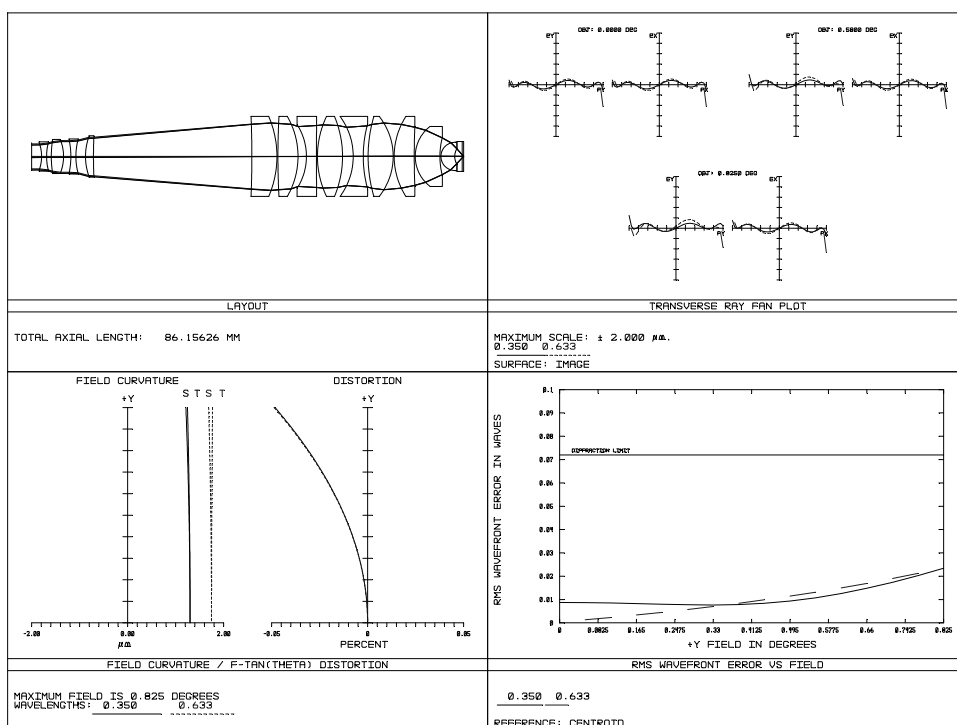


Figure 7.30 Telecentric water-immersion objective lens for confocal imaging at 350 and 633 nm.

used as the illumination and the detection apertures, enabling miniaturization of the imaging head.

The important factors to consider when designing fiber-optic confocal imaging systems include:

- Efficient coupling of light into an optical fiber,
- Proper handling of the fiber output as a light source,
- Efficient projection of the illumination light onto the object surface with the fiber as the illumination pinhole,
- Efficient collection of the light from the object and efficient coupling of the light back to the fiber that acts as a detection pinhole,
- Reducing the background noise related to the fiber, and
- Transmission of short pulses suitable for nonlinear microscopy.³⁷

The coupling of the light into the fiber depends on the following factors: the beam profile from the light source, the mode profile of the fiber, the NA of the fiber, and the coupling lens. The coupling lens should focus the light from the light source to a spot as small as possible and onto the fiber surface perpendicularly, thus maximizing the overlap between the distribution of the energy of the focused spot and the modal field of the fiber.

Any fiber-based confocal system involving illumination and detection via the same fiber, end-face reflections and fluorescence from the fibers may give rise to a background signal, reducing the detection dynamics.

7.5.1 Single-fiber confocal imaging

Single-fiber confocal imaging systems use a single fiber for illumination and detection. The fiber itself works as the illumination and detection pinholes. Both single-mode and multimode fibers have been used in confocal imaging. A single-mode fiber has a very small diameter ($<10\ \mu\text{m}$) and a small acceptance angle. It can only transmit a single-mode laser beam and conserve the coherence of the laser light. The output beam is a series of concentric, diverging wavefronts and can be recollimated and focused to a small spot. A multimode fiber has a larger diameter and a relatively large acceptance angle. Each of the different modes maintains coherence, but the overall output is the sum of all of the propagated modes. The output can neither be restored to a single parallel wavefront nor be focused to a diffraction-limited spot. The spatial distribution of the optical power at the fiber surface of a multimode fiber is equal to the spatial distribution of the refractive index, which means that the spatial distribution of the irradiance at the fiber surface is uniform for a step-index multimode fiber.³⁸

Both ends of the fiber are in the conjugate plane of the pinhole; therefore, the Fresnel reflection of the illumination light at the fiber ends is background noise and must be suppressed by one of the following methods: angle polishing of the fiber surface, index matching of the fiber core, or AR coating on the fiber surface. The input end of the fiber with the surface polished at an angle directs the reflected illumination light out of the detection path. The output end of the fiber can also be polished at an angle so that the light reflected from the angled surface cannot propagate back to the input end and leak through the fiber cladding. By using an index-matching medium, the reflection from the fiber surface can be minimized. An AR coating can also be applied to the fiber surface to reduce the Fresnel reflection. However, it is very challenging to coat a fiber surface.

Single-fiber confocal imaging requires a small raster scan mechanism in the distal end to scan the focused beam across the region of interest. The challenge is to miniaturize the scan mechanism. Several scan methods have been proposed and implemented to construct 2D or 3D images.³⁹⁻⁴¹ Two of the most practical methods are fiber scan and MEMS scan.⁴²⁻⁴⁵

The fiber scanning mechanism vibrates the fiber tip, as shown in Fig. 7.31. The fiber is attached to a piezoelectric-tube actuator with two pairs of electrodes.⁴⁶ The movement of the fiber is driven by two amplitude-modulated waves. The light from the fiber tip is focused onto the tissue and is scanned across the tissue by vibrating the fiber tip. A portion of the light backscattered from the tissue is collected by the lens and coupled back into the fiber and delivered to the detector. In addition to the x and y scan, the fiber can also be moved along the axis direction to achieve 3D scanning. Except for the piezoelectric tube, the movement of the fiber tip can be provided by an

electromagnetic actuator. An alternative to the fiber scanning configuration discussed in Fig. 7.31 is to mount a very small lens on the tip of the fiber or to fabricate a fused-fiber lens on the tip instead of a separate scan lens. The issues are the low quality of the lens and the small achievable NA.

Depending on the movements of the fiber tip, the requirements on the scan lenses are different. As shown in Fig. 7.32(a), when the fiber tip is moved laterally in the x and y directions and the tip moves on the same plane, the light emerging from the fiber is always a cone with the central ray parallel to the optical axis. Therefore, the scan lens should be telecentric on the fiber side. If a lens is not telecentric on the fiber side, less light from the fiber will be focused on the tissue surface, and less light from the tissue will be possibly coupled back into the fiber as the fiber moves away from the optical axis.

When the fiber is vibrated at a pivot point, as shown in Fig. 7.32(b), the locus of the fiber movement is a curved surface instead of a flat plane. The central rays of the light cones from the fiber virtually come from the center of the curved surface. Therefore, the scan lens should be designed so that its chief rays converge at the same point as the center of the curved surface where the fiber tip vibrates. Figure 7.33 shows a lens optimized for an optical fiber with an NA of 0.13; the radius of the locus of the fiber tip is 10 mm.³⁹ The FOV is 0.45 mm in diameter, and the NA is 0.34 on the tissue side. The chief rays converge at the center of the locus of the fiber tip, and the lens is telecentric on the tissue side. Overall, it is a diffraction-limited objective lens. The major remaining aberrations are spherical aberration and distortion. This objective lens is fairly complicated and could be simplified by sacrificing aberration correction.

As an alternative to the fiber scan, MEMS mirrors, which provide a one-axis or two-axis rotation about a pivot point at the center of a mirror, have been also used to scan the focused spot across the tissue. MEMS mirrors can perform large-angle scanning with arbitrary scan patterns. They can also be actuated with resonance to provide a high-speed raster scan. Furthermore, MEMS devices can operate reliably with low power consumption, making them suitable for integration into *in vivo* imaging devices.

Figure 7.34 shows two types of fiber-optic confocal imaging systems using MEMS mirrors: side view and straight view. Figure 7.34(a) is a side-view configuration in which the viewing direction is orthogonal to the insertion direction of the confocal probe. Figure 7.34(b) is a straight-view configuration in which the viewing direction is the same as the insertion direction of the confocal probe. The optical path is folded to redirect the beam and shorten the length of the probe. The first mirror is a folding mirror. The scan is achieved by rotating the second mirror. The light reflected by the fixed mirror is scanned two-dimensionally by the MEMS gimbal scanner and is focused onto the tissue by the objective lens. The reflected light or fluorescence light goes back along the same way into the fiber and is led to the detector through a fiber coupler or beamsplitter. To prevent the light reflected from the lens surfaces or mirror surface from entering the fiber, a very small central portion of one of the two mirrors can be treated to have low reflectivity.

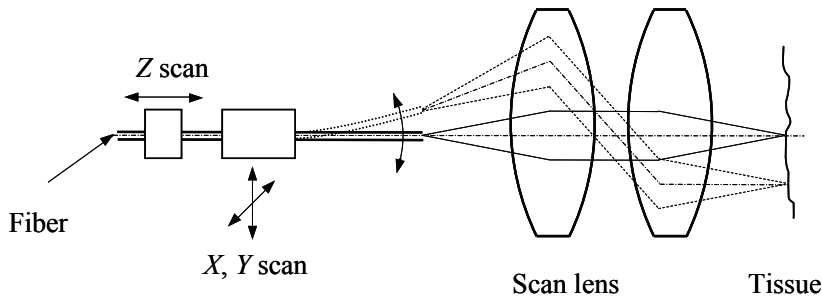


Figure 7.31 Diagram of a single-fiber scanning system. A single fiber is used as the illumination source and detection pinhole.

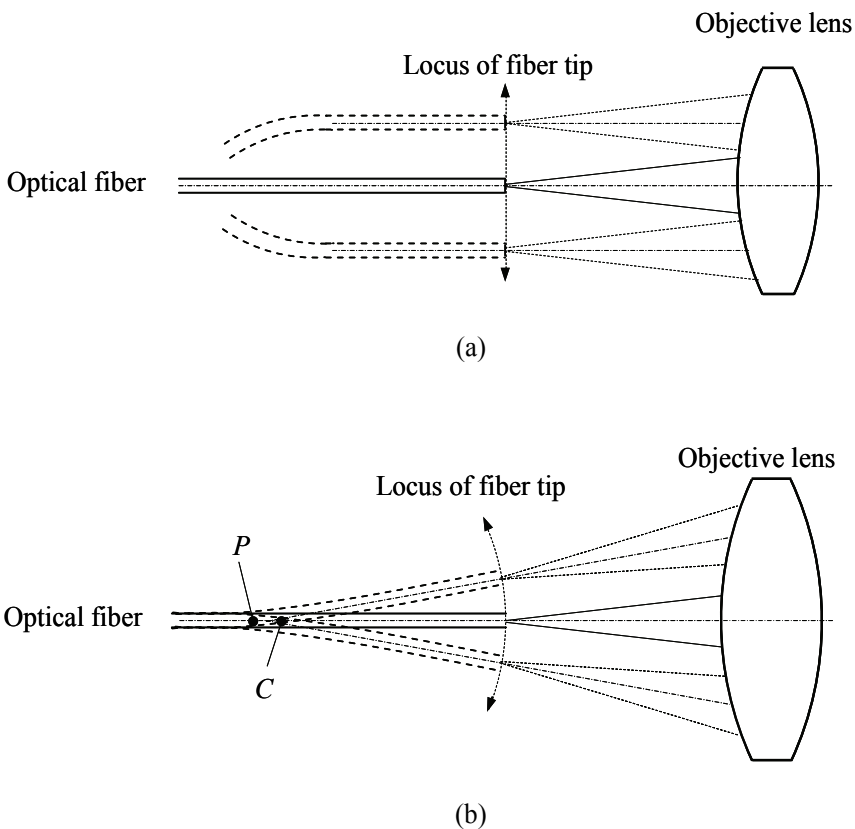


Figure 7.32 Ray diagram of a scanning-fiber imaging system. (a) The fiber is moved laterally with the tip on the same plane. The central rays are parallel to the optical axis. (b) The fiber is vibrated at a pivot point P . The central rays are always perpendicular to the locus of the fiber tip. C is the center of the curved surface where the fiber tip vibrates.

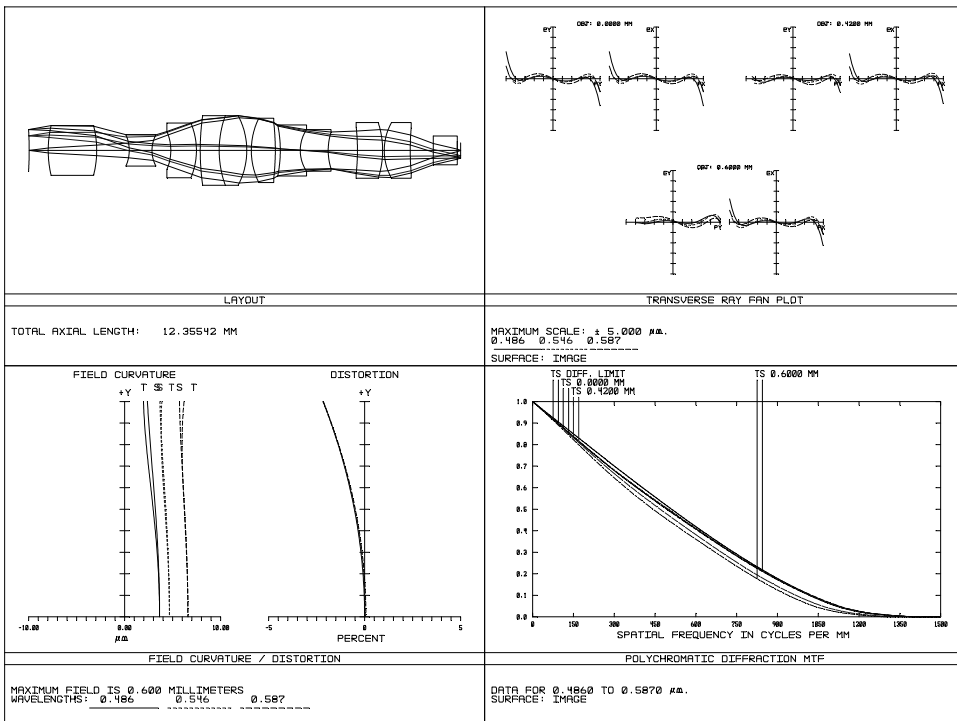


Figure 7.33 Objective lens for single-fiber confocal imaging.

Figure 7.34(c) shows another straight-view configuration. The light from the fiber is first reflected by a small mirror, which can be part of the lens surface or a separate component attached to the lens, and then it is scanned by the second mirror. The scan mirror can be either flat or spherical, as shown in the figure. The two mirrors share the same optical axis with the objective lens. Similarly, a small central portion of the first mirror is treated to have a much lower reflectivity. One issue with this configuration is central obscuration, which reduces image contrast and light collection efficiency. Light collection efficiency is extremely important for fluorescence imaging because the fluorescence signal is usually very weak.

Because of the difficulty in coating the fiber end surface and housing the index-matching medium between the lens and the fiber, the fiber end surface is usually polished at an angle such that the light reflected from the polished surface cannot propagate back to the proximal end of the fiber.

The ideal objective lenses for the configurations in Fig. 7.34 should have the aperture stop outside of the lenses so that the MEMS scan mirror can be placed at the aperture stop to achieve telecentricity on the tissue side.

When a fiber scanning probe is used to image deep into the tissue, the objective lens should be optimized with water as the immersion medium to minimize the effect from the difference in refractive indices. The volume-

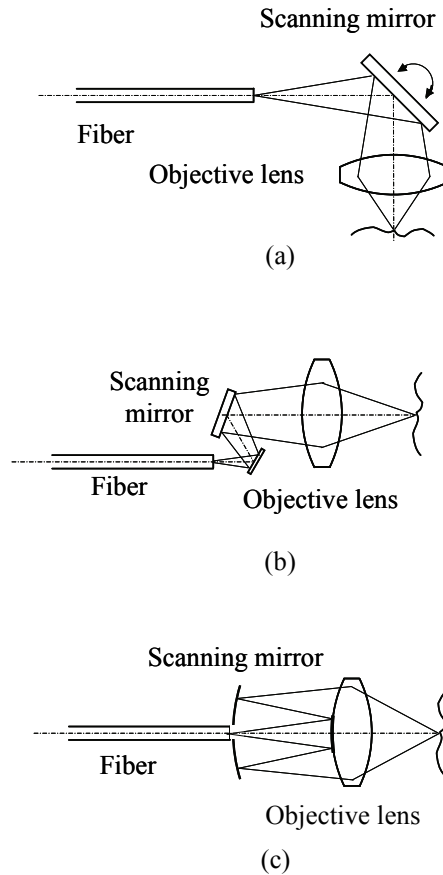


Figure 7.34 Configurations of fiber scanning systems using MEMS mirrors: (a) side-view configuration, (b) straight-view configuration without central obscuration, and (c) straight-view configuration with central obscuration.

averaged refractive index of most biological tissue falls within 1.34–1.62, which is greater than the refractive index of water (1.33). In addition, the index of the tissue varies with depth and from individual to individual. With a compact and simple objective lens used in fiber scanning confocal imaging, there generally are not enough variables to compensate for the aberrations induced by the variation in the refractive index.

Figure 7.35(a) shows an objective lens used in the configuration shown in Fig. 7.34(a). The lens has an NA of 0.13 on the fiber side and 0.36 on the tissue side. The FOV is $170 \times 170 \mu\text{m}$. Both lens surfaces are aspherical to minimize spherical aberration. For the same purpose, the glass usually has a high refractive index. This objective lens has a diffraction-limited performance over the entire FOV at the working wavelength, as shown in Fig. 7.35(b). Because there is no effort to control chromatic aberration, this lens can only be used in single-

wavelength reflectance confocal imaging; it is not suitable for fluorescence imaging.

Figure 7.36(a) shows the same objective lens discussed above but is used in the configuration shown in Fig. 7.34(c). Because of the central obscuration, the MTF at mid-frequency is reduced, which means that the image contrast is reduced, as shown in Fig. 7.36(b), but the resolution is the same. Another major issue is the reduced light collection efficiency.

Figure 7.37(a) shows an objective lens that can be used for fluorescence confocal imaging. It is a relatively simple design and is only corrected for two wavelengths, 400 and 500 nm; 400 nm is the excitation wavelength, and 500 nm is the emission wavelength. The objective lens consists of one cemented doublet and two singlets. It is a water-immersion objective lens, with an NA of 0.5 on the tissue side. The cemented doublet helps to reduce chromatic aberrations, and the

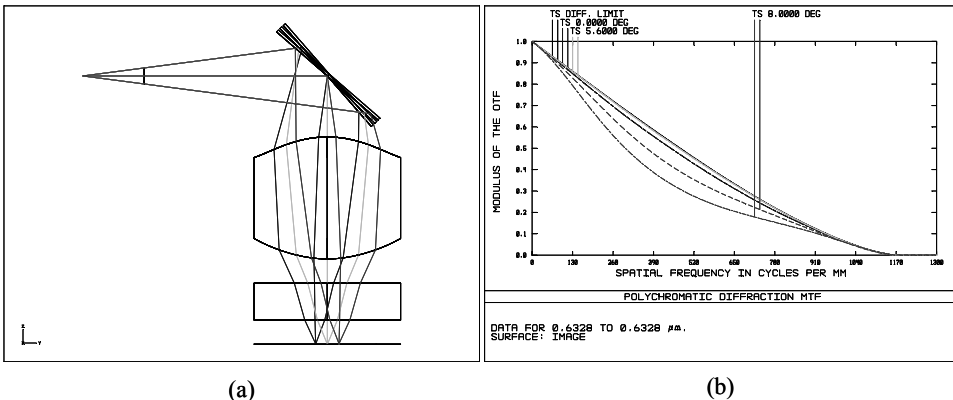


Figure 7.35 (a) Ray diagram and (b) MTF of an objective lens for the configuration in Fig. 7.34(a).

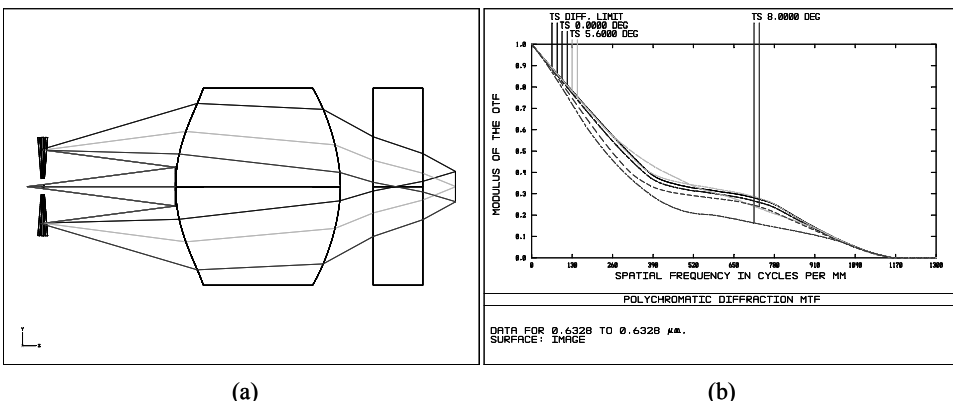


Figure 7.36 (a) Ray diagram and (b) MTF of an objective lens for the configuration in Fig. 7.34(c). The NA is 0.36, and the FOV is 170 μm . The central portion of the lens is blocked by the first reflective mirror.

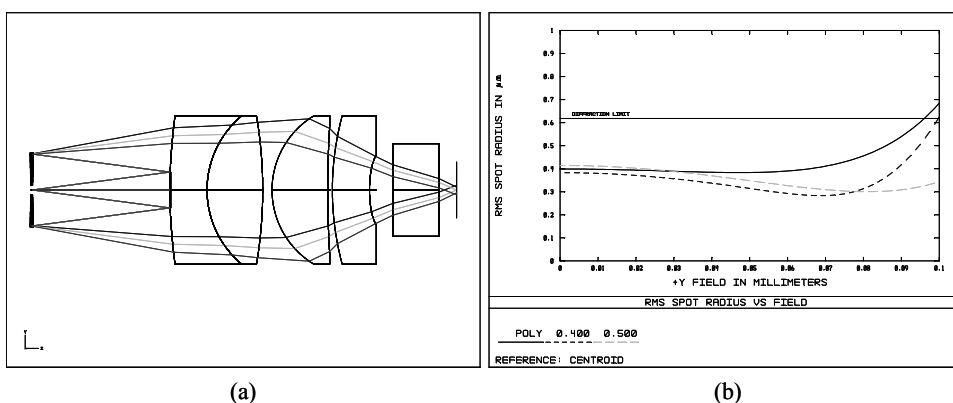


Figure 7.37 (a) Ray diagram and (b) RMS spot radius plot of an objective lens for fluorescence confocal imaging. The NA is 0.37, and the FOV is 200 μm .

two singlets contribute to the correction of spherical aberration. To further reduce spherical aberration, the first and last surfaces of the objective lens are conic surfaces. Even with a doublet to correct the chromatic aberrations, the axial chromatic aberration is still relatively large, about 7 μm . The RMS spot radius plot indicates that this lens has good performance over the entire FOV.

The advantages of the fiber scanning method are its high lateral resolution and good image quality. However, there are some limitations related to this distal scanning method.⁴⁷ Generally, it is challenging to miniaturize the scanning probe below 5 mm because of the space required to hold the piezoelectric-tube actuator, the electromagnetic actuator, or the MEMS mirror. Another limitation related to the scanning mechanism located at the distal end is the scanning speed; it is a challenge to achieve real-time acquisition. One more limitation is its robustness and resistance to shocks, vibrations, and decontamination. A probe with scanning mechanisms at the distal end may have problems with resistance to pressure and temperature.

7.5.2 Fiber-bundle confocal imaging

7.5.2.1 Introduction

To address the issues related to the scanning mechanism at the distal end of a single fiber, a proximal scanning architecture using a fiber bundle has been developed. Beam scanning across a fiber bundle allows the scanning mechanism to be placed at the proximal end of the fiber bundle. Both coherent and incoherent fiber bundles have been used for confocal imaging.^{48,49} The resolution of the transmitted image is limited by the pixilated nature of the fiber bundle as well as by the degree of coupling between the adjacent fiber cores.

Figure 7.38 is a diagram of a typical confocal imaging system using a fiber bundle. The light from the light source is collimated and scanned by the fast scan mirror in one direction. The pivot point of the fast-scan mirror is imaged onto the slow scan mirror by a relay system. Therefore, the angular movement achieved

by the fast scan mirror is relayed to the slow-scan mirror. The recollimated beam is again scanned in the orthogonal direction by the slow-scan mirror and relayed to the entrance pupil of the scan lens by a second relay system. The scan lens, which is telecentric on the fiber side, converts the angular movement of the light at the entrance pupil plane to the lateral movement in the focal plane and couples the light into one fiber at a time on the proximal end of the fiber bundle. The light propagates through the fiber bundle and is focused onto the tissue surface by the objective lens after it emerges from the fibers. On the distal end of the fiber bundle, each fiber serves both as a point light source and a detection pinhole. The objective lens collects the light backscattered from the tissue and couples it into the fiber bundle. The light carrying the tissue information travels in a reverse path relative to the illumination light and reaches the detector through the beamsplitter after being descanned by the slow- and fast-scan mirrors. The resulting 2D images are produced by raster scanning the illumination spot across the proximal face of the fiber bundle.

System optimization depends heavily on the properties of the fiber bundle. A typical fiber bundle has a $\sim 4 \mu\text{m}$ core diameter and $\sim 6 \mu\text{m}$ core-to-core spacing.

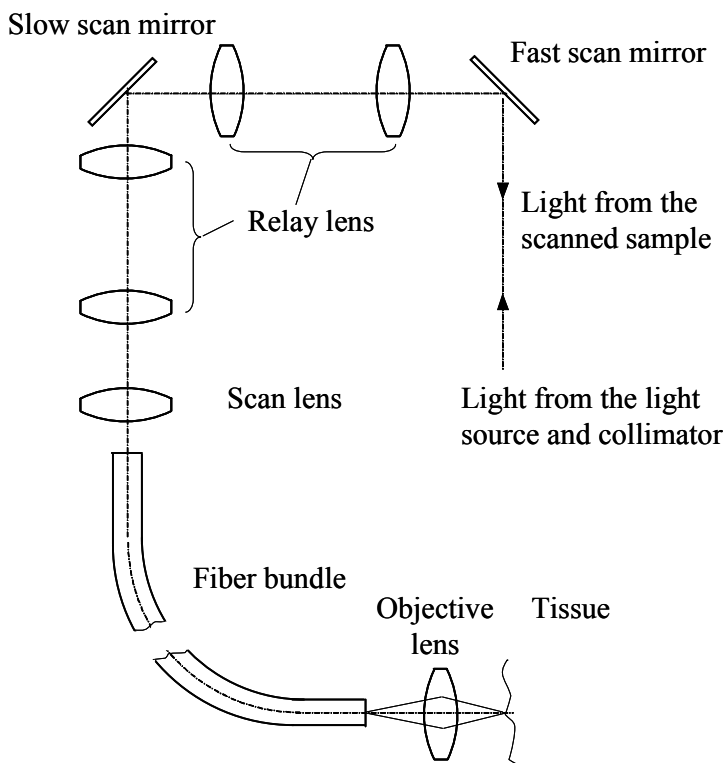


Figure 7.38 Schematic diagram of a confocal imaging system using a fiber bundle.

If the cores are much smaller, they begin to act like single-mode fibers, particularly at longer wavelengths, reducing the throughput of noncoherent fluorescence light. To reduce crosstalk, the core-to-cladding refractive index difference must be made as large as possible for a high NA.

There are several advantages to placing the scan mechanism at the proximal end. First, distal optics can be designed with fewer space constraints to achieve better image quality. Second, a more robust and reliable rapid-scanning solution, such as resonant and galvanometric mirrors, can be used. Third, the probe is more compatible with any cleaning or decontamination procedure since there is no active element in the probe head. There are also several disadvantages related to fiber bundles. Image pixelation, which is due to the nature of fiber bundles, will induce image artifacts, aliasing, and resolution loss. The crosstalk between adjunct fiber cores due to the thin cladding layers further reduces image resolution and contrast. The crosstalk may be reduced by using an incoherent fiber bundle in which the relative fiber positions are randomized and not maintained throughout the bundle. This approach is especially useful for line-scanning configurations in which adjacent fibers at one side are illuminated concurrently, but the images originate from physically separate locations in the object.⁴⁹

Typical refractive index differences in biological tissue range from 0.05–0.10, so the signal (the light reflected from the tissue) is low. Given that both ends of the fiber bundles are in the conjugate planes of the detection pinhole, the reflection from the fiber surfaces must be minimized because the reflection from the fiber surface is much stronger than the signal from the tissue. Polishing the fiber surface at an angle can be a solution at the distal end, but probably not at the proximal end. At the proximal end, the tilted fiber surface can be outside of the depth of field of the scan lens, resulting in nonuniform coupling efficiency and possible crosstalk. At the distal end, the objective lens can be designed for a tilted fiber surface by meeting the Scheimpflug condition; however, the magnification is not constant across the field, and the tissue section under investigation is tilted related to the optical axis of the objective lens. Filling with an index-matching liquid between the fiber surfaces and the lens is a good solution. The drawback is that the lens must be sealed to prevent fluid leakage. Another option is to cement a glass plate on the fiber surface to move the air–glass interface out of the conjugate plane. The refractive indices of the cement and the plate are chosen to match that of the fiber core. The plate has to be thick enough that the light reflected at the air–glass interface can be blocked by the fiber itself.

The scan lens used to couple the illumination light into a fiber bundle must be diffraction limited so that it will only illuminate one fiber at a time to achieve the highest possible coupling efficiency with least crosstalk, leading to the highest image contrast. The NA should be equal to or slightly smaller than that of the fiber. Since the fiber surface is flat, the field curvature should be minimized. The scan lens must be telecentric on the fiber side so that the chief ray is perpendicular to the fiber surface. If this telecentric condition is not satisfied, the

coupling efficiency and the image contrast will vary across the fiber bundle. The FOV of the scan lens is determined by the effective diameter of the fiber bundle. The aperture stop of the scan lens should be at a reasonable distance from the first surface of the lens so that there is enough space to place the scan mirror at the aperture stop. Otherwise, a relay lens is required to image the scan mirror at the entrance pupil of the scan lens.

Figure 7.39 shows a scan lens that meets the above requirements. It is telecentric on the fiber side, with an NA of 0.32 and a FOV of 2 mm. The distance between the aperture stop and the first lens surface is 15 mm, enough to place a scan mirror. As shown in Fig. 7.39, the lens has a diffraction-limited performance across the entire FOV in the visible spectrum. The field curvature is reasonably controlled, and the distortion is less than 0.2%. Spherical aberration is the major limiting factor and can be reduced further by using aspherical surfaces. For laser scanning confocal imaging, this scan lens can be simpler because correction of chromatic aberration is not required.

Besides scan mirrors, a spatial light modulator, such as a DMD, has been used to couple light into individual fibers in a fiber bundle, as shown in Fig. 7.40.⁵⁰ Illumination optics collects the light from the light source, manipulates it, and then projects it uniformly onto a DMD. The DMD selectively turns on a single pixel or a pattern of pixels so that the light can be coupled into the fibers through the coupling lens. The backscattered light from the tissue is collected

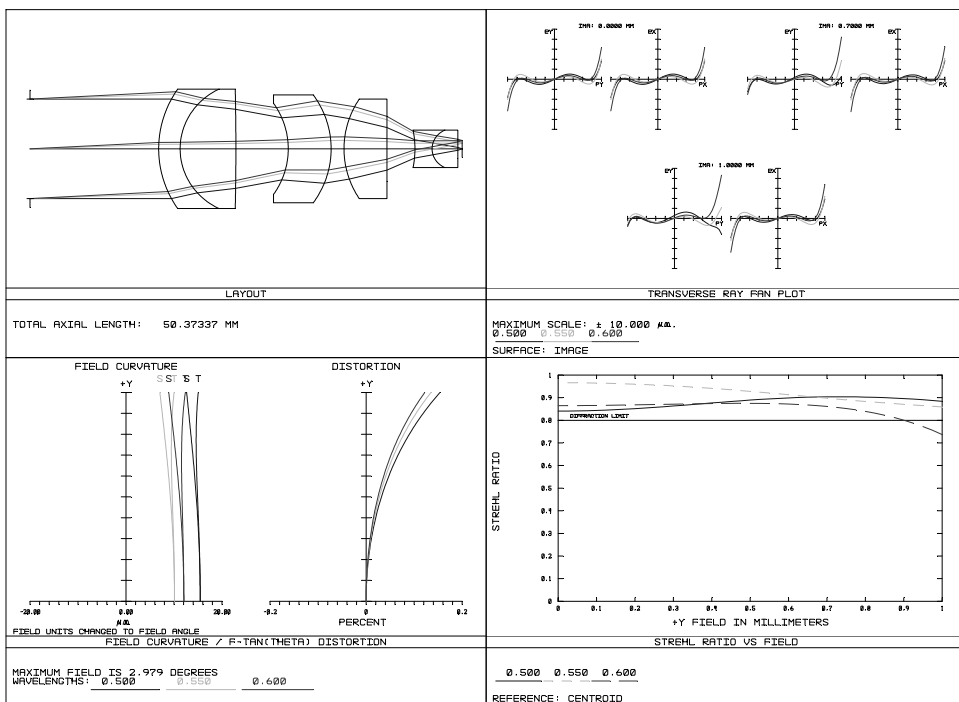


Figure 7.39 Telecentric scan lens to couple light in and out of the fiber bundle.

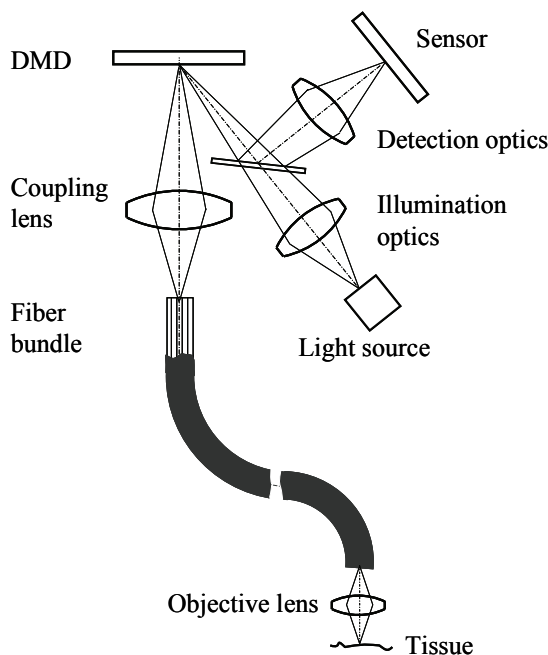


Figure 7.40 Confocal imaging system using a fiber bundle and a DMD.

and focused toward the on-state pixels in the DMD after passing through the fiber bundle. The detection optics focuses the light from the on-state pixels onto the sensor. The coupling lens should be double telecentric in order to collect all of the light directed by the DMD and couple it into a fiber.

The advantages of using a DMD include the following: (1) the active pixel can be random and does not have to follow the raster scan sequence, thereby reducing crosstalk; and (2) a set of fibers can be illuminated simultaneously, which is unachievable by using conventional scanning methods. The disadvantage is that all of the pixels in the DMD are always illuminated, so light efficiency is low. Another issue is that the pixels in the DMD are not always well aligned with the individual fibers in the fiber bundle. Some pixels may be imaged onto the inactive space between the fiber cores.

7.5.2.2 Optical design

The system performance of a fiber bundle confocal imaging system is degraded to some extent because of the pixelation from the imaging fiber bundle. Therefore, it is critical to maximize the optical performance of the objective lens to prevent any additional loss of resolution.

The objective lens should be designed as telecentric, in both the object and image spaces. The telecentricity on the fiber side is required so that the illumination light emerging from the fiber can be efficiently focused onto the tissue and the light backscattered from the tissue can be coupled into the fiber with high efficiency. The telecentricity on the tissue side is preferred to minimize

the change in magnification with respect to the depth position and maximize the collected signal from the tissue.

To maximize the coupling efficiency and image resolution and to reduce the crosstalk between the individual fibers, the NA of the objective lens on the fiber side should be equal or slightly smaller than the NA of the fiber so that all or most of the light emerging from the fiber can be focused onto the tissue and the fiber can be fully filled with the light backscattered from the tissue or with the fluorescence light. An NA smaller than that of the fiber will cause significant light loss and possible loss in the lateral and axial resolutions. An NA much larger than that of the fiber will cause light to escape the cores of the individual fibers, leading to cladding modes and crosstalk between adjacent cores, which will degrade the image quality.

The NA and FOV of the objective lens on the tissue side are dependent on each other because the diameter and the NA of the fiber bundle are fixed. The resolution requirement determines the NA on the tissue side and therefore determines the magnification of the objective lens. The FOV is the diameter of the fiber bundle divided by the magnification.

The performance requirement of the objective lens is that the MTF at the fiber plane should be greater than 50% at Nyquist frequency. For example, when the center-to-center distance of the individual fiber is 4 μm , the MTF of the objective lens should be better than 50% at 125 lp/mm.

As discussed in Sec. 7.5.1, the refractive index of the tissue varies with depth and from individual to individual. Because of space limitations, it is not feasible to integrate a movable element into an objective lens to compensate for the aberrations induced by the variation of the refractive index. To maintain the performance for imaging at a different depth, the optimization of the objective lens should consider the variations in refractive index. One technique is to optimize the lens in multiple configurations to take into account the possible variations of the refractive index in the object space.

Minimizing field curvature is one of the major requirements for the microscopic objective lens given that only part of field is in focus when the field curvature is not fully corrected. For an objective lens used in a confocal imaging system using a fiber bundle, field curvature will cause the fiber plane to be mapped to a curved surface inside the tissue. Therefore, the optical section in the tissue is not a flat surface. As previously discussed, owing to the nature of the tissue and the intention to image a surface within a 3D volume, a small amount of field curvature is acceptable, as long as the sag of the image plane is less than the depth of field of the objective lens.⁵¹⁻⁵³ During the design process, the curvature of the image plane can be treated as a spherical surface or even as an aspherical surface and as a variable to optimize system performance.

For fluorescence confocal imaging or multicolor confocal imaging, the lateral chromatic aberration should be smaller than the distance between the fiber cores. The requirement of axial chromatic aberration is the same as that for the conventional confocal objective lens.

Figure 7.41 shows an objective lens for a fiber bundle with an NA of 0.3. The working wavelength is 850 nm. The objective lens is an oil-immersion lens on the fiber side and a water-immersion lens on the tissue side. The oil between the fiber and the first lens surface contributes to the reduced Fresnel reflection, and the water on the tissue side helps to minimize the spherical aberration induced by the variation in the refractive index of the tissue. The NAs of the lenses are 0.3 (oil immersion) and 0.87 (water immersion). The total length from the fiber surface to the tissue is 10.5 mm, and the FOV is 0.3 mm. This objective lens consists of three lenses made from BK7; two of them have aspherical surfaces. The aspherical surface close to the aperture stop helps to control spherical aberrations, and the aspherical surface in the first lens helps to reduce distortion. The thick meniscus lens close to the fiber bundle helps to reduce field curvature and achieve telecentricity on the fiber side. During the optimization process, the image surface is set as an aspherical surface. The radius, as well as the aspherical coefficients, is allowed to vary. The final radius of the image surface is 2.5 mm.

When only one wavelength is used, a gradient index (GRIN) microlens is an option for the objective lens.⁵⁴ GRIN microlenses with a diameter of 350–2000 μm have been used for one-photon, confocal, and two-photon imaging. Compound GRIN probes usually combine two or more GRIN lenses with different characteristics: one is a high-NA objective lens, and the others are relay

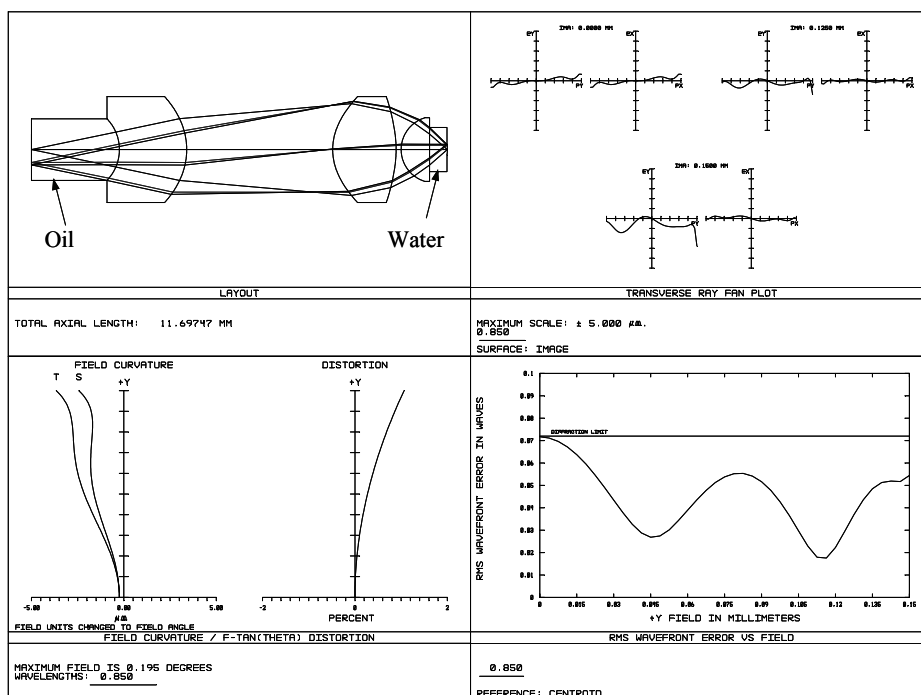


Figure 7.41 Objective lens for a fiber-bundle confocal imaging system.

lenses. Chromatic aberration in current GRIN lenses limits their application to reflectance confocal imaging.

To reduce the cost of high-volume fabrication, optical plastic is usually used. Various features, such as an aspheric surface, can be incorporated into the injection-molded elements with no additional cost. One requirement for optical plastic is low water absorption. Both ZEONEX® and TOPAS® have low water absorption of less than 0.01%. Commonly used optical plastics, such as polystyrene, polycarbonate, and acrylic, are not suitable for objective lenses used in moist environments because the thickness and radius of the curvature may change as a result of moisture absorption. Even a small change in thickness or curvature may cause major performance degradation.⁵⁵

Figure 7.42 shows a double-telecentric objective lens designed for a reflectance confocal imaging probe.⁵¹ The NA on the tissue side is 1.0 with water immersion, and the NA on the fiber side is 0.3 with oil immersion. The FOV is 250 μm , and the radius of the image plane is 1.5 mm. This objective lens uses optical plastic E48R for all lens elements. It is designed for fiber bundles that have an NA of 0.3 and a core diameter of 4.1 μm . As shown in the figure, this objective lens has diffraction-limited performance.

The lenses discussed above only work for reflectance confocal imaging at a single wavelength. Figure 7.43 shows an objective lens optimized for the visible

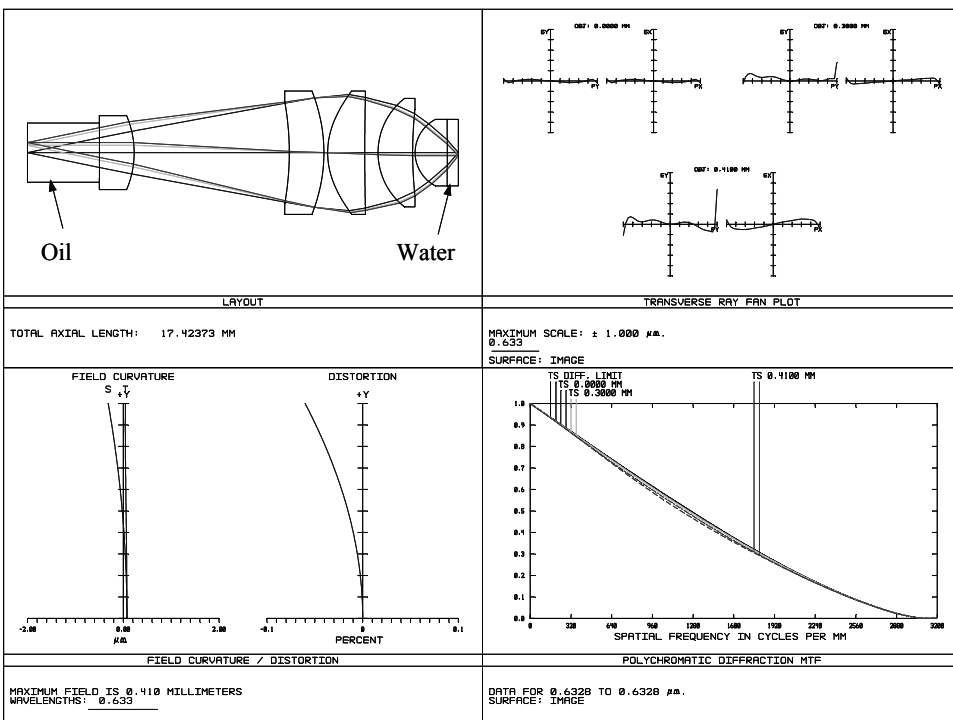


Figure 7.42 Objective lens for a fiber-bundle confocal imaging system.

spectrum, from 400–700 nm.⁵⁶ This lens allows for the simultaneous use of multiple dyes with multiple excitation and emission profiles. The NA of the objective lens is 0.45 (water immersion), the full FOV is 0.45 mm, and the total length is 13.46 mm. To minimize the Fresnel reflection from the fiber surface, a glass plate is glued to the fiber bundle. In order to correct the chromatic aberration, a cemented triplet with special glass, N-PSK53, is used. As shown in the figure, there is about 7.5 μm of axial chromatic aberration, and the lateral chromatic aberration is much smaller than the Airy disk. The system is diffraction-limited for the entire FOV. Spherical aberration is the major remaining aberration and can be further reduced by using aspherical surfaces.

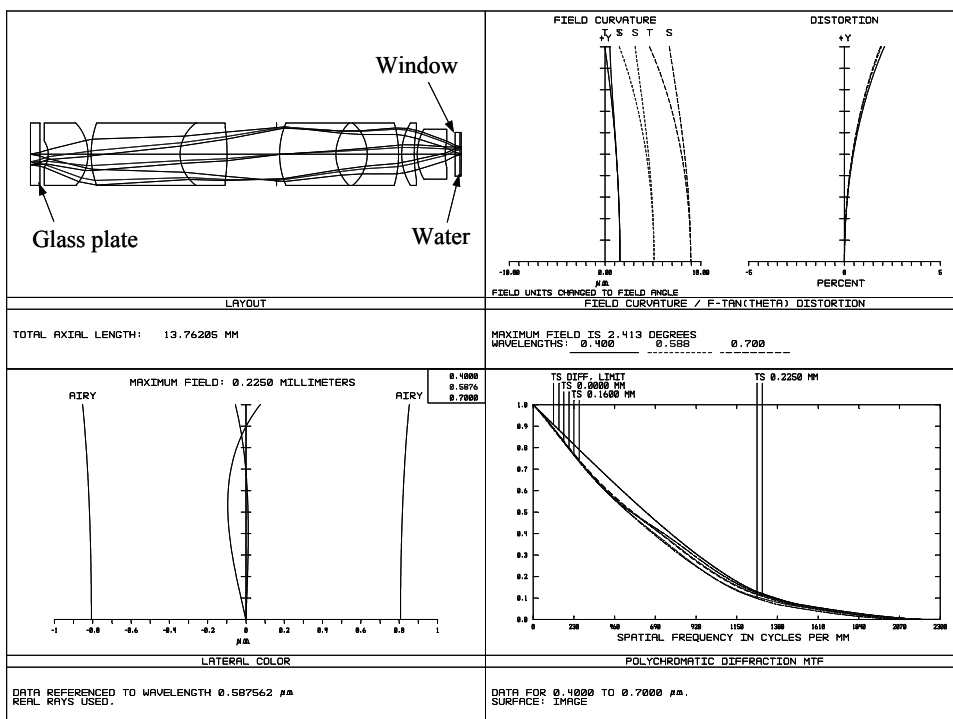


Figure 7.43 Objective lens with aberrations corrected in the visible spectrum.

References

1. M. Minsky, "Microscopy apparatus," U.S. Patent No. 2,013,467 (1961).
2. T. Wilson, *Confocal Microscopy*, Academic Press, London (1990).
3. T. R. Corle and G. S. Kino, *Confocal Scanning Optical Microscopy and Related Imaging Systems*, Academic Press, San Diego, CA (1996).

4. R. H. Webb, "Confocal optical microscopy," *Rep. Prog. Phys.* **59**, 427–471 (1996).
5. S. Wilhelm, B. Gröbler, M. Gulch, and H. Heinz, *Confocal Laser Scanning Microscopy: Principles*, Carl Zeiss, Inc., Jena, Germany (2003).
6. E. Gratton and M. J. vandeVen, "Laser sources for confocal microscopy," Chapter 5 in *Handbook of Biological Confocal Microscopy*, 3rd ed., J. B. Pawley, Ed., pp. 80–125, Plenum Press, New York (2006).
7. A. Nolte, J. B. Pawley, and L. Höring, "Non-laser light sources for three-dimensional microscopy," Chapter 6 in *Handbook of Biological Confocal Microscopy*, 3rd ed., J. B. Pawley, Ed., pp. 126–144, Plenum Press, New York (2006).
8. T. D. Wang, M. J. Mandella, C. H. Contag, and G. S. Kino, "Dual-axis confocal microscope for high-resolution in vivo imaging," *Opt. Lett.* **28**, 414–416 (2003).
9. W. B. Amos, "Optimum optical design characteristics for confocal and multi-photon imaging systems," Technical Note 09, Bio-Rad Laboratories, Hercules, CA (2000).
10. H. Gross, F. Blechinger, and B. Achnert, *Handbook of Optical Systems*, vol. 4, Wiley-VCH, Berlin (2008).
11. G. F. Marshall, *Handbook of Optical and Laser Scanning*, Marcel Dekker, New York (2004).
12. D. L. Dickensheets and G. S. Kino, "Micromachined scanning confocal optical microscope," *Opt. Lett.* **21**, 764–766 (1996).
13. W. Piyawattanametha, H. Toshiyoshi, J. LaCosse, and M. C. Wu, "Surface-micromachined confocal scanning optical microscope," in *Conference on Lasers and Electro-Optics*, San Francisco, CA, Optical Society of America (2000).
14. W. C. Warger II and C. A. DiMarzio, "Dual-wedge scanning confocal reflectance microscope," *Opt. Lett.* **32**, 2140–2142 (2007).
15. J. Sun, L. Liu, M. Yun, L. Wan, and M. Zhang, "Distortion of beam shape by a rotating double prism wide-angle laser beam scanner," *Opt. Eng.* **45**, 043004 (2006) [doi: 10.1117/1.2190672].
16. W. J. Smith, *Modern Optical Engineering*, 4th ed., McGraw-Hill, New York (2008).
17. T. Ruckstuhl, A. Walser, D. Verdes, and S. Seeger, "Confocal reader for biochip screening and fluorescence microscopy," *Biosens. Bioelectron.* **20**(9), 1872–1877 (2005).
18. F. Perraut, A. Lagrange, P. Pouteau, O. Peyssonneaux, P. Puget, G. McGall, L. Menou, R. Gonzales, P. Labeye, and F. Ginot, "A new generation of scanners for DNA chips," *Biosens. Bioelectron.* **17**, 803–813 (2002).

19. M. Petráň, M. Hadravský, M. D. Egger, and R. Galambos, "Tandem-scanning reflected-light microscope," *J. Opt. Soc. Am.* **58**, 661–664 (1968).
20. M. D. Egger and M. Petráň, "New reflected-light microscope for viewing unstained brain and ganglion cells," *Science* **157**, 305–307 (1967).
21. G. Q. Xiao, T. R. Corle, and G. S. Kino, "Real-time confocal scanning optical microscope," *Appl. Phys. Lett.* **53**, 716–718 (1988).
22. S. Yin, G. Lu, J. Zhang, F. T. S. Yu, and J. N. Mait, "Kinoform-based Nipkow disk for a confocal microscope," *Appl. Opt.* **34**, 5695–5698 (1995).
23. A. Ichihara, T. Tanaami, K. Isozaki, Y. Sugiyama, Y. Kosugi, K. Mikuriya, M. Abe, and I. Uemura, "High-speed confocal fluorescence microscopy using a Nipkow scanner with microlenses for 3-D imaging of single fluorescent molecule in real time," *Bioimages* **4**, 52–62 (1996).
24. E. L. Botvinick, F. Li, S. Cha, D. A. Gough, Y. Fainman, and J. H. Price, "In vivo confocal microscopy based on the Texas Instruments digital micromirror device," *Proc. SPIE* **3921**, 12–19 (2000) [doi: 10.1117/12.384208].
25. D. Dudley, W. Duncan, and J. Slaughter, "Emerging digital micromirror device (DMD) applications," *Proc. SPIE* **4985**, 14–25 (2003) [doi: 10.1117/12.480761].
26. W. B. Amos, "Optimum optical design characteristics for confocal and multi-photon imaging systems," Technical Note 09, Bio-Rad Laboratories, Hercules, CA (2000).
27. W. B. Amos, "Achromatic scanning system," U.S. Patent No. 4,997,242 (1991).
28. R. J. Anderson and C. Larson, "Reflective relay optics for use in laser deflection systems," *Appl. Opt.* **10**, 1605–1608 (1971).
29. G. C. de Wit and J. J. M. Braat, "Offner-type pupil relay optics for a scanning system," *Proc. SPIE* **2774**, 553–561 (1996) [doi: 10.1117/12.246702].
30. H. Jacobsen and S. W. Hell, "Effect of the specimen refractive index on the imaging of a confocal fluorescence microscope employing high aperture oil immersion lenses," *Bioimaging* **3**, 39–47 (1995).
31. S. Hell, G. Reiner, C. Cremer, and E. H. K. Stelzer, "Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index," *J. Microsc.* **169**, 391–405 (1993).
32. I. L. Hale and B. Matsumoto, "Resolution of subcellular detail in thick tissue sections: immunohistochemical preparation and fluorescence confocal microscopy," in *Cell Biological Applications of Confocal Microscopy*, B. Matsumoto, Ed., pp. 289–324, Academic Press, San Diego, CA (1993).
33. H. E. Keller, "Objective lenses for confocal microscopy," in *Handbook of Biological Confocal Microscopy*, 2nd ed., J. B. Pawley, Ed., pp. 111–126, Springer Science, New York (1995).

34. T. D. Visser, J. L. Oud, and G. I. Brakenhoff, "Refractive index and axial distance measurements in 3-D microscopy," *Optik* **90**, 17–18 (1990).
35. J. L. Bentley, C. Glazowski, and J. M. Zavislan, "Achromatic immersion objective for in-vivo imaging for low coherence confocal microscopy," *Proc. SPIE* **6342**, 63420G (2006) [doi: 10.1117/12.692204].
36. Y. Kanai, Y. Kanzaki, M. Wakaki, and N. Takeyama, "Design and development of multi functional confocal laser scanning microscope with UV/VIS laser source," *Proc. SPIE* **5878**, 58781D (2005) [doi: 10.1117/12.616637].
37. P. M. Delaney and M. R. Harris, "Fiber optics in scanning optical microscopy," in *Handbook of Biological Confocal Microscopy*, 3rd ed., J. B. Pawley, Ed., Springer, New York (2006).
38. R. M. Verdaasdonk and C. Borst, "Optics of fibers and fiber probes," in *Optical-Thermal Response of Laser-Irradiated Tissue*, A. J. Welch and M. J. C. van Gemert, Eds., Plenum Press, New York (1995).
39. M. Kanai, "Condensing optical system, confocal optical system, and scanning confocal endoscope," U.S. Patent Application No. 20050052753 (2005).
40. D. L. Dickensheets and G. S. Kino, "Micromachined scanning confocal optical microscope," *Opt. Lett.* **21**, 764–766 (1996).
41. E. J. Seibel and Q. Y. L. Smithwick, "Unique features of optical scanning, single fiber endoscopy," *Lasers Surg. Med.* **30**(3), 177–183 (2002).
42. D. L. Dickensheets and G. S. Kino, "Silicon-micromachined scanning confocal optical microscope," *J. Microelectromech. Syst.* **7**, 38–47 (1998).
43. H. Ra, Y. Taguchi, D. Lee, W. Piyawattanametha, and O. Solgaard, "Two-dimensional MEMS scanner for dual-axes confocal in vivo microscopy," *IEEE Int. Conf. Microelectromech. Sys.*, Istanbul, Turkey, 862–865 (2006).
44. K. C. Maitland, H. J. Shin, H. Ra, D. Lee, O. Solgaard, and R. Richards-Kortum, "Single fiber confocal microscope with a two-axis gimbaled MEMS scanner for cellular imaging," *Opt. Express* **14**, 8604–8612 (2006).
45. H. Miyajima, N. Asaoka, T. Isokawa, M. Ogata, Y. Aoki, M. Imai, O. Fujimori, M. Katashiro, and K. Matsumoto, "A MEMS electromagnetic optical scanner for a commercial confocal laser scanning microscope," *J. Microelectromech. Syst.* **12**, 243–251 (2003).
46. E. J. Seibel, Q. Y. J. Smithwick, C. M. Brown, and P. G. Reinhall, "Single fiber flexible endoscope: general design for small size, high resolution, and wide field of view," *Proc. SPIE* **4158**, 29–39 (2001) [doi: 10.1117/12.413801].
47. A. Osdoit, M. Genet, A. Perchant, S. Loiseau, B. Abrat, and F. Lacombe, "In vivo fibered confocal reflectance imaging totally non-invasive morphological cellular imaging brought to the endoscopist," *Proc. SPIE* **6082**, 608208 (2006) [doi:10.1117/12.646659].

48. A. F. Gmitro and D. Aziz, "Confocal microscopy through a fiber-optic imaging bundle," *Opt. Lett.* **18**, 565–567 (1993).
49. C. P. Lin and R. H. Webb, "Fiber-coupled multiplexed confocal microscope," *Opt. Lett.* **25**, 954–956 (2000).
50. P. M. Lane, A. L. P. Dlugan, R. Richards-Kortum, and C. E. MacAulay, "Fiber-optic confocal microscopy using a spatial light modulator," *Opt. Lett.* **25**, 1780–1782 (2000).
51. M. C. Chidley, K. D. Carlson, R. R. Richards-Kortum, and M. R. Descour, "Design, assembly, and optical bench testing of a high-numerical-aperture miniature injection-molded objective for fiber-optic confocal reflectance microscopy," *Appl. Opt.* **45**, 2545–2554 (2006).
52. C. Liang, M. R. Descour, K. Sung, and R. Richards-Kortum, "Fiber confocal reflectance microscope (FCRM) for in-vivo imaging," *Opt. Express* **9**, 821–830 (2001).
53. C. Liang, K. Sung, R. Richards-Kortum, and M. R. Descour, "Design of a high-numerical-aperture miniature microscope objective for an endoscopic fiber confocal reflectance microscope," *Appl. Opt.* **41**, 4603–4610 (2002).
54. J. Knittel, L. G. Schnieder, G. Buess, B. Messerschmidt, and T. Possner, "Endoscope-compatible confocal microscope using a gradient-index lens system," *Opt. Commun.* **188**, 267–273 (2001).
55. K. Carlson, M. Chidley, K. Sung, M. Descour, A. Gillenwater, M. Follen, and R. Richards-Kortum, "In vivo fiber-optic confocal reflectance microscope with an injection-molded plastic miniature objective lens," *Appl. Opt.* **44**, 1792–1797 (2005).
56. A. R. Rouse, A. Kano, J. A. Udovich, S. M. Kroto, and A. F. Gmitro, "Design and demonstration of a miniature catheter for a confocal microendoscope," *Appl. Opt.* **43**, 5763–5771 (2004).

Chapter 8

Endoscope Optics

Endoscopes are used to observe otherwise inaccessible areas within the human body either noninvasively or minimally invasively. Endoscopes have unparalleled ability to visualize lesions within internal organs with high resolution through natural body orifices, such as the mouth, nose, anus, and urethra. With high-resolution endoscopes, harvesting biopsy samples for later laboratory analysis may not be necessary.

The history and basic optics of endoscopes will be discussed in Secs. 8.1 and 8.2. Sections 8.3, 8.4, and 8.5 will discuss the optics and optical design of relay lenses, objective lenses, and illumination systems. Section 8.6 will focus on the imaging and illumination systems of wireless endoscopes.

8.1 Introduction

The earliest endoscope-using lens was called the cystoscope. It was invented by Maximilian Nitze in 1877 and was used to examine the interior of the urinary bladder through the urethra.¹ In 1932, G. Wolf introduced the first semiflexible gastroscope, and in 1956, Basil Hirschowitz performed the first image transmission with fiber bundles. In 1959, Harold H. Hopkins invented rod lenses for image transmission, and in 1963, Karl Storz combined rod lenses for image transmission with fiber bundles for illumination.² A video endoscope with a camera on the eyepiece was first introduced in 1987, and a video endoscope with a camera at the distal end of the endoscope was demonstrated in 1992.¹ In the late 1990s, with the advances in light sources, sensors, and electronics, a wireless endoscope was invented for small-bowel imaging.³

Traditional endoscopes comprise an airtight and waterproof elongated tube having a distal end with an objective lens for imaging and a proximal end with an eyepiece for viewing. The elongated tube includes a relay lens system, or a fiber bundle, to transmit the image formed by the objective lens to the proximal end of the tube. The function of the eyepiece is to magnify the image at the proximal end for the observer. Endoscopes used in clinical applications further consist of a work channel and an irrigation channel.

In the last five decades, the introduction of optical fibers, rod lenses, and then the electronic detector has advanced the development of endoscopes dramatically. Various endoscopes have been developed for general applications and to meet some special requirements. Some commonly used endoscopes are:

- Arthroscopes to examine joints,
- Bronchoscopes to examine air passages and lungs,
- Colonoscopes to examine the colon,
- Cystoscopes to examine the urinary bladder,
- Gastrosopes to examine the small intestine, stomach, and esophagus, and
- Hysteroscopes to examine the uterus.

8.2 Basic Optics for Endoscopes

In some form or another, all endoscopes use optical elements to guide light to a target and transfer an image to the eye or to a detector. The basic optical elements of an endoscope include an illumination system, an imaging system, an image transmission system (relay system), and a viewing system (eyepiece or electronic sensor).

Figure 8.1 is a diagram of a conventional optical system in a rigid endoscope where, for simplicity, only one relay stage is shown. The system consists of three basic and separate optical components, as follows:

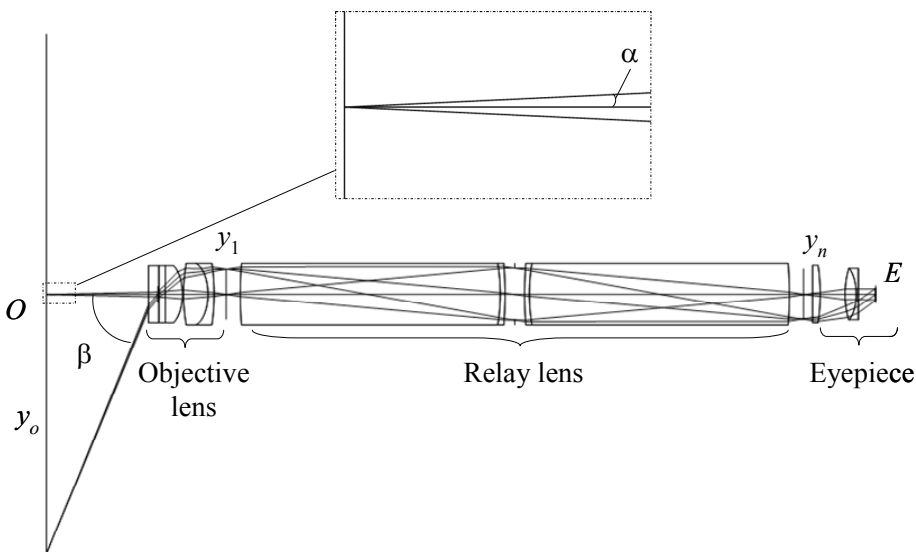


Figure 8.1 Typical optical layout of a conventional endoscope with a one-stage relay lens. It consists of three lenses: the objective lens, relay lens, and eyepiece. Here y_o is the height of the field, y_1 is the first intermediate image height, and y_n is the final image height through the relay lens. E is the exit pupil of the eyepiece, and it is the location for placement of the observer's pupil.

- An objective lens, which forms the first inverted intermediate image y_1 of the object y_o .
- A relay lens system, which reimages the first intermediate image y_1 to the final image y_n at the proximal end of the endoscopic tube. In most endoscopes, there are several unit-magnification relay lenses forming the intermediate images y_2, y_3, \dots , and the final image y_n .
- An eyepiece or a focusing lens, which presents the final image y_n to a sensor. Traditionally, an eyepiece is attached to the endoscope and produces a magnified virtual image for the observer. In modern endoscopes, a camera lens is attached to the endoscope, and it produces a real image on the electronic sensor. The image is then displayed on a monitor or other display device.

According to optical transmission systems, we can classify endoscopes into three groups: rigid, fiber optic, and video. In recent years, some new types of endoscopes, such as wireless, scanning, and stereo endoscopes, have been developed.

8.2.1 Illumination and imaging optics

Figure 8.2 illustrates a basic optical configuration at the distal end of an endoscope. The light from the fiber bundle illuminates the observation region. Part of the reflected light from the tissue is captured by the objective lens.

There are two major obstacles to designing an endoscope having a small diameter. The first obstacle is the lack of sufficient illumination. An endoscope with a small diameter does not provide enough space to transmit light. The second obstacle is light collection efficiency. The aperture is small because of the small diameter of the endoscope, which results in a low light collection efficiency.

The amount of light captured by the eye or an electronic sensor is principally determined by three factors:

- The intensity of the light incident upon the observation region,
- The optical characteristics of the observation surface, such as surface reflectivity and curvature, and
- The light collection efficiency of the objective lens and the transmission of the optical system.

The brightness of the image seen in an endoscope depends on the level of illumination falling on the region being viewed, as discussed below.²

The total amount of light captured by the endoscope from a uniformly diffuse reflecting surface is given by

$$\Phi = LR(\pi y_o^2)(\pi \sin^2 \alpha), \quad (8.1)$$

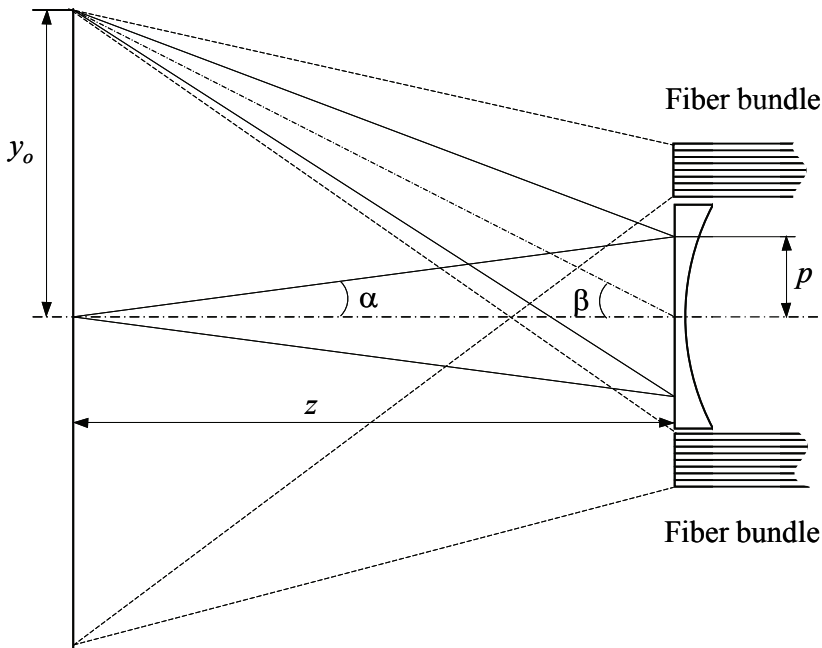


Figure 8.2 Basic optical configuration at the distal end of an endoscope. The light from the fiber bundle illuminates the tissue. Part of the reflected light is captured by the objective lens.

where L is the radiance from the fiber bundle, R is the reflection coefficient, y_o is the object height, and πy_o^2 is the area of the object, assuming that the FOV is circular, α is the cone angle of the light accepted by the objective lens, and $\pi \sin^2 \alpha$ measures the on-axis solid angle of the cone of rays accepted by the endoscope. All of the parameters are shown in Figs. 8.1 and 8.2.

The cone angle α of the light accepted by objective lens is approximately determined by the radius p of the entrance pupil of the objective lens and the working distance z :

$$\sin \alpha = \frac{p}{z}. \quad (8.2)$$

Therefore, Eq. (8.1) becomes

$$\Phi = \pi^2 L R y_o^2 \left(\frac{p}{z} \right)^2. \quad (8.3)$$

The total amount of light accepted by the endoscope is proportional to the radiance L and inversely proportional to the square of the working distance z .

In object space, the optical invariant is defined as

$$H = n \sin(\alpha) y_o, \quad (8.4)$$

where n is the refractive index of the medium in object space. Equation (8.1) becomes

$$\Phi = \frac{\pi^2 LR}{n^2} H^2 \quad (8.5)$$

for the given illumination and the object distance. The total amount of light entering the endoscope is proportional to H^2 .

8.2.2 Rigid endoscopes

As the name suggests, rigid endoscopes have a rigid tube to house the refractive relay lenses and illumination fibers. The relay lenses transfer the image at the distal end to the proximal end of the tube so that the image can be directly viewed through an eyepiece or with an imaging detector. When the endoscope is designed for direct view through the eyepiece, an odd number of relay stages is required in order to correct for the inversion produced by the object because the eyepiece has a positive magnification and does not produce an inverted virtual image. When an electronic sensor is used in an endoscope, the parity of the number of relay stages is not important since the inversion can always be performed electronically or through software.

As seen from Fig. 8.1, the relay system determines the value of the optical invariant H of the entire imaging system. In object space, the optical invariant is determined by Eq. (8.4). While in the plane of the entrance pupil, the optical invariant is

$$H = n \sin(\beta) p, \quad (8.6)$$

where β is the field angle and p is the radius of the entrance pupil, as shown in Fig. 8.2.

Given the H of the relay system and the FOV, either the object height y_o or the field angle β and the semiangle of the on-axis marginal ray α or the radius of the entrance pupil p can be obtained from Eqs. (8.4) and (8.6). Once the FOV is determined, the size of the relay system governs the focal length f of the objective lens by

$$f \tan \beta = r, \quad (8.7)$$

where r is the radius of the clear aperture of the relay lens.

The depth of field of the rigid endoscope is determined by the viewing configuration. When an eyepiece is mounted at the proximal end of the endoscope, the accommodation of the human eye helps to extend the DOF. The normal human eye can focus from infinity to a near point, which is 250 mm from

the eye. Therefore, the depth of field is calculated by the difference between two working distances: the working distance of the objective lens when the final image is located 250 mm from the exit pupil of the eyepiece and the working distance when the image is at infinity.

When a CCD or CMOS sensor is used at the proximal end of an endoscope, the depth of field is smaller than that of an endoscope using an eyepiece. The depth of focus can be calculated from the requirement of the resolution and the size of the pixels in the CCD or CMOS sensor, as discussed in Chapter 2.

One special requirement for objective lenses used in rigid endoscopes is the telecentricity that prevents light loss during image transmission from the distal end to the proximal end. As shown in Fig. 8.3(a), when the objective lens is not telecentric, the off-axis ray may miss the relay lens, for example, ray 1, or it cannot be transferred to the next intermediate image plane, such as ray 2, because it hits the edge of the relay lens. The ray may be absorbed by the lens edge or may be reflected back to the relay lens. Generally, rays reflected from the edge of the relay lens degrade image quality and should be minimized. The objective lens

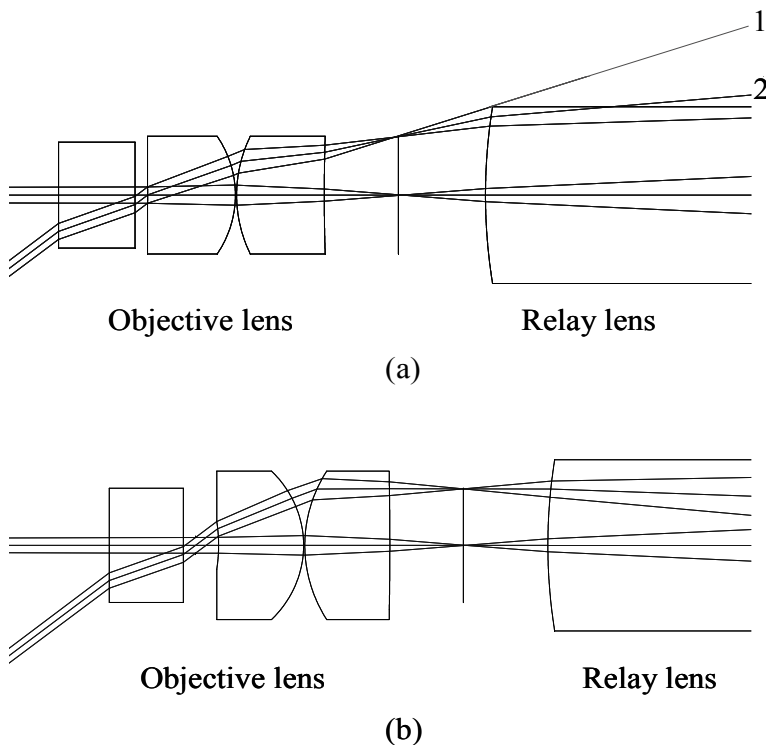


Figure 8.3 (a) Ray diagram of a nontelecentric objective lens. Rays 1 and 2 cannot be relayed to the next intermediate image. (b) Ray diagram of a telecentric objective lens. All of the rays are relayed to the next intermediate image.

in Fig. 8.3(b) is telecentric in the image space where the chief ray is parallel to the optical axis. All of the off-axis rays enter the relay lens and can be transferred to the next intermediate image.

There are three types of relay systems used in rigid endoscopes, namely, traditional lenses, rod lenses, and GRIN lenses. Section 8.3 will discuss the properties and design of each relay system.

8.2.3 Flexible endoscopes

Flexible endoscopes use imaging fiber bundles to transfer the image from the distal end of the endoscope to the imaging lens or eyepiece, as shown in Fig. 8.4. With fiber bundles, the diameter of the space required for the image relay is reduced significantly, leaving more space for other instrument channels. Other advantages of fiber-optic endoscopes include transmitting the image over long distances and observing around corners.

An imaging fiber bundle typically contains about 3000 or more optical fibers packed into a hexagonal array. The spatial arrangement of the fibers at both ends is identical, enabling a spatially coherent transmission of the image. The optical fibers are multimode and preserve the color of light. The core diameter is typically $3\ \mu\text{m}$, the overall diameter of the imaging bundle varies between 0.2 and 3 mm, and the fiber bundle may be over 2 m in length.

The resolution of a flexible endoscope is limited by the core diameter of the fiber bundle. Because of the pixelated nature of the fiber bundle, flexible endoscopes have a smaller depth of field than the conventional endoscopes with refractive relay lenses. The calculation of the depth of field is discussed in detail in Chapter 3. The telecentricity of the objective lens on the fiber side is required in order to couple more light into and out of the fiber bundle.

Unlike endoscopes with refractive relay lenses, where the eyepiece can be designed to compensate for the residual aberrations accumulated from the objective and relay lenses, the eyepiece in fiber endoscopes can only give a sharp image of what appears at the proximal end of the fiber bundle, and it cannot

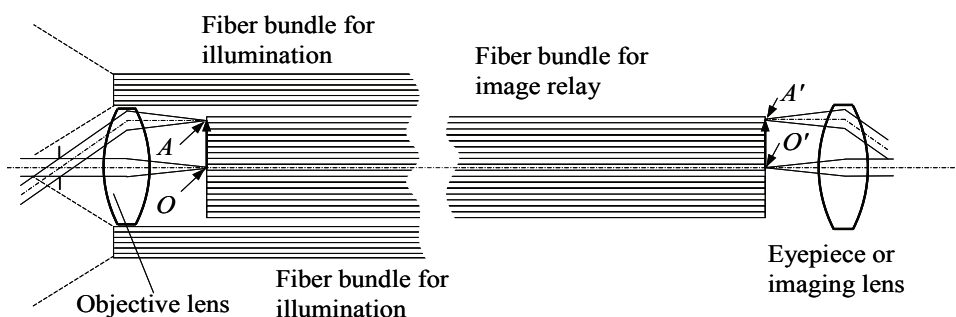


Figure 8.4 Optical layout of a typical fiber-optic endoscope.

correct the blur of the image formed at the distal end by the objective lens and transmitted by the fiber bundle.

8.2.4 Video endoscopes

One of the most important advances in endoscopes is the introduction of video-chip technology on the distal end of the device. Image relay is not necessary given that the objective lens images the observation surface directly to the sensing elements in the imaging detector. As shown in Fig. 8.5, the imaging head of a video endoscope generally consists of an imaging detector, objective lens, and illumination optics. Two types of imaging detectors, CCD and CMOS, are commonly used in video endoscopes.

Video endoscopes have revolutionized many surgical procedures by improving visualization while minimizing the tissue damage that is generally caused by invasive approaches. The advantages of a video endoscope over a fiber-optic endoscope include:

- The number of pixels in an imaging detector is typically greater than the number of fibers in a fiber bundle, giving a sharper image.
- In a conventional fiber-optic endoscope, the hexagonal sampling, which is due to the arrangement of the fibers, is followed by a rectangular sampling through the imaging detector. The mismatch between these two sequential sampling operations makes processing fiber-optic endoscope images difficult.

The objective lens of a video endoscope must have a short structural length, a small diameter, and a wide field angle. Compared to objective lenses used in rigid and flexible endoscopes, telecentricity is not a requirement for lenses in video endoscopes, even it is desired to obtain uniform QE across the field of view. The incident angle of the chief ray on the image surface increases in proportion to the field. The chief ray angle should be smaller than the specification of the imaging detector for a reasonable QE.

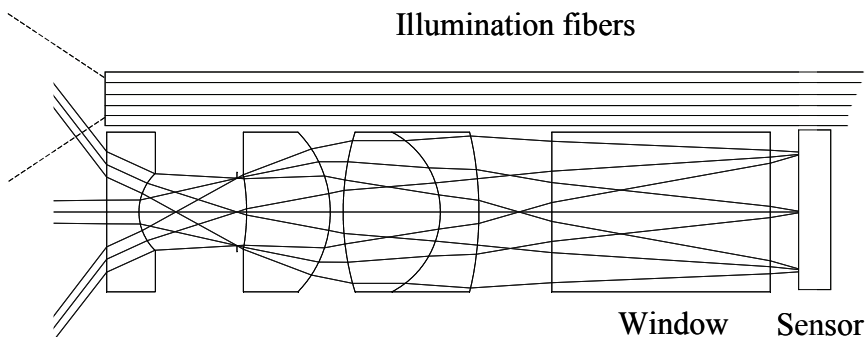


Figure 8.5 Optical layout of a video endoscope.

Another requirement of the objective lens in a video endoscope, compared to the objective lenses in fiber-optic and rigid endoscopes, is that the back working distance should be longer. This increase is necessary because of the window glass and the distance between the window and the sensing pixels. When the size of a sensor does not allow the objective lens to be placed perpendicular to the optical axis, a folding mirror or prism is required so that the sensor can be placed parallel to the optical axis of the objective lens. Because of the limited number of pixels in a small sensor, the objective lens for a video endoscope should have an RMS spot size smaller than the size of the pixels in order to obtain adequate resolution.

The resolution and sensitivity of imaging detectors are constantly improving. Smaller sensors with higher resolutions will provide the opportunity to design smaller endoscopic probe heads and to place two or more sensors into probe heads for stereoscopic imaging.

8.3 Relay Lenses

Rigid endoscopes require the use of a relay system to transfer an image through a distance inside a tube of limited diameter to the proximal end of the tube where the eyepiece or imaging lens is located. Depending on the tube length, generally four or more relay stages are necessary to transfer the image from the distal end to the proximal end. The brightness of an image in a rigid endoscope is ultimately determined by the optical invariant of the relay lens, which is proportional to the NA and the diameter of the relay lenses. Because of the various constraints related to the relay lenses, such as size, multiple relay stages, and NA, the relay system should be the starting point in designing rigid endoscopes.

Two optical parameters are important for understanding relay systems: lateral magnification m and optical invariant H . Magnification m determines how aberrations are transferred from one intermediate image plane to the next. Longitudinal aberrations, such as spherical aberration, are transferred to the next image plane with a factor of m^2 . When $|m| < 1$, longitudinal aberration can be reduced in each stage but the NA will increase, making the aberration correction more challenging; therefore, the optimal lateral magnification is -1 . The major advantage of relay lenses with unit magnification is that the optical elements in a relay lens are symmetrical and identical to the aperture stop. A symmetrical layout with unit magnification has superior advantages for aberration correction, fabrication, and assembly. Odd aberrations, namely, coma, distortion, and lateral chromatic aberrations, from the lens group in front of the aperture stop are completely cancelled out by those from the lens group behind the aperture stop because odd aberrations are dependent on the odd power of the chief ray height.

Optical invariant H is a function of the refractive index n , aperture angle α , and image height y . It determines whether a relay system fits the objective lens,

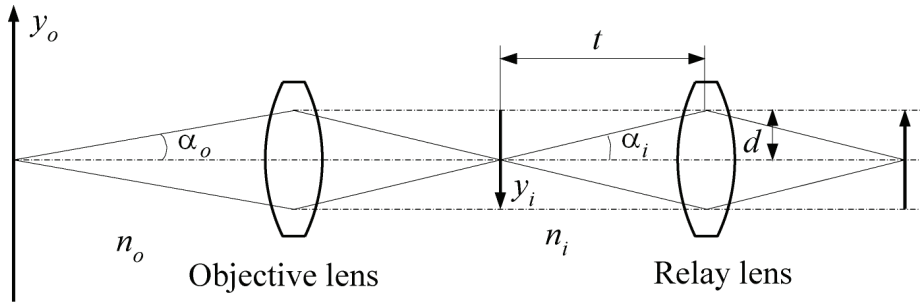


Figure 8.6 Simplified diagram of an objective lens and one stage of a relay system.

eyepiece, or coupling lens. In order to transport all of the information in the image formed by the objective lens, the optical invariants of the objective lens and the relay system must match. As shown in Fig. 8.6, H for the relay system is determined by

$$H = n_o \sin(\alpha_o) y_o = n_i \sin(\alpha_i) y_i, \quad (8.8)$$

where n_o and n_i are the refractive indices, α_o and α_i are the marginal ray angles in the object and intermediate image spaces, respectively, and y_o and y_i are the object height and intermediate image height, respectively. To maximize the H of the relay system and therefore the light transmitting capability, both $\sin(\alpha_i)$ and y_i must be at their maximum values, which are (d/t) and d , where t is the distance between the intermediate image and entrance pupil of the relay lens and d is the clear radius of the relay lens. Therefore,

$$H = n_i \sin(\alpha_i) y_i = n_i \left(\frac{d}{t} \right) d = n_i \frac{d^2}{t}, \quad (8.9)$$

and Eq. (8.5) becomes

$$\Phi = \left(\frac{\pi^2 LR}{n_o^2} \right) H^2 = \left(\frac{\pi^2 LR}{n_o^2} \right) \left(n_i \frac{d^2}{t} \right)^2 = \left(\frac{\pi^2 LR}{n_o^2 t^2} \right) n_i^2 d^4. \quad (8.10)$$

This equation gives the theoretical maximum total light that can be transmitted by the relay system. This maximum is proportional to the square of the refractive index in the intermediate image space and to the fourth power of the clear radius of the relay lens.

The resulting image brightness will be further reduced by the absorption in the optical materials and unwanted reflection losses at the interfaces of the optical elements. Therefore, in addition to applying an AR coating to each

surface, reducing the number of air/glass interfaces can also improve the image brightness.

To design an efficient relay system, one has to understand the optical properties of each type of relay lens, namely, the conventional relay lens, rod lens, and GRIN lens. In addition, one must consider the lens diameter, NA, length of a single relay, lens complexity, image quality, and cost. Given that most endoscopes have multiple relay stages, in order to transmit the light more efficiently by minimizing the vignetting, the symmetry of the system requires that each stage be telecentric in both its object and image spaces.

The following section will discuss the design principles of relay systems. To simplify the discussion, only a single relay stage will be discussed.

8.3.1 Conventional relay systems

In order to meet the double-telecentric requirement, at least two elements, the relay objective lens and the field lens, are needed at each half stage of a conventional relay system, as shown in Fig. 8.7. The lenses close to the aperture stop are relay objective lenses, and the two lenses near the object and the image planes are field lenses. The relay objective lens provides the first-order requirement, such as the magnification and the NA. The field lens helps achieve the double-telecentric condition.

For endoscopes used over a broad spectrum, such as in white-light observation, the relay lens needs to be achromatized. In a conventional relay system, achromatization is best achieved by splitting the relay objective lens into a doublet. A doublet can also reduce the spherical aberration and compensate for the small amount of axial chromatic aberration introduced by the field lenses if the lenses are displaced from the intermediate image planes.

Figure 8.8(a) shows a conventional relay lens with a 1-mm object height, 0.1 NA, 45-mm total track, and 2 mm between the intermediate image plane and the lens surface. The diameter of the relay system is 2.7 mm. For purposes of comparison, the same object height, NA, lens diameter, and total track will be used for other relay systems in this section.

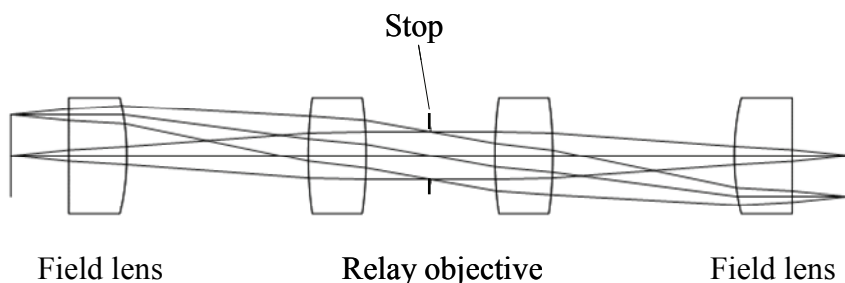


Figure 8.7 One stage of a conventional telecentric relay system.

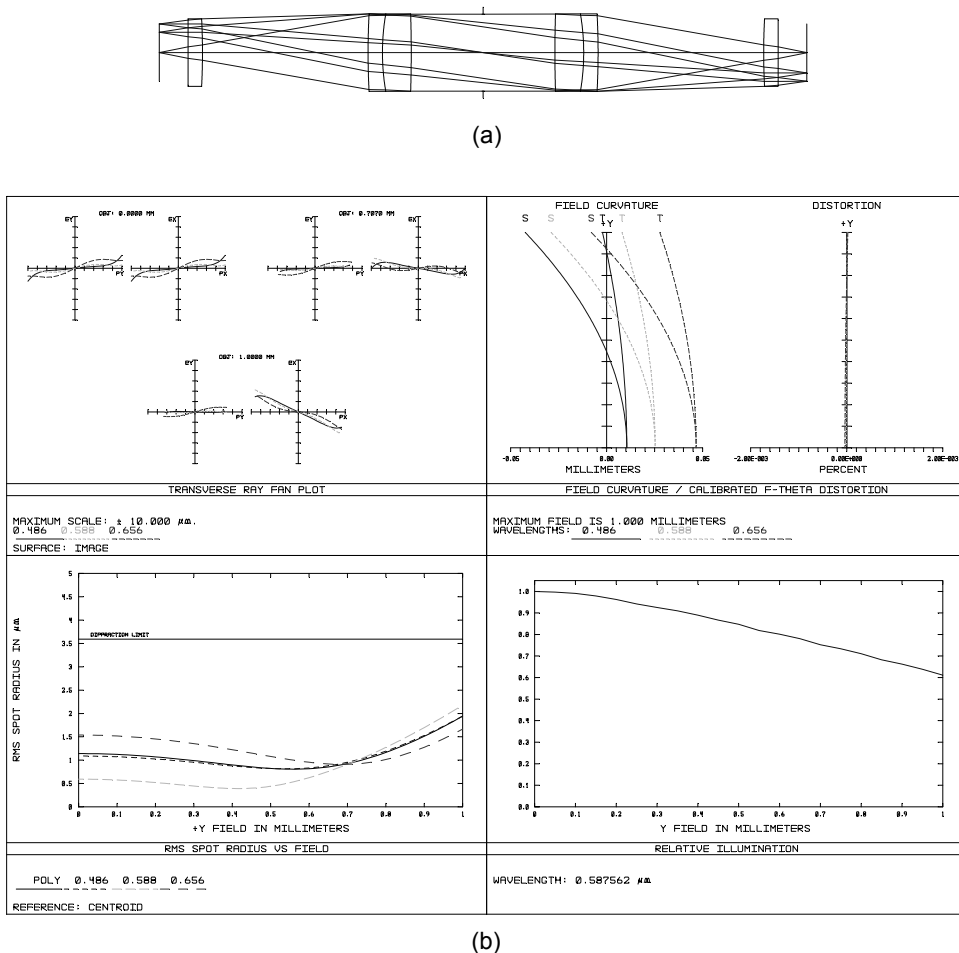


Figure 8.8 Conventional relay system: (a) system layout and (b) system performance: transverse ray fan plot, field curvature/distortion, RMS spot radius, and relative illumination.

Conventional relay systems generally consist of lenses of positive power only, and hence, such systems generate a significant positive field curvature. The spherical and axial chromatic aberrations are corrected by the use of an appropriately designed doublet. Coma, distortion, and lateral chromatic aberration are canceled completely because of the symmetry of the relay system, as shown in the performance plots in Fig. 8.8(b). The RMS spot radius is smaller than the Airy radius. However, there is a relatively large amount of vignetting, approximately 33%. The plot of the relative illumination reflects the light loss from the vignetting. By replacing the plano-convex field lens with a meniscus lens, the field curvature and astigmatism can be significantly reduced.⁴

As a result of the difficulties in loading thin lenses into a long, rigid tube, the lenses are generally mounted in a thin-walled tube that can be slid into a long,

rigid outer tube. Therefore, the clear aperture of the conventional relay lens system is relatively small, resulting in a low light throughput, as predicted by Eq. (8.10).

8.3.2 Hopkins rod-lens relay

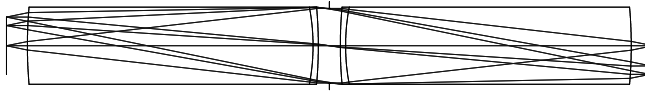
In late 1950s and early 1960s, H. H. Hopkins developed a rod-lens relay system.^{2,5} In contrast to the conventional relay system where the medium between the field lens and the relay objective lens is air, a Hopkins rod-lens relay system has a glass rod between the relay objective lens and the field lens.

Once the outer diameter of the rigid tube and the number of relay stages are determined, as indicated in Eq. (8.10), we can increase the light throughput by increasing either the refractive index n_i or the clear radius of the relay lens d . The Hopkins rod lens, which has a rod lens between the relay objective lens and field lens, can increase both the refractive index and the clear aperture of the relay lens.

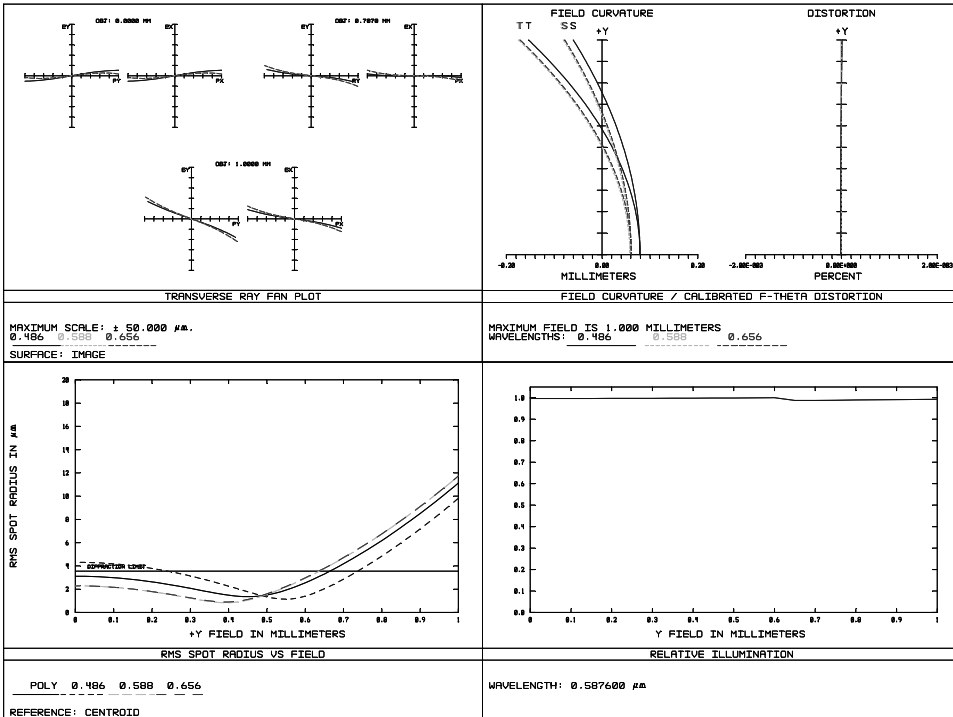
In the Hopkins rod-lens relay system, as shown in Fig. 8.9(a), the roles of the glass and air are interchanged.⁵ Compared to the conventional relay, where $n_i = 1$, the light throughput of the rod-lens relay increases by a factor of n_i^2 . If BK7 is used as a rod lens, the light transmitted is increased by a factor of $n_i^2 = 1.5163^2 = 2.2992$. Another advantage of the rod lens relay system is that the rod lens permits a greater diameter for a given outer diameter of a rigid tube. To mount the small and thin lenses into a long tube, another short and thin tube must be used to mount the lenses first, and then the short thin tube is inserted into the long tube. With a long rod lens, no additional short tube is necessary; therefore, the diameter of the lens can be larger. Considering that the light transmitted Φ is proportional to the fourth power of the clear radius of the relay lens, even a slight increase in the clear aperture will increase the throughput dramatically. Therefore, the effect of using glass spaces and air lenses in place of air spaces and glass lenses is twofold.

The relay lens in Fig. 8.9(a) is a scaled version of the lens in Hopkins' patent.⁵ It has a 2-mm FOV, 0.1 NA, 2.7-mm lens diameter, and a total track of 45 mm.⁵ Each half has identical doublets consisting of a thin negative lens and a long positive rod to correct chromatic aberrations. The performance of this relay lens is plotted in Fig. 8.9(b). The RMS spot radius is smaller than the Airy radius for the field up to 65% of the FOV. As a result of the symmetrical configuration, coma, distortion, and lateral chromatic aberration are essentially zero. One significant improvement in the Hopkins rod-lens relay is the negligible vignetting.

The Hopkins rod lens employs only two positive doublets; therefore, it suffers from the effects of astigmatism and field curvature, which may significantly degrade the off-axis image quality. The large amount of astigmatism is the result of fewer surface curvatures available for aberration correction. The small amount of positive astigmatism from the field lens is balanced by negative



(a)



(b)

Figure 8.9 Hopkins rod lens relay system. (a) System layout and (b) system performance: transverse ray fan plot, field curvature/distortion, RMS spot radius, and relative illumination.

astigmatism from the cemented surface of the cylindrical rod. The accumulated astigmatism and field curvature in the relay system may be compensated for, to some extent, by the endoscope objective lens.

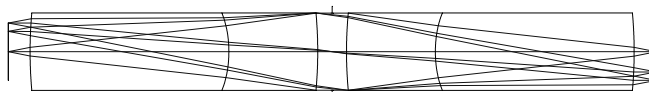
As shown in Fig. 8.9 and other examples using rod lenses in the next section, the rod lens is thick with curved surfaces at both ends, possibly making it problematic to manufacture. An alternative is to split the rod lenses into two plano-convex lenses and one flat-rod lens and cement them together.

8.3.3 Variations of the Hopkins rod lens

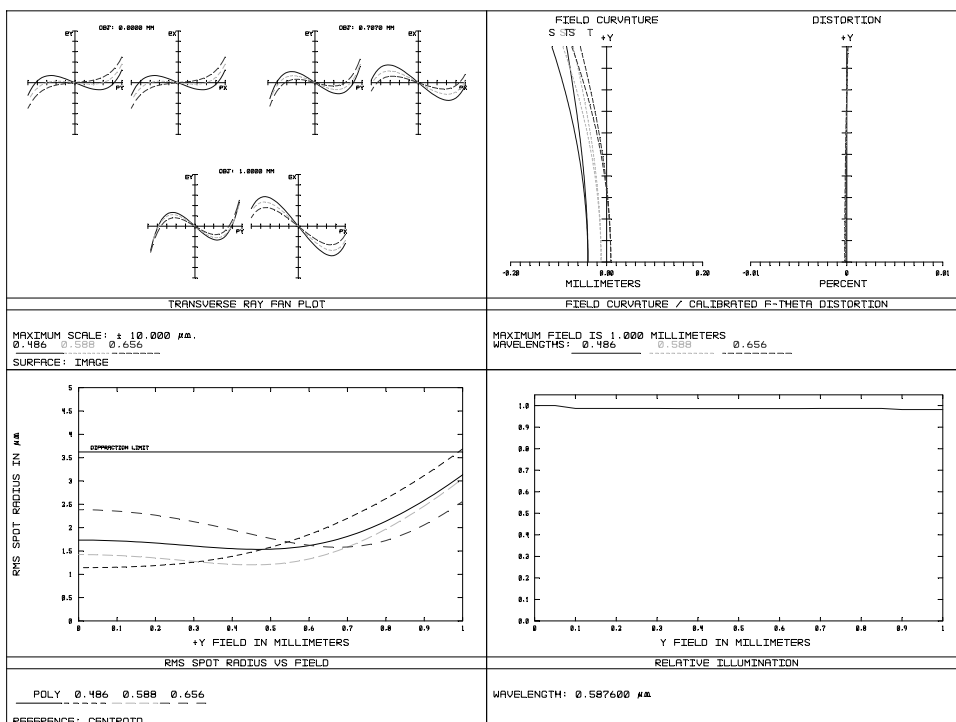
If the negative lenses in the cement close to the aperture become thicker, astigmatism of the relay system is reduced significantly. Figure 8.10 shows a relay lens with a thick negative lens. Compared to a traditional rod lens relay, the

RMS spot size is smaller than the Airy radius over the entire field. The falloff of the relative illumination is very small. There is some axial chromatic aberration, but the field curvature is the major residual aberration.

Figure 8.11 shows another variation of the Hopkins rod-lens relay system.⁶ The relay lens includes two identical meniscus lenses cemented to the opposite side of a biconvex rod lens. With an additional cemented surface, this relay lens has a better correction of the axial chromatic aberration, but the overall performance is similar to that the previous design. The distance between the object plane and the first lens surface is the same as the distance from the last surface of the front group to the aperture stop. This design has the advantage of reducing assembly and manufacturing costs because the cemented lenses and the spaces between the lenses are identical.



(a)



(b)

Figure 8.10 Rod-lens relay with thick negative lenses. (a) System layout and (b) system performance: transverse ray fan plot, field curvature/distortion, RMS spot radius, and relative illumination.

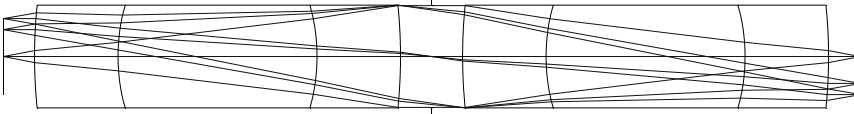


Figure 8.11 Rod-lens relay with two negative lenses.

For the relay lenses discussed above, the spherical aberration, coma, distortion, and lateral chromatic aberration are corrected favorably. However, field curvature and astigmatism are not corrected satisfactorily. When the number of relays increases, field curvature and astigmatism accumulate. As a result, it is impossible to focus on the entire field at the same time.

Most refractive lens systems have a considerable amount of field curvature. Field curvature is related to the Petzval sum,

$$P = \sum \frac{\varphi}{n}, \quad (8.11)$$

where φ is the power of the refractive surface with refractive index n . In the absence of astigmatism, the field curvature is equal to $-P$. Generally, there are three basic methods for correcting the field curvature for relay lenses, namely, spacing, bending, and index difference.

Spacing is based on the simple idea that an optical system, which comprises a positive lens and a negative lens with equal but opposite optical powers and the same refractive index, has a Petzval sum equal to zero, and the overall power of the system is determined by the separation of the elements. Bending is a common technique used to reduce aberration. When an element is bent such that it has equal curvatures at its opposite surfaces, the Petzval sum is zero, and the overall power is provided by the separation of the two surfaces. Index difference utilizes the difference between the refractive indices of the positive and negative elements. Generally, a combination of three methods is used to correct field curvature.

Figure 8.12(a) shows a relay lens with two additional negative lenses between the cemented rod lenses. The two negative lenses generate some negative Petzval curvature and astigmatism. As shown in Fig. 8.12(b), the whole relay system has a much smaller astigmatism, and the Petzval curvature is reduced by approximately 50% compared to the relay lens in Fig. 8.10. Compared to other configurations of rod-lens relays, this variation has 17% vignetting at the edge of the field, which is relatively large.⁷

Figure 8.13(a) shows one example of a relay system using index difference to reduce field curvature. Each half of the relay lens consists of two identical doublets and a rod lens.⁸ The doublets are so-called “new achromats,” where the refractive index of the positive lens is higher than that of the negative lens. This

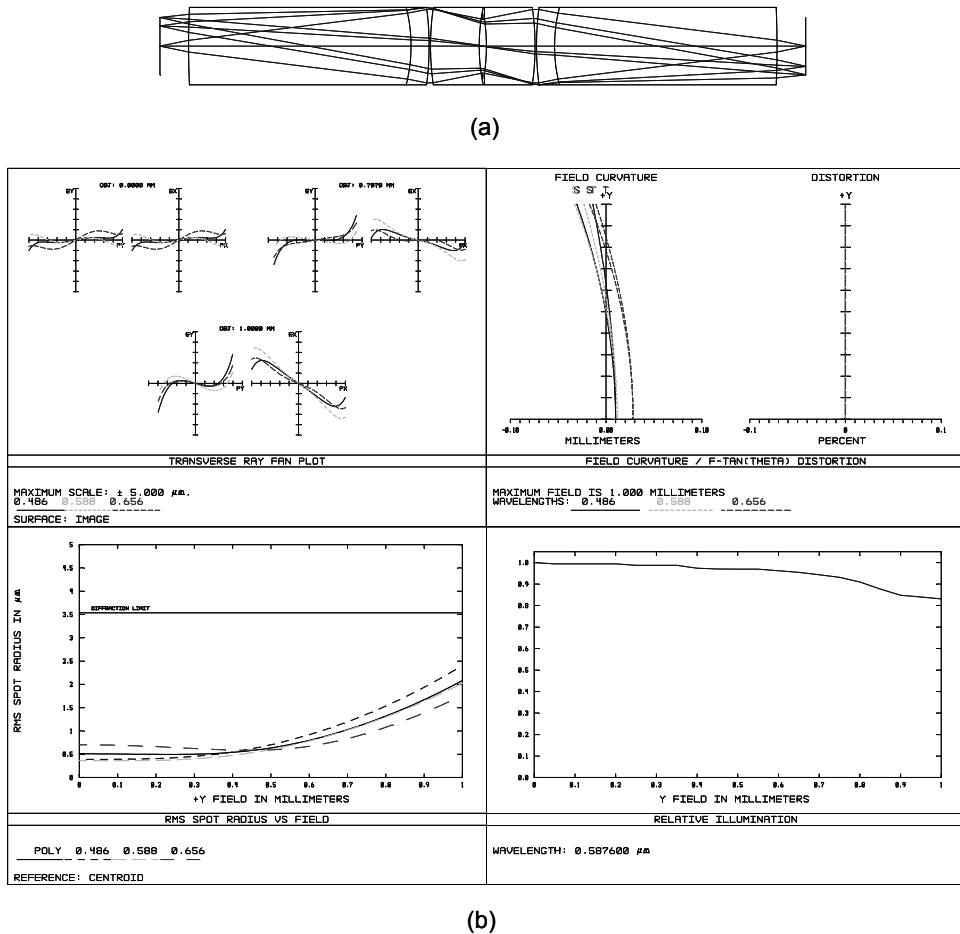


Figure 8.12 Rod-lens relay with two negative lenses in the relay objective lens. (a) System layout and (b) system performance: transverse ray fan plot, field curvature/distortion, RMS spot radius, and relative illumination.

relay lens has an excellent correction of field curvature. However, it may have some vignetting at the edge of the field, limiting the FOV. Another issue with this design is that it requires as many as eight lens elements to compose a relay unit, and it has twelve air/glass surfaces, thereby increasing the manufacturing cost and reducing the total transmission.

To reduce the number of air/glass surfaces, the doublet and the rod lens can be cemented together, as shown in Fig. 8.13(b).⁸ This relay system consists of two identical halves arranged symmetrically. Each half has a rod lens and two identical doublets. The doublets are cemented to the rod lens on each side so that each half is symmetrical. Like the relay system in Fig. 8.13(a), the distance from

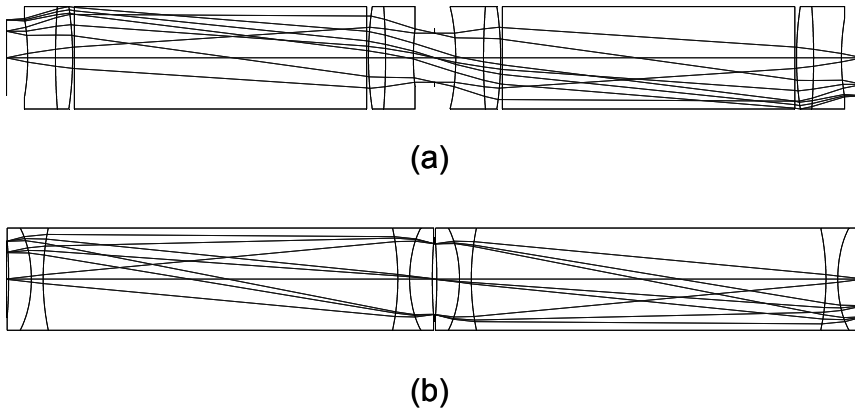


Figure 8.13 (a) Relay lens with reduced field curvature and (b) relay lens with reduced air/glass surfaces.

the object plane to the first surface is the same as the distance between the last surface of the front group and the aperture stop. There are only four air/glass surfaces in this design compared to twelve in the previous design. All of the aberrations are corrected satisfactorily with the help of one aspherical surface on each doublet. This configuration has some additional space to increase the NA and the FOV.

To avoid the risk of exposing patients to others' diseases, there is a need for disposable endoscopes. Relay lenses made of glass are expensive because of the cost of the materials themselves as well as the manufacturing cost. Since the late 1980s, relay systems have been developed with polymeric materials, such as acrylics, polystyrenes, polycarbonates, and others.⁹⁻¹² In order to increase image brightness, a high-refractive-index plastic rod lens with flat surfaces is placed in between the other polymeric lenses. Figure 8.14 shows one example of a disposable relay system. Each half of the relay lens comprises a polycarbonate rod with two flat surfaces and three other plastic elements made from acrylic and polystyrene. This relay lens has a reasonable aberration correction and has an acceptable performance in up to 70% of the field. The limiting aberrations are spherical aberration, field curvature, and astigmatism.

Figure 8.15 shows two more variations of a rod-lens relay. Figure 8.15(a) shows a nonsymmetrical relay lens.¹³ It consists of a doublet and two rod lenses, wherein a cemented lens is placed between the rod lenses. The odd aberrations, such as coma, distortion, and lateral chromatic aberration, are relatively small but not zero because the relay system is not completely symmetric. The main residual aberrations are spherical aberration and field curvature.

Figure 8.15(b) shows a three-stage relay system. Each of the first two stages consists of only two rod lenses, and the two rod lenses in the third stage have a cemented surface on each end. The axial chromatic aberrations in the first two stages are not corrected, and they are compensated as a whole by the third stage.

The aberrations of each group are not well controlled, but as a relay system, the aberrations are balanced.¹⁴

One commonly used relay lens is a GRIN lens. Generally, the NA of a GRIN lens is larger than that of a conventional relay system. GRIN lenses can be fabricated down to 0.2 mm in thickness or diameter. Plano lens surfaces, instead of spherical or aspherical surfaces, are typically used in GRIN lenses. Their simple geometry allows for cost-effective production and simplifies the assembly process. The optical invariant, and therefore the maximum brightness, of a GRIN-lens relay is larger than in conventional systems of equal diameter. GRIN lenses are typically fine ground on their outer surfaces to reduce stray reflection.

When using a GRIN lens, one has to know the available types of GRIN lenses; otherwise, an unmanufacturable design may result. Despite the cost

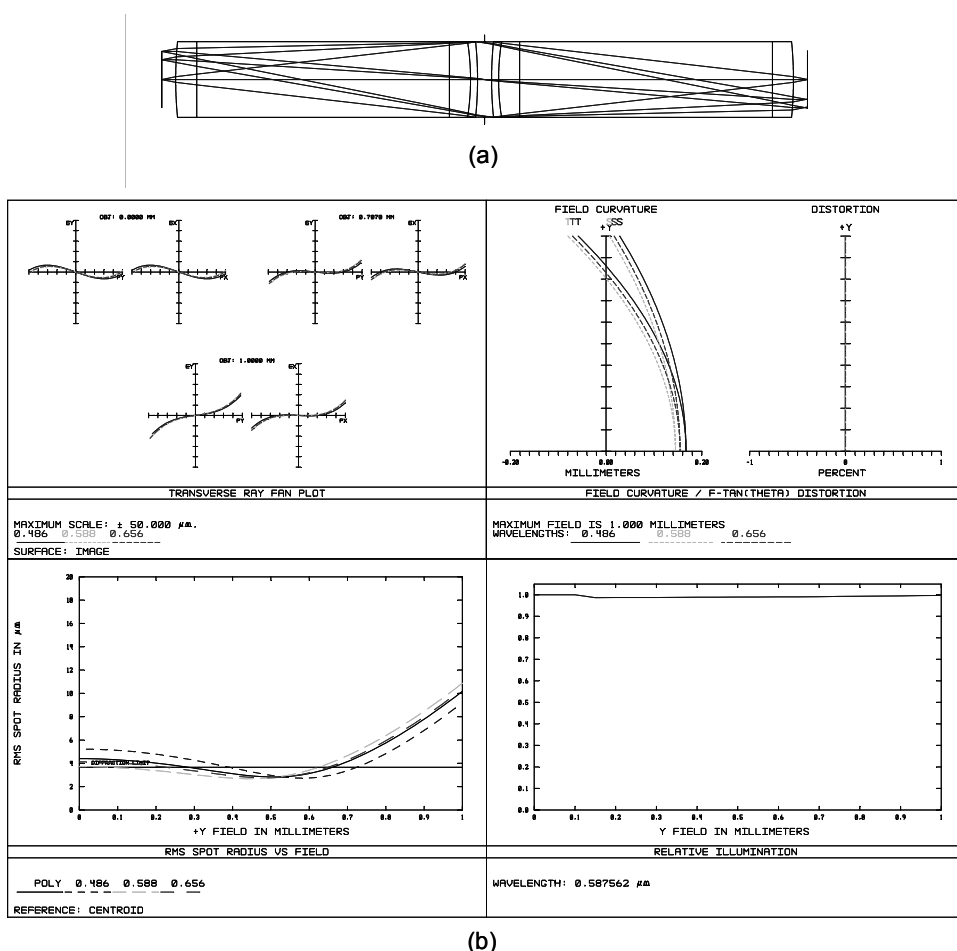


Figure 8.14 Disposable relay lens: (a) system layout and (b) system performance: transverse ray fan plot, field curvature/distortion, RMS spot radius, and relative illumination.

advantage and simplicity, GRIN-lens relay systems cannot replace the Hopkins configuration. The primary issue of GRIN-lens relays is that they suffer from excessive dispersion, making it impossible to produce a color-corrected relay system. The overcorrected axial chromatic aberration in the relay system may be compensated by the objective lens and the eyepiece or focusing lens.

Figure 8.16 shows a GRIN-lens relay (GRINTECH's GT-IFRL-200-inf-50cc). The numerical aperture in this example is set to 0.1. The overall performance is not as good as that of the rod-lens relay systems. The RMS spot size is larger than the Airy disk over the entire FOV. The axial chromatic aberration is much larger, but the lateral chromatic aberration is within the Airy radius.

The relay system in an endoscope transports the image over a long distance from the distal end to the proximal end with minimal image quality degradation. Of the three different types of relay systems discussed above, relay systems using Hopkins rod lenses are used most often because they have better image quality and higher throughput.

8.4 Objective Lenses

8.4.1 Introduction

In order to observe tissues inside the body without missing useful information, the objective lens must have a large FOV, for example, 100 deg, and a large depth of field so that objects at a distance between 3 and 50 mm can be observed without using the focus adjustment. Enlarging the FOV of a lens in order to create a wider angle while keeping the image size constant requires a shorter focal length. The diameter of the endoscope objective lens should be small enough for noninvasive or minimally invasive observation. A high degree of telecentricity is required for objective lenses used in endoscopes with refractive relay lens systems or imaging fiber bundles. Because relay systems unavoidably degrade the images to some extent, the objective lens must provide adequate images.

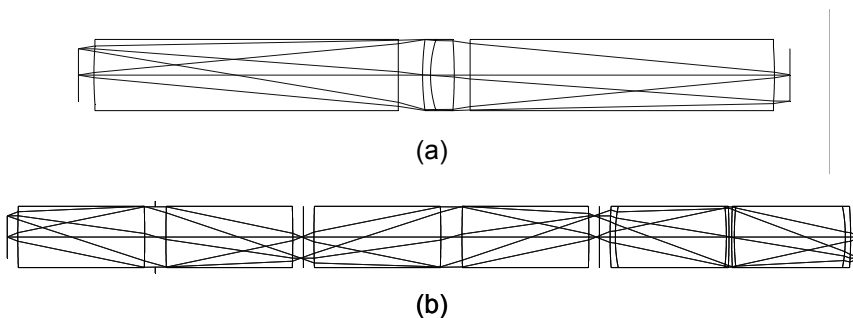
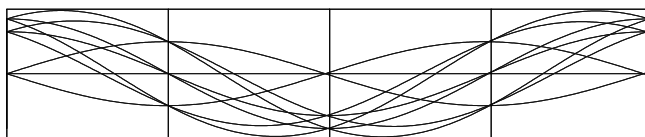
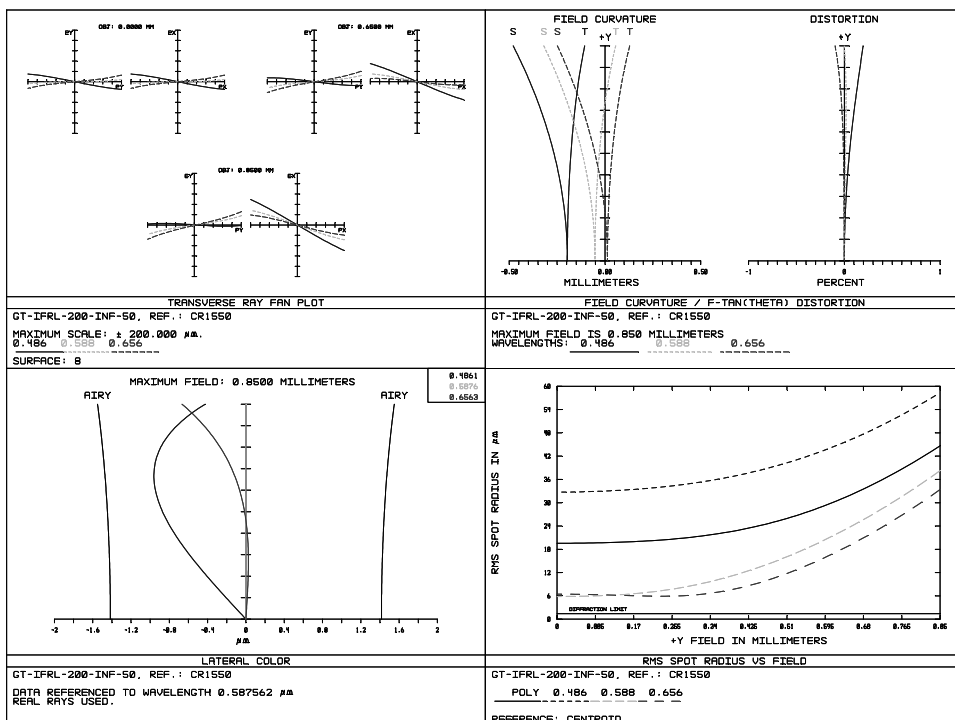


Figure 8.15 (a) Asymmetrical relay lens and (b) relay lens with three relay stages.



(a)



(b)

Figure 8.16 GRIN relay lens: (a) layout and (b) system performance: transverse ray fan plot, field curvature/distortion, lateral chromatic aberration, and RMS spot radius.

To meet the above requirements, a retrofocus objective lens is generally used in endoscopic imaging systems. The retrofocus objective lens primarily consists of two lens groups that are separated by an aperture stop, as shown in Fig. 8.17(a). The back focal length (BFL) is longer than the effective focal length (EFL). In order to achieve a large field angle, the front lens group has a negative refractive power as a whole and often has only one negative lens. The rear group usually consists of several lenses and has a positive refractive power as a whole. The objective lens is designed to obtain a wide-angle view by strongly refracting the chief ray through the first lens group.

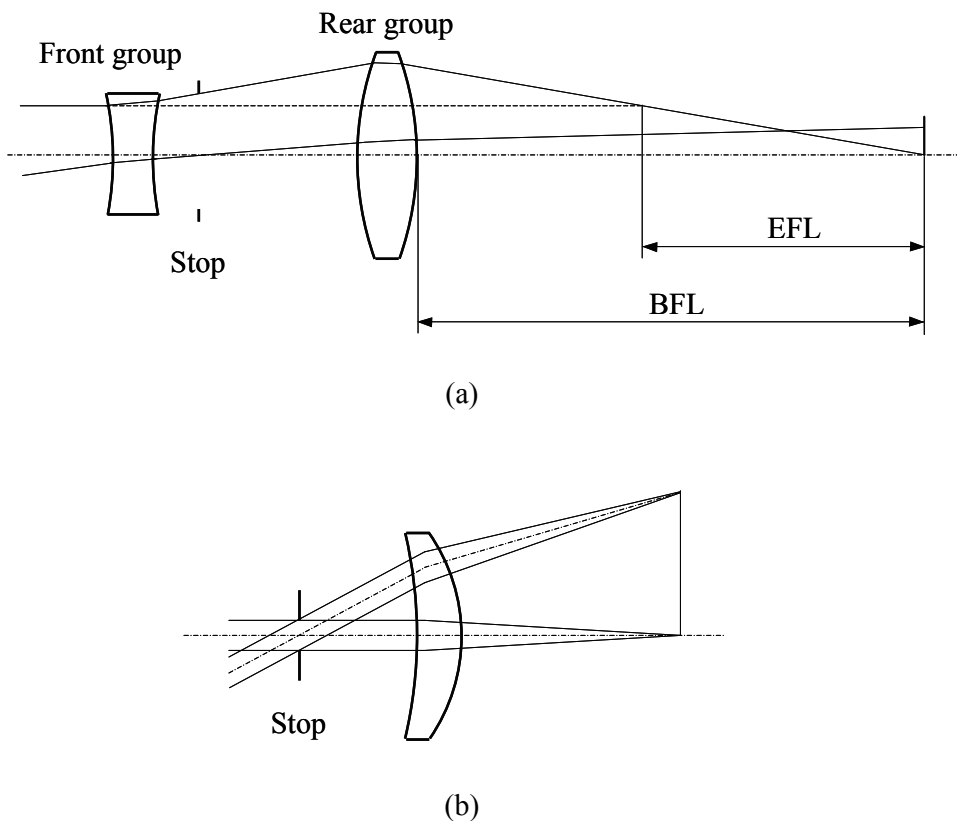


Figure 8.17 (a) Retrofocus lens characterized by a long back focal length and (b) landscape lens with aperture in front.

For the objective lenses used in endoscopes with refractive relay lenses or imaging fiber bundles, the aperture stop is placed at the front focal point so that the chief rays in the image space are parallel to the optical axis. This placement is required to ensure the coupling efficiency of the off-axis rays into the relay lens or fiber bundle. On the other hand, telecentricity is not required for an objective lens used in video or wireless endoscopes. Thus, the objective lenses can be simpler. Their distortion and lateral chromatic aberration can also be compensated digitally because the tissue is imaged onto the sensor directly without any relay component.

The power of a typical objective lens mainly relies on a front lens group that is located close to the object to be observed. If the front lens group is eliminated, such as in the landscape lens shown in Fig. 8.18(b), the overall length of the objective lens can be reduced. However, the power of the rear lens group must be increased accordingly, which leads to deterioration in optical performance. In order to correct chromatic aberration, the objective lens commonly uses one or more cemented doublets.

Given that the objective lens has a negative front group and a positive rear group, the correction for coma, lateral chromatic aberration, and distortion is particularly difficult. When satisfying the telecentric requirement for rigid and flexible endoscopes, a significant distortion will also be produced.

The correction of distortion in a telecentric objective lens system may influence other off-axis aberrations. To correct distortion by using aspherical surfaces, the surface may abnormally depart from the reference sphere. Further, for a compact, wide-angle objective lens, it becomes difficult to correct aberrations, especially off-axis aberrations. Therefore, correction of aberrations other than distortion is the key to designing a wide-angle and compact objective lens for endoscopes.

Lateral chromatic aberration not only generates color blurring of the peripheral image but also reduces the resolution of the peripheral region. A large lateral chromatic aberration prevents precise observation, which is essential for an accurate diagnosis based on color imaging through the endoscope. As the core diameter of the fiber bundle becomes smaller and the size of the pixel in the electronic sensor decreases, sufficiently reducing the lateral chromatic aberration is also required. For an objective lens used with a fiber bundle, the lateral chromatic aberration should be minimized with reference to the distance between the optical fiber cores. If the lateral chromatic aberration is not corrected sufficiently, the transmitted image will produce noticeable color diffusion.

The first element, lens, or window in an objective lens should be very

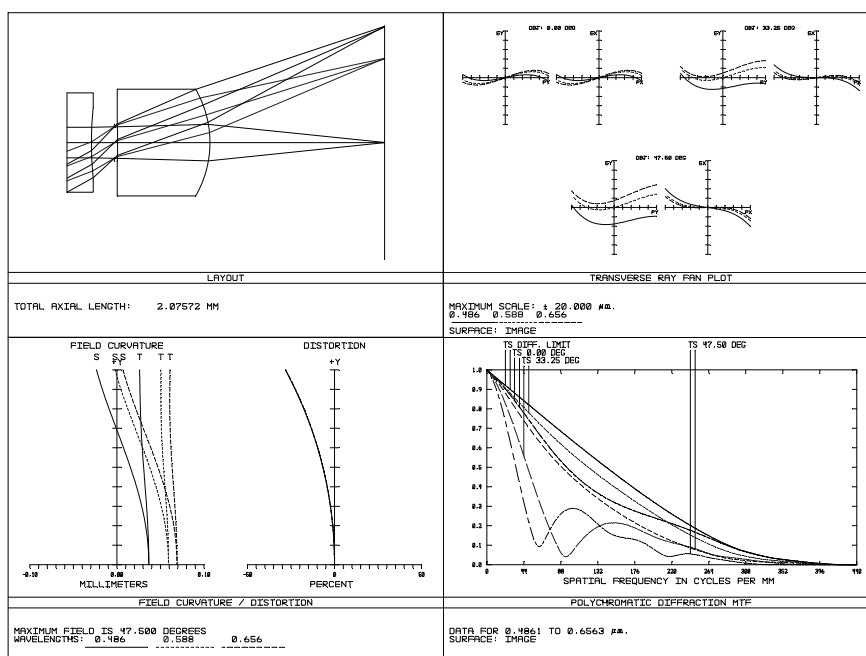


Figure 8.18 Endoscopic objective lens with two singlets.

durable. This element usually has a flat front surface facing toward the tissue in order to avoid any change in optical power when the endoscope is immersed in fluid.

The design of endoscopic objective lenses in this section does not consider electronic or software correction of distortion and lateral chromatic aberration. Some objective lenses are not telecentric in the image space. However, the lens formats, design principles, and techniques can be applied in designing telecentric objective lenses.

8.4.2 Objective lenses with a fixed field of view

Figure 8.18 shows a simple objective lens that consists of a negative singlet, a positive singlet, and an aperture stop between them. This objective lens is designed for the object in infinity and has a focal length of 1 mm, NA of 0.1, and a half FOV of 47.5 deg. The chief ray angle is 19.4 deg, which is smaller than the maximum chief ray angle recommended for most CCD sensors. This type of objective lens has a relatively long back working distance; however, the NA is limited. Another drawback, as shown in the ray fan plot, is that it provides almost no freedom to correct chromatic aberrations. The lateral chromatic aberration is 12 μm . The axial chromatic aberration is about 25 μm .

One way to reduce chromatic aberration is to use optical materials with a large Abbe number since the chromatic aberration is inversely proportional to the Abbe number. Because of the limited freedom for aberration correction, none of the five monochromatic aberrations, especially distortion, field curvature, and astigmatism, are well corrected.

Another type of objective lens that has only two elements and can be used for endoscopes is a landscape lens with a stop in front. Figure 8.19 shows a two-element landscape lens with a window in front of the stop. The optical parameters are the same as for the lens in Fig. 8.18. This landscape lens has an all-positive configuration and therefore has an inherent field curvature that can be

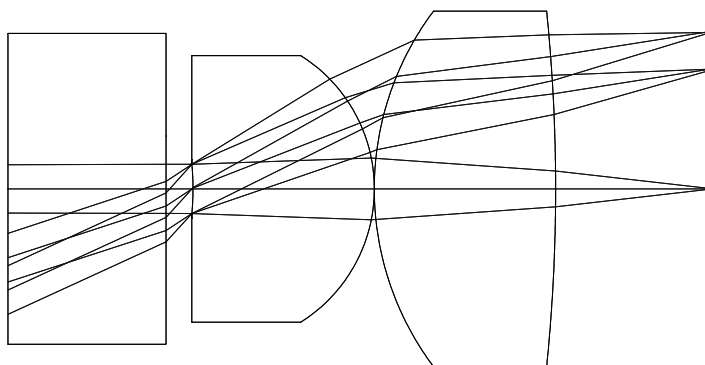


Figure 8.19 A landscape lens for endoscopes. The first optical element is the window without optical power.

reduced by using very high-refractive-index glasses. However, glass with a high refractive index usually has low Abbe numbers, resulting in a relatively large lateral chromatic aberration. Sapphire, which has a refractive index of 1.768 and an Abbe number of 72.2, could be a solution, but it is expensive and can only be made with diamond abrasives. Since there is no negative lens in front of the aperture stop, field-dependent aberrations, such as astigmatism, field curvature, distortion, and lateral chromatic aberration are relatively larger than for the objective lens in Fig. 8.18.

Figure 8.20 shows an objective lens with a doublet in the rear group instead of a singlet to correct the chromatic aberrations. The system parameters, such as NA and FOV, are the same as for the objective lens in Fig. 8.18. With the cemented doublet, both axial and lateral chromatic aberrations can be improved significantly.

The chief rays of the objective lenses discussed above are not parallel to the optical axis, which means that the exit pupil is not at infinity. These objective lenses are only suitable for video endoscopes with electronic sensors, such as CCD and CMOS sensors, at the distal end. To achieve a telecentric condition in the image space, the aperture stop should be placed on the front focal plane of the objective lens. Figure 8.21 shows a three-element telecentric objective lens with 0.17 NA and 60-deg half FOV. The image surface is at the last surface of the objective lens so that it can be glued or joined to the incident end of the fiber bundle. If this objective lens is used together with a relay lens system, the image plane should be away from the lens surface to avoid a tight quality requirement on the last surface. Distortion is considerably large and is proportional to the field angle. Spherical aberration and coma are corrected sufficiently, but the astigmatism, field curvature, and chromatic aberration are still large. Typically, both the axial and lateral chromatic aberrations at shorter wavelengths are on the negative side of those at a reference wavelength because the focal length in short wavelengths is smaller.

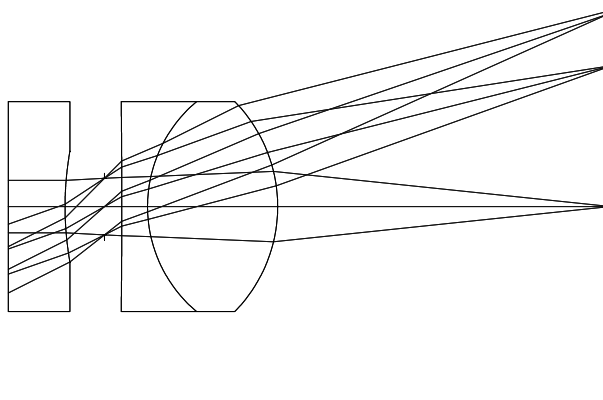


Figure 8.20 Objective lens with a doublet to correct chromatic aberrations.

There are two approaches to reduce lateral chromatic aberration. The first approach reduces the lateral chromatic aberration on the basis of the principle of symmetry. It is effective in reducing the lateral, but not axial, chromatic aberration. Figure 8.22 shows one design using this approach. The objective lens is designed with a working distance of 10 mm, an NA of 0.137, and a half FOV of 60 deg. The lateral chromatic aberration is well balanced. Coma and astigmatism are reasonably controlled. Other aberrations, such as spherical aberration, field curvature, and distortion, need further improvement.

The second approach uses a cemented doublet, which is an effective way to reduce both axial and lateral chromatic aberrations. To correct for the lateral color, it is preferable to place the cemented doublet at a position that is a distance from the aperture stop where the ray height is large. Figure 8.23 shows an objective lens with a doublet as the last element close to the image plane. The working distance, NA, and FOV are the same as those of the objective lens in Fig. 8.22. The center of the cemented surface is on the opposite side of the aperture stop, so the off-axis rays enter the cemented surface at an angle far from the perpendicular direction. The major aberrations remaining in the objective lens include coma, field curvature, and distortion. The curvature of the cemented surface can be increased to ensure a more effective compensation for the lateral

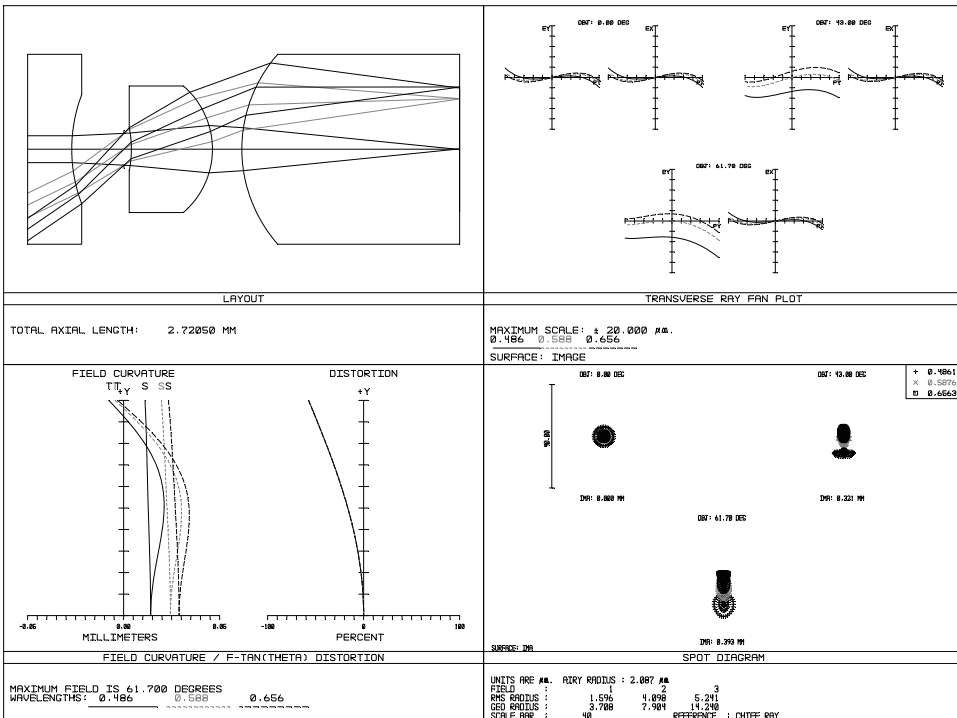


Figure 8.21 Three-element telecentric objective lens.

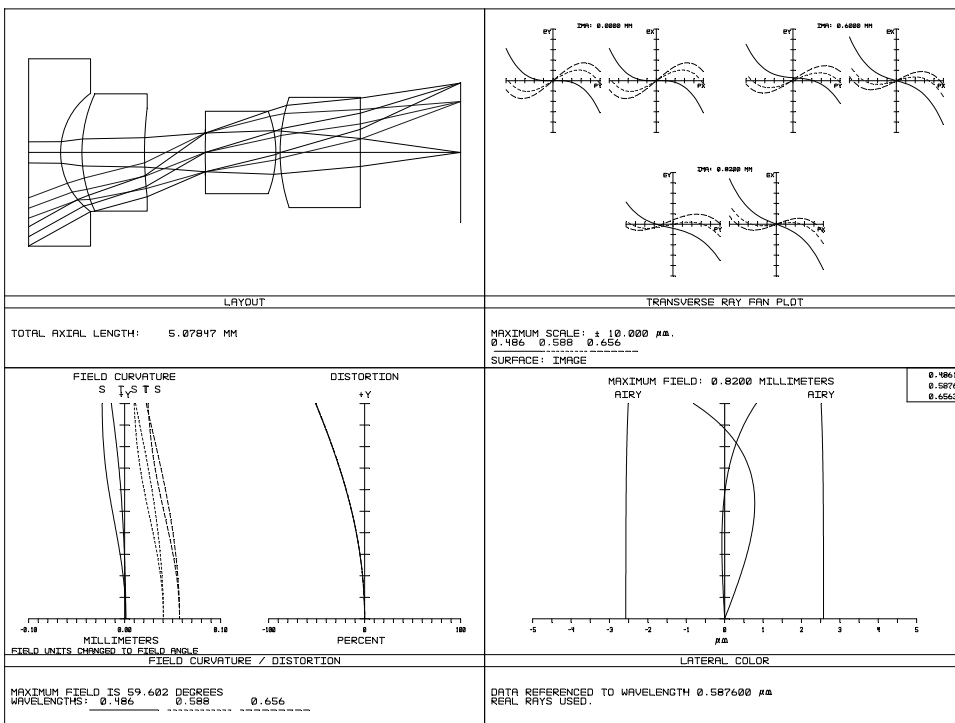


Figure 8.22 Objective lens with a symmetrical lens configuration to reduce lateral chromatic aberration.

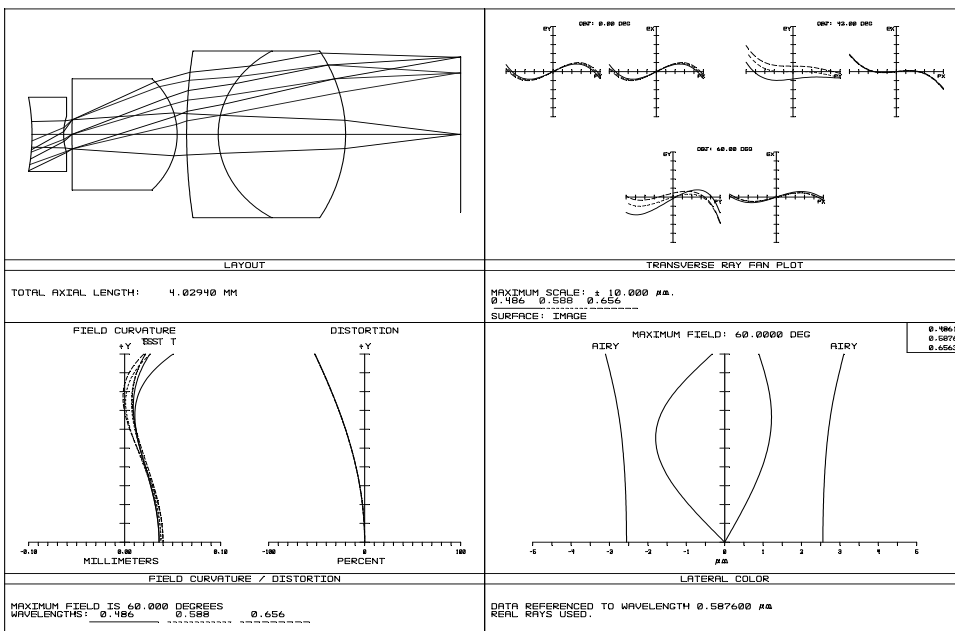


Figure 8.23 Objective lens with a cemented doublet to correct lateral chromatic aberration.

chromatic aberration. However, the degree of meniscus of the negative lens elements in the doublet is increased, making the manufacture of these lenses more difficult.

The objective lens in Fig. 8.24 is a combination of adding a positive lens in the front group to moderate the asymmetry of the optical system and using a cemented doublet in the rear group to reduce both axial and lateral chromatic aberration. The objective lens has an NA of 0.137 and a half FOV of 60 deg. The coma and chromatic aberration are well balanced. The negative spherical aberration induced by the concave surface of the first negative lens is compensated by the positive lenses in two doublets near the image plane. However, the distortion is still relatively large and is mainly from the first surface of the negative lens in front of the aperture stop.

One potential issue with the cemented doublet is that the cement material may deteriorate and slough off as a result of repeated high-temperature sterilization. To address this issue, some objective lenses only comprise a plurality of noncemented lenses with at least one negative lens made from a high-dispersion material at each side of the aperture stop for chromatic aberration correction.

To achieve better aberration correction with a simple configuration, one or more aspherical surfaces can be used. When the aspherical profile is applied on

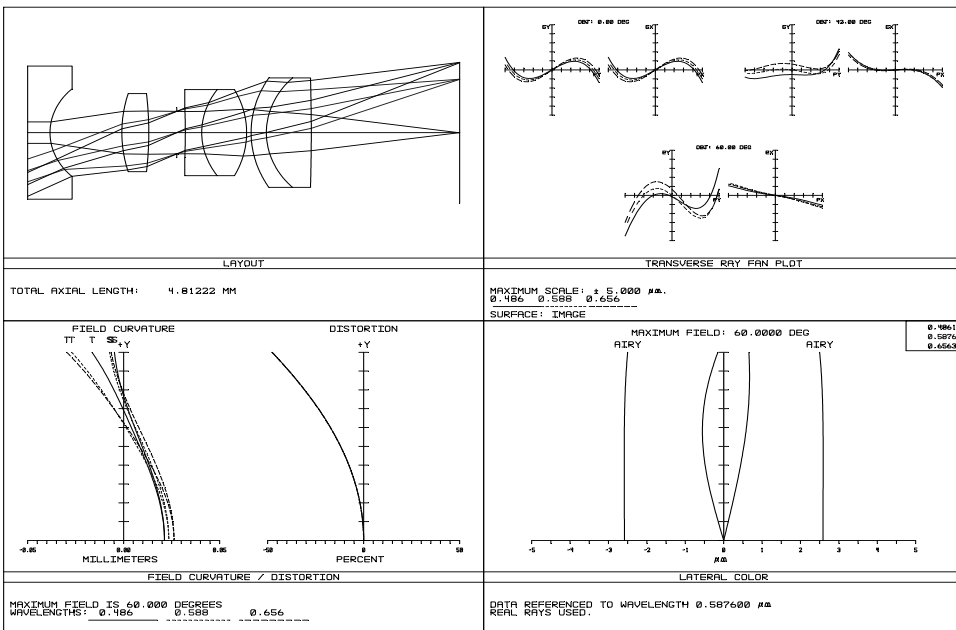


Figure 8.24 Objective lens with a positive lens in the front group and two doublets in the rear group for chromatic aberration correction.

the surface away from the aperture stop, the field-dependent aberrations, such as astigmatism and distortion, can be controlled more effectively. On the other hand, if the aspherical surface is located on the surface close to the aperture stop, the correction of spherical aberration can be improved.

Figure 8.25 shows an example of using an aspherical surface to improve the aberration correction. The configuration and system parameters are the same as those for the lens in Fig. 8.18. The aspherical surface is on the concave surface of the first element with some distance from the aperture stop. The performance plots clearly show that the overall performance is improved. The spherical and lateral chromatic aberrations are reasonably smaller than those for the lens without an aspherical surface. Residual aberrations consist primarily of zonal astigmatism and distortion; the distortion of this design approaches 29% at the edge of the field. This lens has the potential for wide-angle application.

The objective lenses discussed above have a negative lens component as the first element to obtain a wide FOV. A considerable negative distortion occurs at the first surface, which is usually a flat surface. The radius of the first surface can be decreased to reduce the negative distortion. However, the small radius of the first surface tends to cause water drops to remain on the surface, degrading image quality.

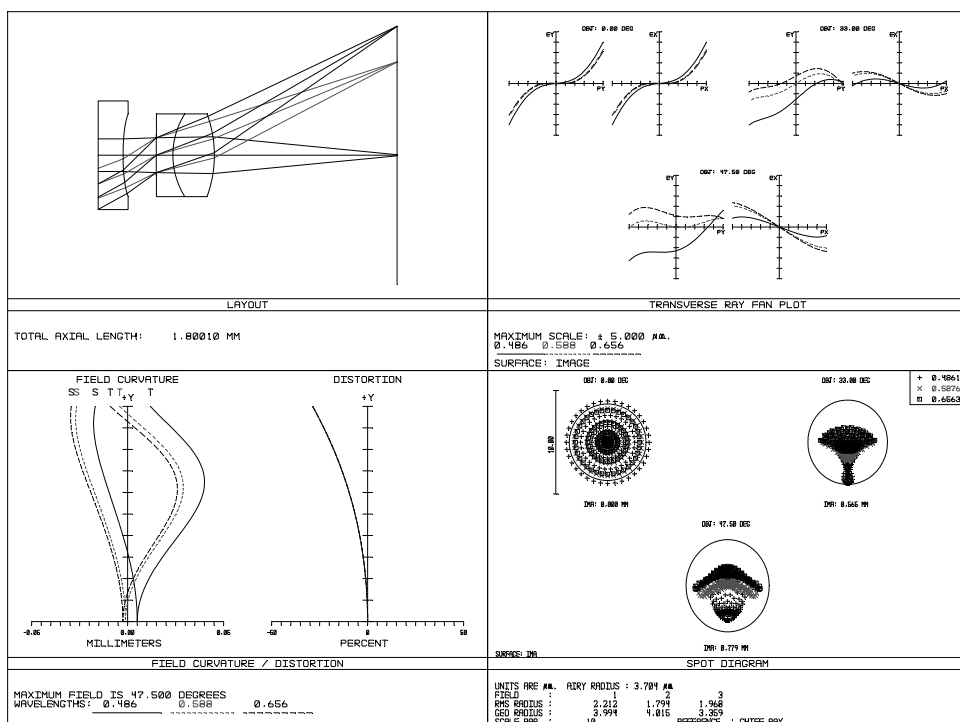
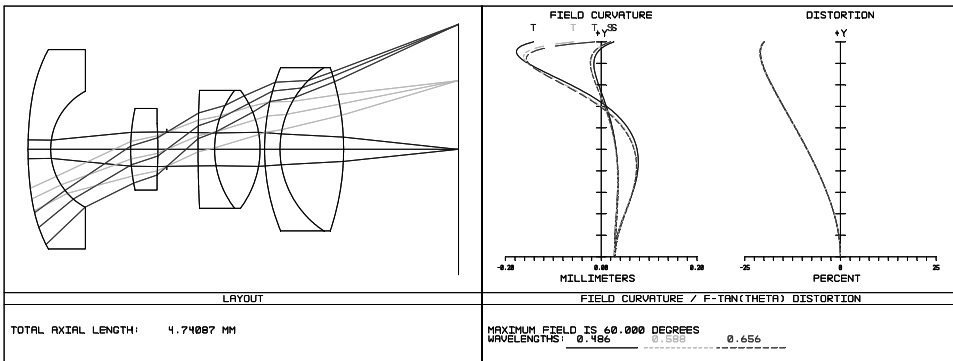


Figure 8.25 Objective lens with a doublet to correct chromatic aberrations and an aspherical surface to improve monochromatic aberrations. The concave surface of the first element is aspherical.

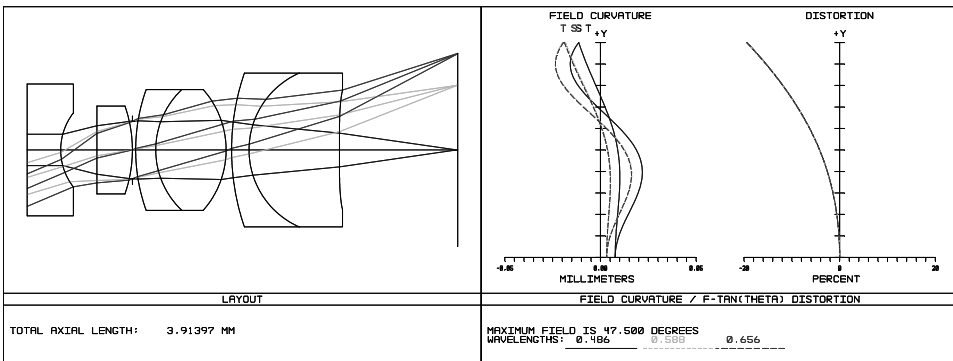
When correcting the distortion in an endoscopic objective lens, the other odd aberrations—coma and lateral chromatic aberration—are influenced directly. An endoscopic objective lens usually has a large FOV, and as a result, off-axis aberration is even more difficult to correct. Therefore, correction of odd aberrations other than distortion is the key in designing a compact and wide-angle objective lens.

Distortion can be effectively controlled by using an aspherical surface since curvature at a point on an aspherical surface can be continuously varied. Therefore, it is possible to vary the curvature at a point on the aspherical surface through which the chief or principal ray can attain an optimal point on the image plane. Generally, it is sufficient to arrange the aspherical surfaces at locations where the chief ray is high.

Figure 8.26(a) shows an objective lens modified from that in Fig. 8.24. The front surface of the first lens is formed as an aspherical surface whose curvature gradually gets stronger as it moves away from the optical axis. Compared with



(a)



(b)

Figure 8.26 Objective lens with an aspherical surface to correct distortion. (a) An aspherical surface in the front group and (b) an aspherical surface in the rear group.

the objective lens in Fig. 8.24, the distortion is reduced significantly, from 45 to 21%, and the spherical aberration is also reduced, but the coma, astigmatism, and field curvature are not as controlled as in the previous design.

Figure 8.26(b) shows another variation of the objective lens in Fig. 8.24. The rear surface of the last lens element is aspherical. The aspherical surface for correcting distortion has a gradually weakening refractive function as the lens components are positioned farther away from the optical axis toward the edge of the field. Similar to the objective lens in Fig. 8.26(a), the distortion is reduced significantly, from 45 to 19.4%, and the spherical aberration is smaller, but coma, astigmatism, and field curvature are slightly degraded.

Similar to lateral chromatic aberration, distortion can also be improved by moderating the asymmetry. Figure 8.27 shows an objective lens with an NA of 0.1 and FOV of 80 deg.¹⁵ All of the surfaces are spherical. The front lens group contains one positive lens and one negative lens, with the positive lens in front. The plates between the front and back lens groups are prisms for side viewing. The distortion produced by the front positive lens appreciably compensates for the negative distortion that the objective lens would otherwise have. The maximum distortion is well controlled and less than 6.5%. As the plot shows, lateral chromatic aberration is also small.

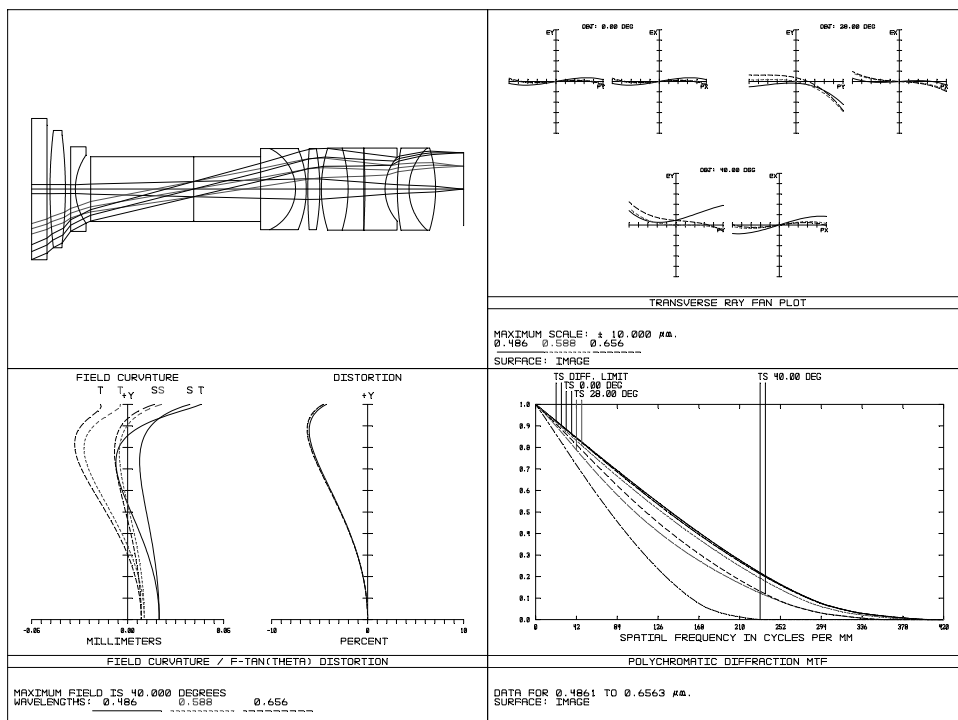


Figure 8.27 Objective lens with a semisymmetrical configuration to correct distortion and lateral chromatic aberration.

Because of its unique properties, a GRIN lens is sometimes used as an endoscopic objective lens. The single-element objective lens eases the tight tolerance requirement for both optical and mechanical components, and it dramatically reduces the time for assembly. The flat surfaces on a GRIN lens also simplify the assembly of complex optical systems, such as prisms, beamsplitters, and fiber bundles. For example, the image plane of the objective lens can be configured to be on the flat surface of the GRIN lens so that the optical fiber bundle can be glued directly onto the lens surface.

A GRIN lens has a significant advantage when the diameter of the optics has to be smaller than 2 mm, especially below 1 mm, because it is very difficult to manufacture and assemble conventional lenses smaller than 2 mm in diameter. Another advantage of a GRIN lens is that it only has two air/glass interfaces, so the Fresnel reflection loss is minimized.

Figure 8.28 shows one example of a GRIN lens (GRINTECH's GT-IFRL-200-020-50-CC). The lens diameter is 2 mm, the FOV is ± 30 deg, and the NA is 0.25. As can be seen from the performance plots, the major remaining aberrations are lateral and axial chromatic aberrations. The other aberrations, namely, spherical aberration, coma, astigmatism, field curvature, and distortion, are comparable to those of conventional objective lenses.

The primary limitation of a GRIN lens as an endoscopic objective lens is that it suffers from excessive dispersion, making it impossible to produce a color-

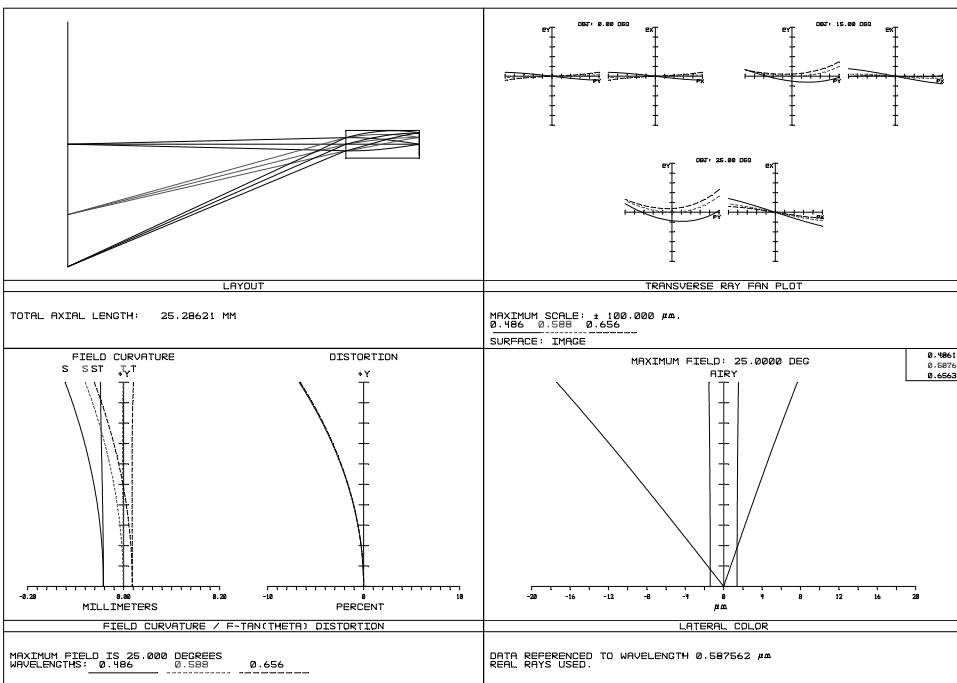


Figure 8.28 A GRIN objective lens.

corrected objective lens for the visible spectrum. Another major limitation of a GRIN lens is that the FOV is relatively small compared to the objective lenses in conventional endoscopes. In addition, there is a limited amount of freedom in designing objective lenses with parameters other than those provided by the manufacturers.

8.4.3 Objective lenses with variable fields of view

While an endoscope provides a wide-angle overview of the subject, there is a need to observe certain regions in detail given that medical treatment demands highly precise observation and a more accurate diagnosis. In recent years, endoscopic objective lenses that are transferable between a wide-angle view and a close-up partial view have been developed. The wide-angle mode, which is the normal viewing mode, refers to a mode having a longer working distance and a wider viewing angle, thus making it suitable for viewing over a wide area. On the other hand, the close-up viewing mode refers to a mode having a shorter working distance and a small field of view, thus making it suitable for detailed viewing. In the wide-angle viewing mode, the system uses its lowest magnification; in the close-up viewing mode, the system uses its highest magnification. A close-up view allows for detailed observation of the morbid region in order to inspect the microstructure, measure the degree of infiltration, and determine the extent to which the region should be excised. Depending on the design, there are two types of objective lenses with variable fields of view. One is only suitable for wide-angle viewing and close-up viewing; the image quality between these two ends is not optimized. The other is a zoom lens that is suitable for continuous viewing between the two extremes.

The objective lens shown in Fig. 8.29 is made up of two lens groups.¹⁶ Group I is the objective lens group that forms an intermediate image. Group II is a relay lens with two field lenses; one is located in the vicinity of the intermediate image in order to converge the light from the imaging lens, and the other is near the image plane in order to reduce the chief ray angle. The imaging

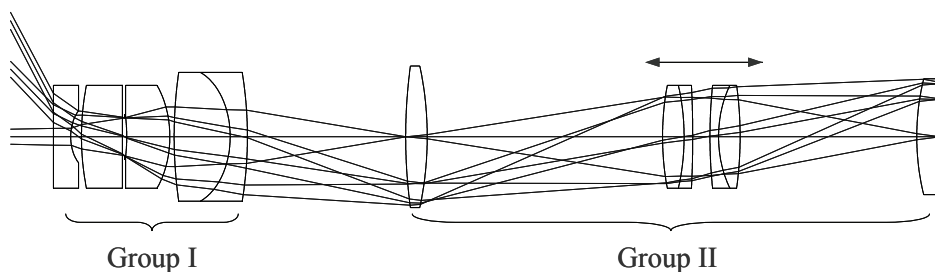


Figure 8.29 Objective lens with two groups of lenses: imaging lens and relay lens. The two doublets in the relay lens are movable to change the FOV captured by the sensor. The working distance is fixed while the lenses move.

lens in Group II is movable between two positions, while the distance from the first element of the objective lens to the image sensor is fixed.

Since the distance between the object and the image plane is fixed when the lens is moved to change the viewing mode, the FOV for Group I is the same for the wide-angle and close-up modes; therefore, the intermediate image is the same for both modes. The FOV captured by the sensor is determined by the size of the sensor. When Group II moves from a wide-angle position to a close-up position, only a portion of the intermediate image is relayed to the sensor. Therefore, the partial view of the object does not increase the resolution of the local image; the image is only magnified. The depth of field is the same for both modes.

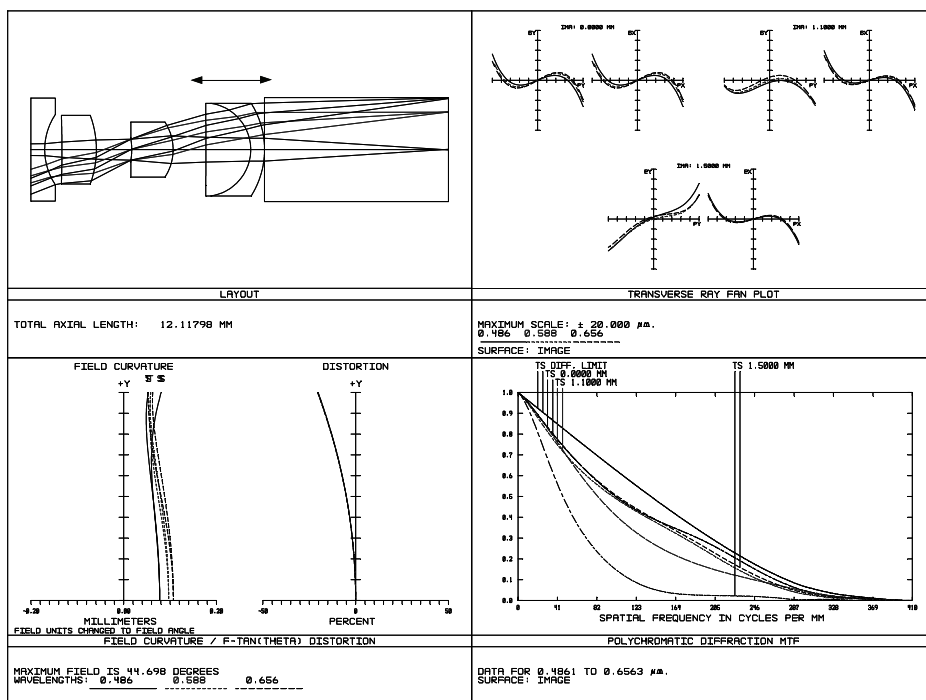
Figure 8.30 shows an objective lens with a variable viewing depth.¹⁷ The objective lens is optimized only for wide-angle mode and close-up mode. The last doublet is movable in order to change the field of view, thereby changing the working distance. This objective lens is not a zoom lens, so the performance between the far point (wide-angle mode) and near point (close-up mode) may not be acceptable.

The full FOV for the wide-angle mode and close-up mode are 25 and 5 mm, respectively, and the working distances are 12 and 1.75 mm, respectively. The focal length of the objective lens is shorter in the close-up mode than in the wide-angle viewing mode. The resolution in a close-up observation is much higher than in a wide-angle observation; however, the depth of field is much smaller. The optical performances of the system for wide-angle and close-up observations are plotted in Figs. 8.30(a) and 8.30(b). The spherical aberration and coma are better controlled in the close-up configuration, but the astigmatism is degraded.

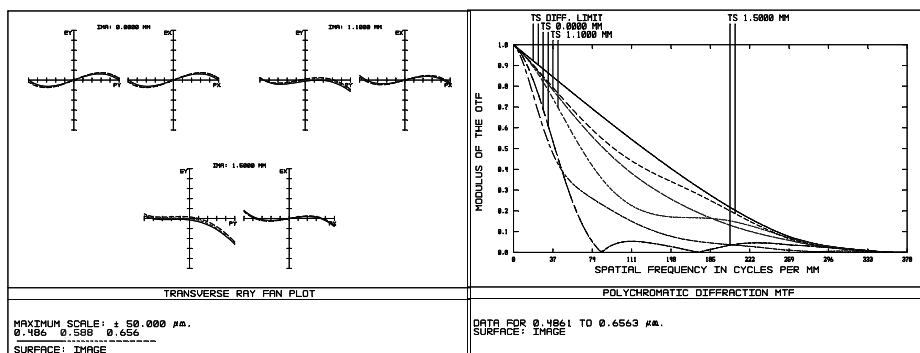
Figure 8.31 shows an example of an objective lens that can continuously change its magnification between the wide-angle and close-up viewing modes.¹⁸ It comprises (successively from the object side) a first lens group that has a negative refracting power, a second lens group that is a positive lens and is movable along the optical axis, a third lens group that is a negative lens and is also movable, and a fourth stationary lens group consisting of the last three lenses. The second lens group is the variator, and the third group is the compensator. The FOV of this objective lens is from 30 to 2.0 mm, and the working distance is from 10.4 to 1.4 mm.

In this example, the aperture is fixed and moves with the lens. This objective lens has a typical zoom-lens design with one variator group and one compensator group. The total length, which is the distance from the very first lens surface to the sensor, does not change while the two lens groups move. The performances of the objective lenses in the wide-angle and close-up modes are plotted in Figs. 8.31(a) and 8.31(b). As the plots of RMS spot radius show, the system has good performance in both modes. In the wide-angle mode, distortion and lateral chromatic aberration need further improvement, while coma and astigmatism are the limiting aberrations in the close-up mode.

Figure 8.32 shows a straightforward objective lens that focuses on objects at different working distances by moving the sensor along the optical axis.¹⁹ The



(a)



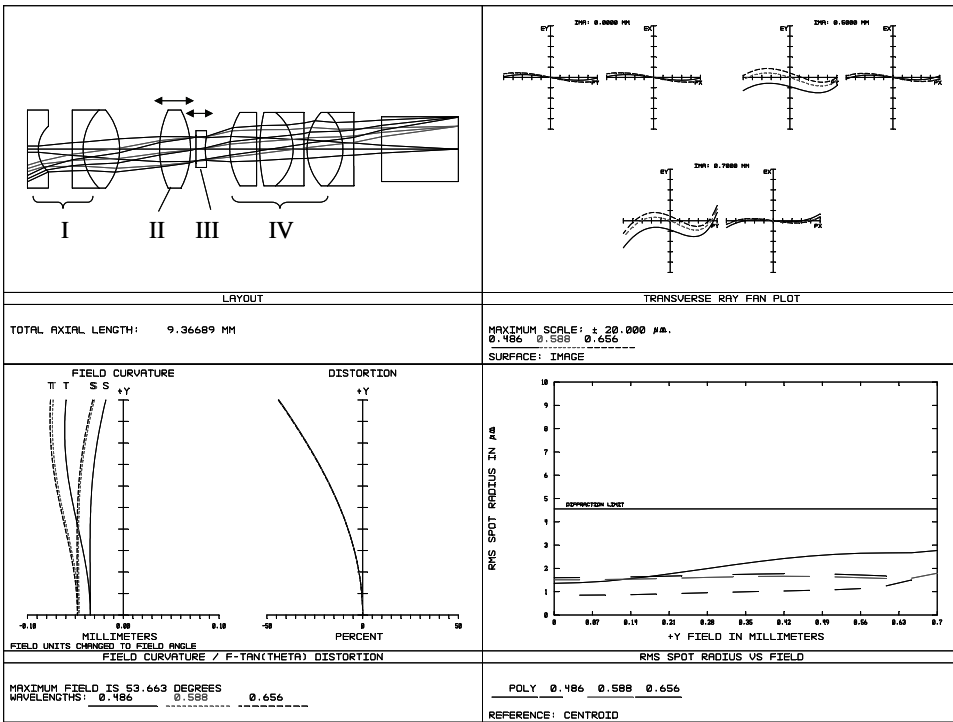
(b)

Figure 8.30 Objective lens with one moving lens element in the rear group: (a) wide-angle mode and (b) close-up mode.

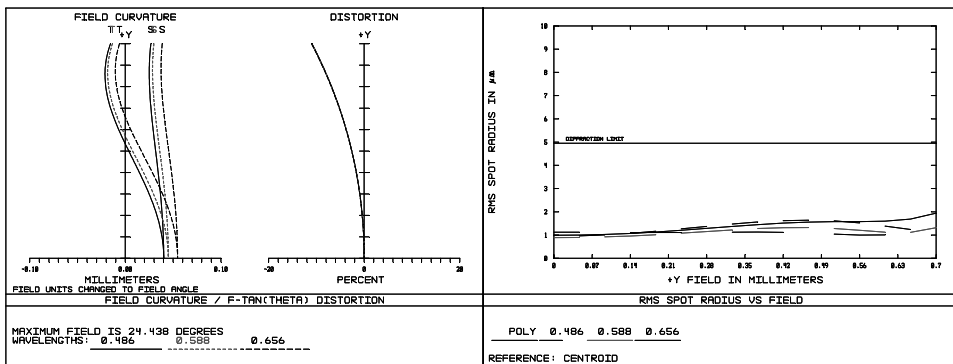
optical system is telecentric with the help of a field lens. The MTF plots in Figs. 8.32(a) and 8.32(b) show that this simple system has similar performance in both modes.

8.4.4 Design options

One approach in designing the endoscopic objective lens and the relay



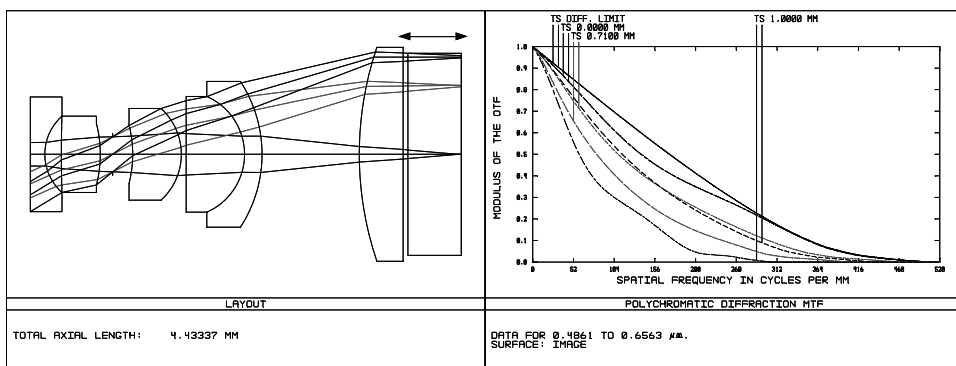
(a)



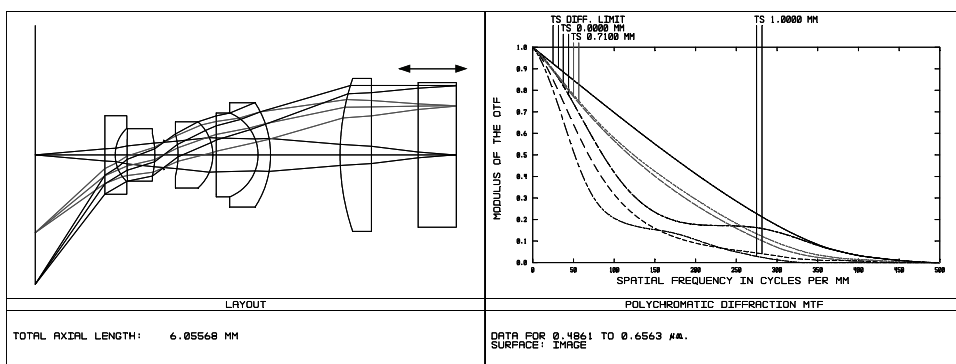
(b)

Figure 8.31 Objective lens with two moving lens elements to change the FOV: (a) wide-angle mode and (b) close-up mode.

system is to design them separately and then integrate them without further optimization. This approach is very straightforward; however, it requires that each subsystem be well corrected for aberrations, especially field curvature and astigmatism. The correction of these aberrations usually requires lenses with small radii of curvature; such lenses are difficult to fabricate and require tight tolerances.



(a)



(b)

Figure 8.32 Objective lens with movable sensor to change the FOV: (a) wide-angle mode and (b) close-up mode.

Since an endoscope imaging system usually consists of several relay stages, and the field curvature of the complete relay system is equal to that of a single relay unit multiplied by the number of relays, each relay unit should have a good correction of field curvature. Relay lenses for endoscopes also should be double telecentric to avoid severe vignetting. In addition, the lens surfaces in a relay system should not be too close to the internal images; otherwise, dust and scratches will become a major issue. All of these requirements make it difficult to design a high-performance imaging system by designing the objective lens and relay system separately.

Another approach is to design the objective and relay system as a system, thus relaxing the flat-field requirement for the relay subsystem. Figure 8.33 shows an endoscopic imaging system designed with the above approach.²⁰ This design consists of an objective lens and five relay units. The half FOV is 30 deg, and the NA on the image side is 0.075.

The relay unit has a pair of thick cemented doublets in which the Petzval sum is reduced but not fully corrected by the use of a moderately high-refractive-

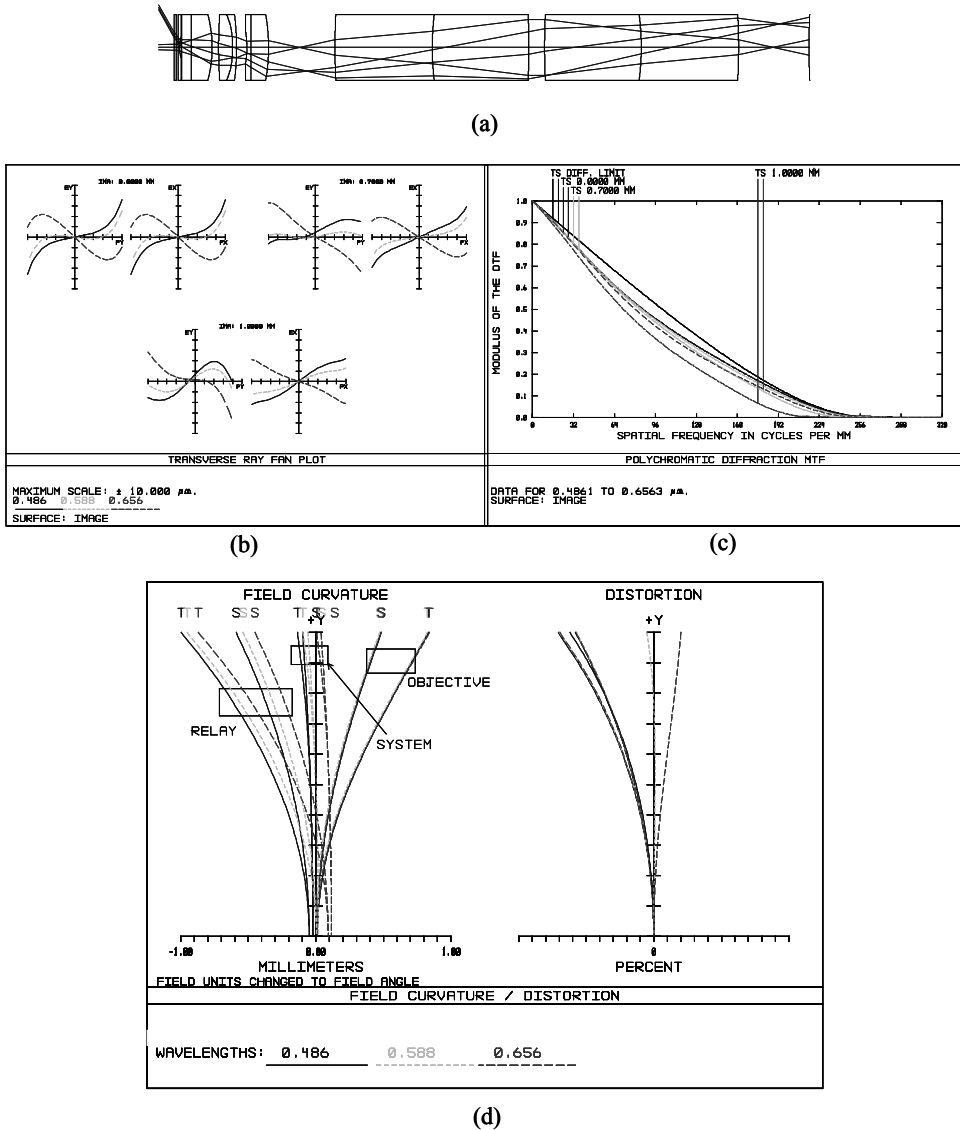


Figure 8.33 Endoscopic imaging system with aberrations balanced through the objective and relay lenses: (a) optical layout, (b) ray fan plot, (c) MTF plot, and (d) field curvature and distortion.

index crown glass. Figure 8.33(a) shows the layout of the objective lens and one relay unit. Figures 8.33(b) and 8.33(c), which are the ray fan and MTF plots, show that the overall performance is nearly diffraction limited with some residuals of axial chromatic and spherical aberrations. As shown in Fig. 8.34(d), the undercorrected astigmatism and field curvature of the relay lenses are compensated by that of the objective lens.

8.4.5 Prisms in endoscopic objective lenses

The objective lenses discussed above are designed for forward viewing. However, in some situations, the region of interest is not directly in front of the objective lens, and it can be inclined from the axis of the endoscope. Therefore, there is a need to change the viewing direction as well as the illumination direction at the distal end of the endoscope. A refracting prism, reflecting prism, and mirror can be used to change the viewing direction.

When the deflection angle is small, for example, less than 15 deg, a wedge plate (refracting prism) or a noncentered front lens can be used as a deflecting element. However, a wedge plate or decentered element unavoidably generates aberrations that degrade image quality. Figure 8.34(a) is a diagram of a refracting prism, which is used in some endoscopes to change the viewing direction. The refracting prism is usually used in front of the objective lens. When the wedge angle is increased to enlarge the incline angle, the off-axis rays can intersect the edge of the prism (ray *a*) or encounter TIR (ray *b*). Therefore, the FOV is limited when using a refracting prism. The lateral chromatic aberration and distortion increases quickly as the wedge angle increases. Due to the limited incline and aberrations, refracting prisms are only used in rigid endoscopes with outer diameters of less than 2 mm for which reflecting prisms are extremely difficult and expensive to fabricate.

Even a mirror is a simple component that changes the viewing direction; it is

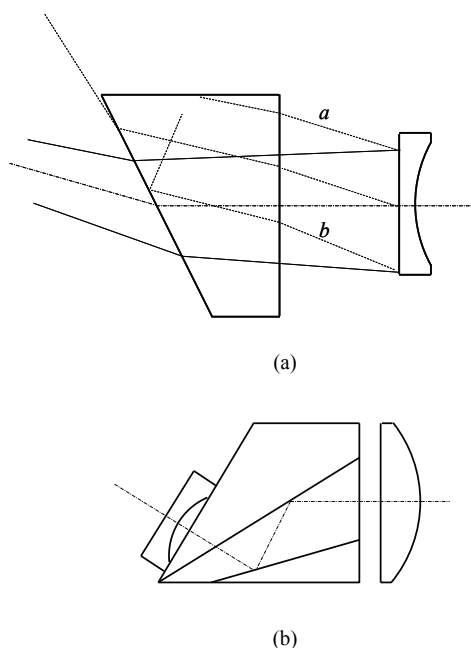


Figure 8.34 Prisms for changing the viewing direction: (a) refracting prism and (b) prism assembly.

seldom used at the distal end of endoscopes because of the difficulty in mounting a mirror at the distal end. A reflecting prism is the element in endoscopes that is most frequently used to change the viewing direction. In order to minimize the prism size, the prisms are usually placed between the front negative lens and the remaining lenses or between the front lens group and rear lens group, as shown in Fig. 8.34(b). The objective lens in Fig. 8.27 is one example of an objective lens with prisms placed between the front and rear lens groups.

Typical reflecting prisms and prism assemblies used in endoscopes are shown in Fig. 8.35. Each prism has an input and output surface perpendicular to the optical axis. Figure 8.35(a) shows a 90-deg prism with only one reflection. Figure 8.35(b) shows a roof prism with two reflections to correct the mirror inversion. The angles of the prisms in Figs. 8.36(a) and 8.36(b) can be other than 90 deg to obtain a forward- or backward-viewing field. Figures 8.35(c)–8.35(f) are configurations of reflecting prisms for viewing directions less than 90 deg.

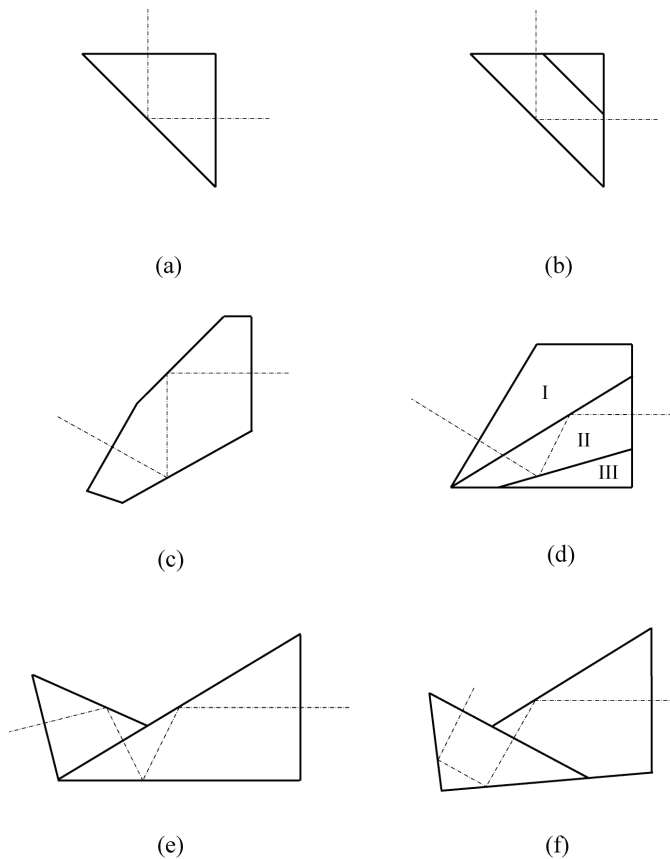


Figure 8.35 Prisms and prism assemblies for inclining the FOV: (a) right-angle prism, (b) roof prism, (c) prism with two reflection surfaces, (d) prism assembly with three prisms, (e) prism assembly with three reflections, and (f) prism assembly for backward viewing.

The configurations in Figs. 8.35(c) and 8.35(d) have two reflecting surfaces that retain the proper reversion of the image. The prism in Fig. 8.35(c) is advantageous because it has only one element and is more robust. The prism assembly shown in Fig. 8.35(d) has three components: I, II, and III. The interface between components I and II is coated with a reflecting layer over only the upper half of the clear aperture since the light ray also transmits through the lower half of this interface. This prism assembly can be simplified to elements I and II with an air gap between them, but the FOV is limited as a result of the TIR. The prism assemblies in Figs. 8.35(e) and 8.35(f) have two prisms with three reflective surfaces, with the former for forward viewing and the latter for backward viewing. With current plastic-molding technologies, the prisms in Figs. 8.35(c), 8.35(e), and 8.35(f) can be easily molded with only one component, eliminating a cemented surface.

While there are many different prism designs that allow for changing the viewing direction, the key requirement for designing a prism is the FOV. Because of space constraints, the FOV of an objective lens with a prism to change the view direction is smaller than that without a prism. For an endoscope equipped with an eyepiece for direct viewing, the produced image must be “right side up.” To achieve this, the endoscope must have an even number of reflections. Those reflections can occur at the distal or proximal end of the endoscope. For an endoscope whose final receiver is an electronic sensor, mirror inversion is not necessarily corrected by an optical method and can always be done electronically or with software.

When using an electronic image sensor at the distal end, sometimes there is a need for an endoscopic objective lens with a longer back working distance to accommodate the thickness of the glass window, filters, and the distance between the window and the sensing pixels. The packaging of the sensor may be too large to be placed perpendicular to the optical axis of the endoscope. For this reason, the sensor may be arranged parallel or obliquely, relative to the longitudinal axis of the endoscope, so that the sensor occupies a minimum space in the diametric direction. To implement this arrangement, the BFL should be long enough that the folding component, i.e., prism and mirror, can be inserted between the optical lens and the image sensor. One example of a folding prism is shown in Fig. 8.36.

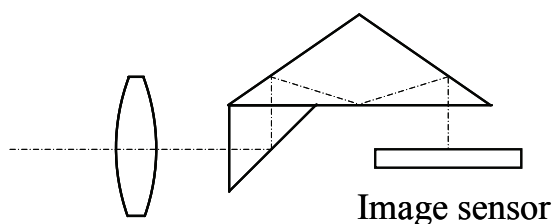


Figure 8.36 Folding prisms in a video endoscope that accommodates for sensor size.

8.5 Illumination Systems

The general requirements of illumination systems for endoscopic imaging are sufficient light level, uniform illumination, high light utilization efficiency, large FOV, and low cost. A number of light sources and illumination methods have been developed for endoscopic imaging and have been used successfully in clinical applications.

8.5.1 Light sources

Xenon or halogen lamps are the traditional light sources used in endoscopes. A fiber or fiber bundle is usually used for delivering the light from the lamp to the observation region. With a long transmission path, usually 2 m or longer, from the light source to the distal end of the endoscope, light loss due to the coupling and absorption is significant; therefore, a lamp with higher power to balance the loss is needed.

Most broadband light sources produce large amounts of UV and IR light. IR filters are necessary to prevent the transfer of IR radiation to the tissue. With the development of high-intensity white-light sources, dichroic mirrors that block infrared radiation from light sources, and fiber bundles with low absorption, it is possible to have high levels of illumination with high color temperature without danger from the heat generated by the light source.

LEDs are a new type of light source used in endoscopes. They have several advantages over conventional light sources. LEDs can provide light at any region of the spectrum from UV to IR. White light can be generated by mixing red, green, and blue LEDs or by phosphor-converting white LEDs. The LED lifetime is more than an order of magnitude longer than that of bulb-type light sources. The power consumption is also much lower than for other high-power light sources. LEDs do not require a warm-up time and are capable of providing instant illumination.

With three primary RGB LEDs, the color gamut of the illumination can be tuned for different applications. For example, shifting the illumination color can increase the visibility and differentiation needed in the diagnostic evaluation of a surgical scene. LED illumination systems are more robust against shock, vibration, and other extreme environmental conditions than fiber illumination systems.

8.5.2 Fiber illumination systems

Before the introduction of optical fibers into endoscopes in the early 1960s, the illumination of the endoscope relied on a small tungsten filament lamp installed at the distal end of the endoscope. There are several issues with this illumination method: insufficient light for illumination, poor illumination uniformity, low color temperature, and the heat generated by a tungsten lamp. The color temperature is very important in endoscopy because the detail in tissue often depends on relatively slight differences in color.

The fiber bundles for illumination are incoherent fiber bundles. The individual fibers do not have a particular order between their two ends, providing the benefit of reducing the nonuniformity of light distribution at the output end when the light is not uniformly coupled into the fiber at the input end.

The efficiency of a fiber bundle for illumination depends on the core diameter and NA. With a fixed diameter of the fiber bundle, the core diameter of each individual fiber determines the packing fraction and therefore the light-coupling efficiency. The larger the core diameter is, the larger the packing fraction, which means the smaller the loss of light when coupling it into a fiber bundle. However, the larger the core diameter is, the less flexible the fiber bundle. Usually a fiber with a large NA is used in illumination systems in order to convey more light from the light source to the distal end of the endoscope. Theoretically, the ability of a fiber to collect light from a light source is proportional to the NA^2 .

Another important requirement of the fiber used for endoscopic illumination is the spectral response of light transmission. The true rendition of color, largely depending on the spectral content of the illumination, is of particular importance in medical applications where the diagnosis of disease depends on the appearance and color of the tissues under investigation. Because of the strong absorption in the UV and blue regions, optical fibers made of plastic materials are not extensively used as illumination fibers in endoscopes with broad-spectrum imaging despite their flexibility and low cost. This absorption creates a need for a special fiber, such as a fiber made of fused silica, for high transmission in the UV and blue regions.

One advantage of a fiber bundle in illumination is its flexibility in the shape of the output end. The output end can be arranged in different shapes, such as a ring or a line, for special applications. The output or input end of an optical fiber bundle can also be split into multibranch fiber bundles, such as bifurcated or trifurcated bundles. In multibranch fiber bundles, the fiber bundles should be comb randomly to produce evenly divided outputs.

The fiber bundle can be used directly in endoscope illumination without additional optical elements. The size of the illumination field is determined by the NA, which is much larger than that of the imaging fiber bundle but may not be large enough for a wide-angle endoscope. For example, when the refractive indices of the core and cladding are 1.62 and 1.48, the NA of the fiber is 0.659, corresponding to a divergent angle of 41.2 deg. Typically, the intensity of a ray having a large angle from the optical axis is smaller than that of a ray close to the optical axis. This nonuniform intensity further limits the usable illumination field.

When a fiber bundle with a diameter of 1.5 mm and an NA of 0.659 illuminates an object 20 mm away, the illuminated region has a diameter of 36.52 mm. Assuming that the light at the output surface of the fiber has a Lambertian angular distribution and a uniform spatial distribution, the illumination intensity and irradiance in the peripheral FOV region are much

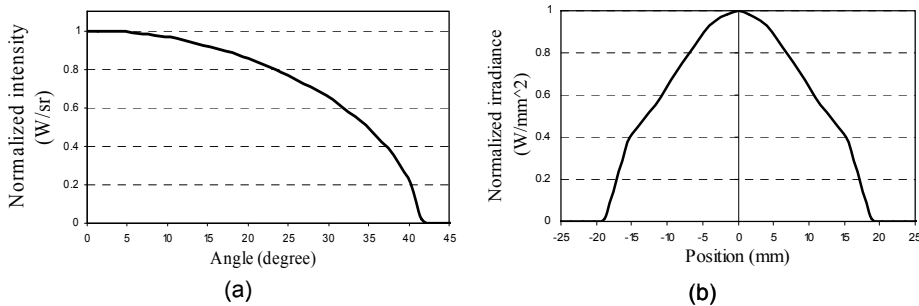


Figure 8.37 (a) Normalized intensity and (b) normalized irradiance on an object with a fiber-bundle illumination system.

weaker, as shown in Fig. 8.37. The irradiance at 15 mm from the optical axis is only 40% of the on-axis irradiance. Therefore, this intensity distribution characteristic cannot be adapted to a wide-angle endoscope.

When only one fiber bundle is used, the peak of the illumination does not appear on the optical axis of the objective lens because the optical axis of the fiber bundle shifts away from the optical axis of the objective lens. Using a wedge prism or tilting the fiber bundle can shift the center of the illumination to the center of the observation field at a fixed working distance.

To obtain a more uniform illumination, the illumination fiber bundle can be split into two or more groups that are arranged symmetrically at the distal end of the endoscope. The fiber bundle can be packed into a circular ring at the distal end. The exit surfaces of each subbundle can be made convex, concave, or inclined in different directions to achieve a uniform illumination. Figure 8.38(a) shows one example in which the fibers are split into two groups with surfaces obliquely polished relative to the optical axis so that the illumination light is expanded to the edge of the field, achieving a wider and more uniform illumination.

Figure 8.38(b) shows another configuration that increases the illumination field and improves illumination uniformity. The fibers from the fiber bundle are divided into several groups and arranged symmetrically around the optical axis. Each group is arranged to point in a different direction to achieve a wider illumination field and better illumination uniformity.

One simple and effective method commonly used in a conventional endoscope is to randomize the pointing direction of each fiber at the output end by twisting or bending the fiber bundle. The pointing directions of the fibers in the outer layer change more than the fibers close to the center. This method can be applied to both rod-shaped and ring-shaped fiber bundles.

8.5.3 Illumination systems with fiber bundles and lenses

In order to illuminate a larger region, a negative lens with a strong curvature can be placed in front of the output surface of the fiber bundle. Figure 8.39 shows the way in which the negative lens diverges the rays and the potential issues this may

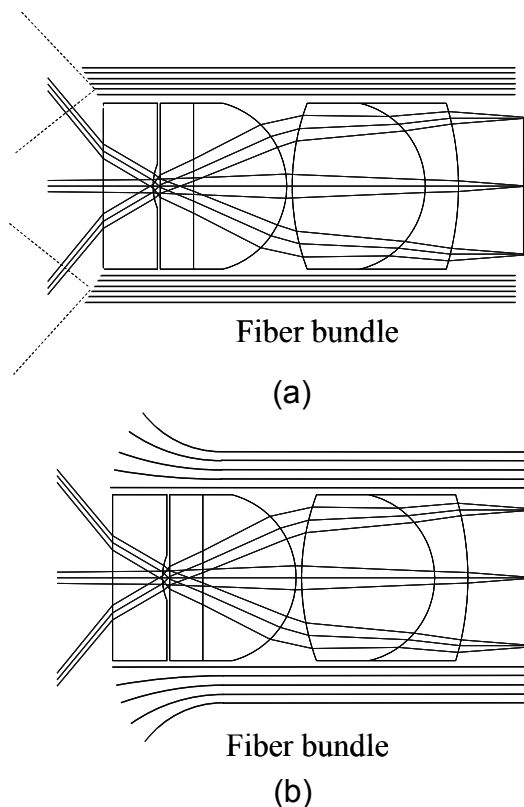


Figure 8.38 Variations of illumination systems using split fiber bundles. (a) The fiber surfaces are obliquely polished relative to the optical axis, and (b) each set of fibers points in a different direction.

cause. The ray paths with and without the negative lens are shown by the solid and dashed lines in Fig. 8.39(a). The plano-concave lens bends the rays toward the edge of the negative lens, resulting in a wider divergent angle. One potential issue is that the edge rays may encounter the lens edge given that the diameter of the plano-concave lens is limited by the available space. The divergent angle can be further extended by the biconcave negative lens, as shown in Fig. 8.39(b). However, the edge rays may be reflected back because of the TIR on the second concave surface, or they may encounter the edge of the lens. Because of the TIR, the concave surface always faces the output surface of the fiber light guide.

As the radius of the plano-concave lens becomes smaller, the divergent angle becomes larger, and a larger area is illuminated. However, the intensity of the light in the marginal portion is considerably low. The irradiance of the peripheral region is much smaller than that of the central region. The simple negative lens only bends the ray toward the edge gradually, as shown in Fig. 8.39(a); it is unable to redistribute enough light from the central region to the edge to obtain a uniform illumination over the entire illumination field.

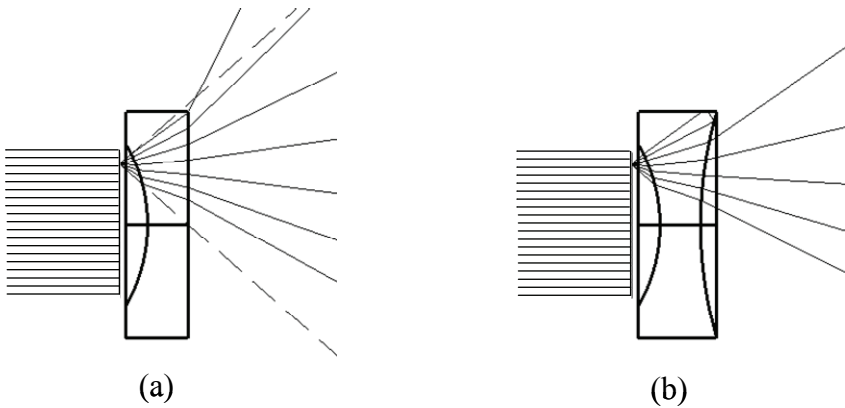


Figure 8.39 Single spherical lens that increases the field of illumination: (a) plano-concave lens and (b) biconcave lens. The dashed rays are the two edge rays from the fiber without a lens.

An aspherical surface has a different local curvature and can be designed to redistribute more light from the central region to the peripheral region. Figure 8.40 shows one example in which the first aspherical surface bends the light toward to the peripheral region unevenly, with the stronger bending close to the optical axis. Figure 8.40(a) shows that an aspherical surface is very effective in distributing the collimated light to a specific direction. Even when the input light is divergent, the aspherical surface is still more effective than the spherical lens in bending light to the edge, as shown in Fig. 8.40(b). In order to avoid the TIR on the lens surface, the aspherical surface usually faces the fiber bundle.

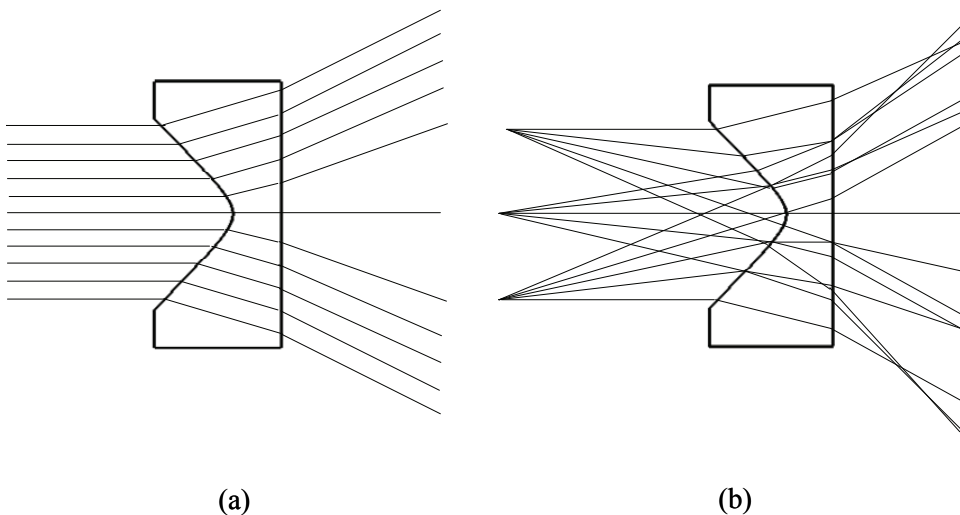


Figure 8.40 Aspherical lens redistributes light from the center to the edge when the input light is (a) collimated and (b) divergent.

To illustrate the improvement of illumination uniformity from a regular concave lens and an aspherical lens, normalized irradiance and intensity of the light distribution on the object plane are plotted in Fig. 8.41. The fiber bundle has a diameter of 1.5 mm and an NA of 0.659. The object plane is 20 mm from the lens. As a comparison, the irradiance and intensity from the fiber bundle with a flat window are also plotted ($R1 = R2 = 0$). Both conventional spherical and aspherical concave lenses can improve the illumination field and uniformity, but an aspherical surface is more effective in providing uniform illumination by diverging rays more strongly close to the optical axis, as shown in Fig. 8.40. The radius and the conic constant c can be optimized for different inputs. However, the uniformity is still not optimal because the ray angle from the fiber bundle is large and the angular distribution is not uniform.

An aspherical surface is more effective in providing uniform illumination when the incident light is small or when the emitting surface is small. This result occurs because the variation of the incident angle on each point of the aspherical surface is smaller, so the local curvature can be tuned to direct the incident light to a desired region. Figure 8.42 shows the normalized irradiance and intensity plots on a plane 20 mm away from the aspherical lens when the divergent angle of the light from the fiber bundle is ± 20 deg. The aspherical lens increases the divergent angle dramatically, from 20 to 65 deg, and the intensity is higher at a large angle. The object is illuminated uniformly up to ± 20 mm, compared to ± 5 mm without the aspherical lens. When the divergent angle of the light from the fiber bundle becomes smaller, the aspherical surface can be fine tuned to provide more uniform intensity and irradiance distribution over a larger area.

When a laser or LED is used as the light source in endoscopic illumination, the aspherical lens is a perfect beam-shaping element that provides uniform illumination.

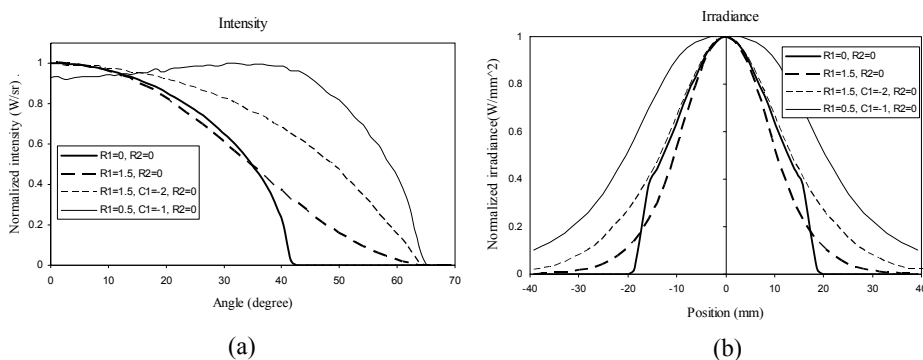


Figure 8.41 (a) Normalized intensity and (b) normalized irradiance distribution of the light on the object plane when an additional lens is placed in front of the fiber bundle. The distributions of the light with a flat window are the curves where $R1 = R2 = 0$. $R1$ and $R2$ are the radii of the first and second lens surfaces. $C1$ is the conic constant of the first surface. The object plane is 20 mm from the lens.

Positive lenses have also been used to redistribute light. Figure 8.43(a) shows one example. Instead of diverging the light from the fiber bundle directly, the positive lenses focus the light and then let the light diverge. Since the light is focused first and then diverges, the diameter of the positive lens is smaller than that of the negative lens that is used for the same purpose. In order to further reduce the lens diameter and reduce Fresnel loss, the first positive lens is usually designed as a rod lens and glued to the surface of a fiber light guide, as shown in Fig. 8.43(b). The first lens also acts as a light guide to reflect the off-axis light back to the illumination field through TIR or reflective coating. A positive lens group has a better performance in uniformity; however, this lens group usually needs more than one positive lens, and it has a longer total length.

The illumination methods discussed above are based on nonimaging methods. There is no direct imaging relationship between the surface of the fiber

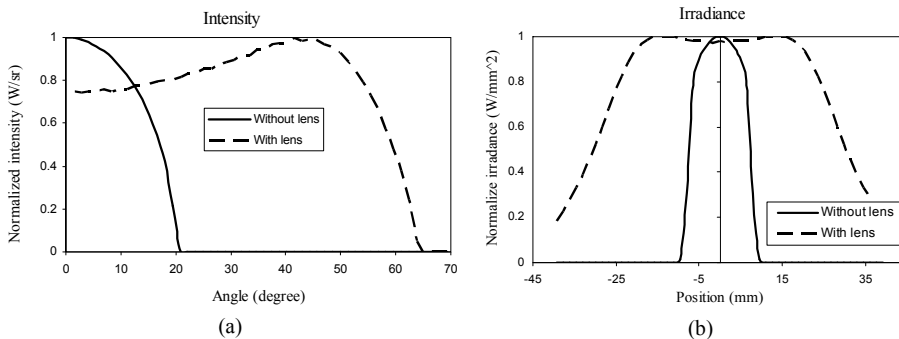


Figure 8.42 (a) Normalized intensity and (b) normalized irradiance distribution of light on the tissue plane for a low-NA fiber bundle with an aspherical lens.

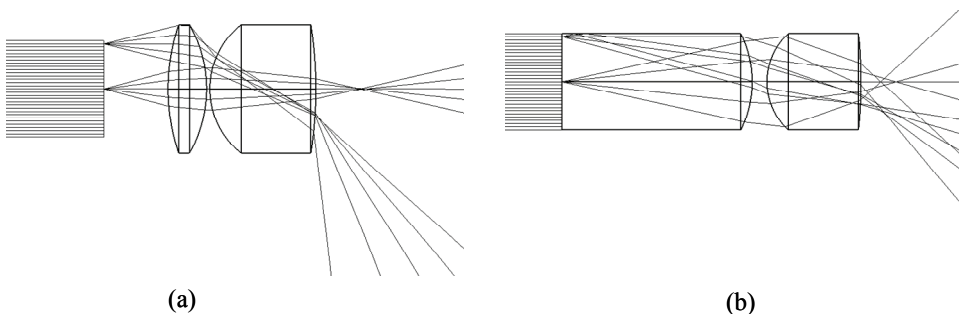


Figure 8.43 Fiber-optic illumination systems with positive lenses: (a) two thin singlets and (b) one thick and one thin positive lens. The first singlet is thick and works as a light guide.

bundle and the object plane. The illumination by a nonimaging configuration has a significant falloff in the peripheral region. Therefore, the effective illumination field is much smaller than the full extent of the illuminating light's reach. In order to achieve enough illumination uniformity within the FOV of the imaging lens, the illumination field should be reasonably wider than the FOV of the imaging system. Therefore, a significant amount of light is wasted.

A different type of illumination system for endoscopic imaging involves imaging the output surface of the light guide onto the observation plane. There exists an imaging relationship between the output surface of the light guide and the observation region. Therefore, the illumination field can be well controlled, and the illumination light does not reach any region other than the imaging field. This configuration is advantageous in that it not only has higher light efficiency but also reduces the background light scattered from outside of the observation region. As a result, the observed image is brighter, and image contrast is higher.

Figure 8.44 shows two illumination systems based on these imaging methods. The configuration in Fig. 8.44(a), where two positive lenses image the output surface of the fiber bundle onto the object plane, provides a well-defined illumination region, but it causes a major issue in that the structural features of the fiber light guide show up on the observation plane. Figure 8.45(a) shows the simulated irradiance on the observation plane; the core diameter is exaggerated to clearly illustrate this drawback. The fiber cores, which are bright regions, are clearly shown in the central FOV, and the images are distorted in the peripheral region as a result of the aberrations in the illumination system. The structural patterns make it very difficult to conduct a proper observation of the tissue.

There are two approaches to removing the structural pattern. One is to move the object away from the conjugate plane. In order to remove the pattern completely, a significant displacement is required. Also, endoscopic imaging systems usually have a large depth of field. The structural pattern may still be seen within its depth of field. Therefore, this method is not practical for endoscopic illumination.

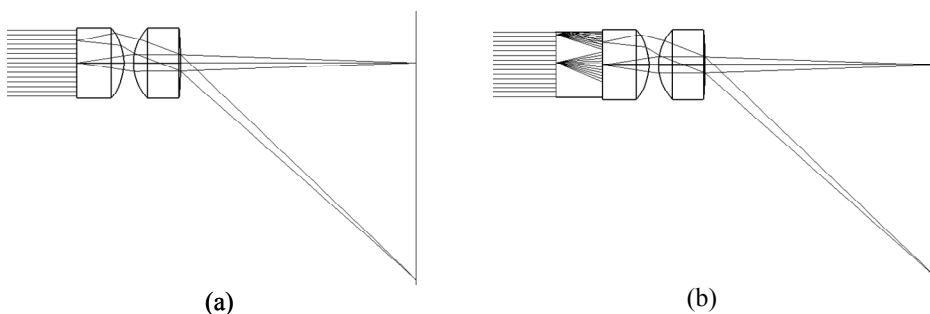


Figure 8.44 Illumination systems based on imaging methods. (a) Two positive lenses image the output surface of the fiber bundle to the object plane, and (b) two positive lenses image the output surface of the light guide to the object plane.

The other method is to move the fiber bundle away from the illumination lenses and place a light guide between the fiber light guide and the lenses, as shown in Fig. 8.44(b). The light guide homogenizes the spatial distribution of the light emerging from the fiber bundle. It can be either a solid light guide made of plastic or glass or a hollow light guide. Instead of imaging the fiber-bundle surface onto the observation surface, the optical system images the light guide surface onto the observation surface. While retaining the advantages, namely, the high light efficiency and clear edge of the illumination field, this illumination system can eliminate the nonuniformity caused by the structural features of the fiber bundle. Figure 8.45(b) illustrates simulated irradiance on the observation surface. The fiber structure is removed, and the illumination is more uniform with some falloff due to the cosine-fourth factor.

One potential advantage of the imaging-based illumination method is that the light distribution on the light guide surface can be apodized to provide uniform illumination on the observation surface, or it can even provide a specified illumination to compensate for inherent cosine-fourth falloff.

8.5.4 Illumination with LEDs

LEDs are ideal light sources for endoscopes because of their compact size and low power consumption. A wireless system is possible with LEDs as the illumination light source because LEDs can be operated by a battery.

LEDs can be either mounted in the proximal end or mounted directly in the distal end of the endoscope. When LEDs are mounted in the proximal end, a fiber bundle or a plastic light guide is required to convey the light to the distal end.

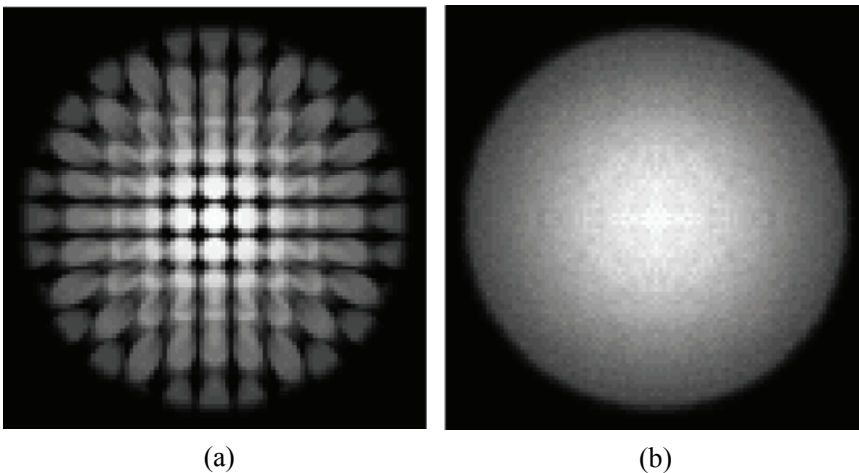


Figure 8.45 Simulated irradiance on the observation surface for the illumination system in Fig. 8.44. (a) Two positive lenses image the fiber-bundle surface onto the observation plane, and (b) two positive lenses image the light guide surface onto the observation plane.

The lenses discussed in the previous section can be used here to increase the illumination field and improve uniformity.

When three primary LEDs are used to capture a color image, RGB LEDs can be either turned on sequentially to obtain one image for each color or turned on simultaneously to obtain a color image. For both methods, the illumination light distribution should be similar for each color. This can be achieved by several configurations. The first configuration involves mixing the spectrum before the light is coupled into the light guide, as shown in Figs. 8.46(a) and 8.46(b). Two dichroic beamsplitters combine the light from RGB LEDs, and the light is then coupled into a fiber bundle or light guide, which also makes the light uniform, to convey the light from the LEDs onto the observation field. In Fig. 8.46(a), the light from LEDs with different wavelengths is coupled into the fiber bundle without mixing the colors completely. There may be some color nonuniformity in the illumination.

Figure 8.46(c) shows an alternative configuration. The light from LEDs with different wavelengths is coupled into the cylindrical light guide directly without the dichroic combiner and illuminates the object. The cylindrical light guide in this configuration makes the light uniform as well as mixing color. The light distribution on the object plane is similar to the illumination system with ring-shaped fiber bundles.

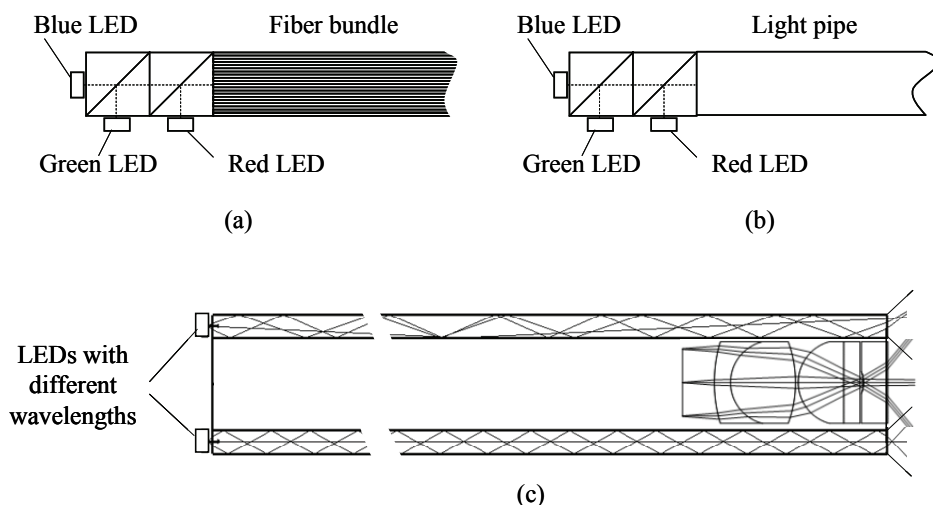


Figure 8.46 Illumination systems using LEDs at the proximal end. (a) A fiber bundle delivers the light from the proximal end to the distal end, (b) the square or round light guide delivers the light from the proximal end to the distal end, and (c) the cylindrical light guide delivers the light from the proximal end to the object. In (a) and (b), the light from the LEDs with different wavelengths is combined with the dichroic combiners before it enters the fiber bundle or light guide. In (c) the cylindrical light guide mixes the light from LEDs with different wavelengths.

There are several advantages when LEDs are mounted outside the tube. There are fewer space constraints when mounting more or larger LEDs, the heat generated by LEDs is dissipated and is not transferred to the distal end close to the tissue, and better uniformity of illumination can be achieved through the light guide. However, some light is lost as a result of coupling and absorption.

The use of LED chips at the distal end of the endoscope eliminates the need for a fiber bundle or light guide to deliver the light to the desired object. The light loss due to the light coupling and absorption is also avoided.

Unlike fiber bundles with a moderate divergent angle, an LED chip has a Lambertian distribution over 180 deg. When a plurality of LED chips is mounted symmetrically at the distal end of an endoscope, the illumination field is wider and more uniform than when using the fiber bundle to deliver the light from the proximal end to the distal end because the NAs of the fiber bundles are limited.

The heat generated by an LED is from the PN junction in the LED chip, and is not in the form of radiative heat. The heat can be conducted out of the endoscope tip using a conductive layer or through the endoscope tube.

Miniature optical elements, such as polarizers, filters, and beam-shaping elements, can be integrated with LEDs at the distal end to provide special illumination for some applications, for example, fluorescence and polarization imaging. A variety of illumination configurations exist for mounting LED chips at the distal end of an endoscope.²¹ The extent of illumination and uniformity are controllable by positioning and angling the LED chips. In Fig. 8.47(a), the LED chips are mounted directly in the distal end of the endoscope. In Fig. 8.47(b), LEDs are tilted relative to the optical axis to expand the light distribution. In Fig. 8.47(c), LEDs are mounted between the aperture stop and the first negative lens. This configuration is possible because the clear aperture of the concave surface is much smaller than the convex surface. This configuration is compact; however, it is not practical when the distance between the first negative lens and the aperture stop is too small to mount the LEDs. The other two drawbacks are Fresnel reflection from the first surface and the limited freedom to adjust the illumination field and uniformity. The light reflected from the first surface may pass through the aperture and reach the sensor, resulting in a degraded image.

8.6 Wireless Endoscopes

8.6.1 Introduction

Wireless endoscopy, also referred to as capsule endoscopy, is a significant technical breakthrough in small-bowel imaging. With conventional technologies, only the proximal (esophagus, stomach, and duodenum) and the distal (colon) portions of the gastrointestinal tract are visible. In the past, the 20 ft or so of small intestine between these two portions was essentially unreachable.

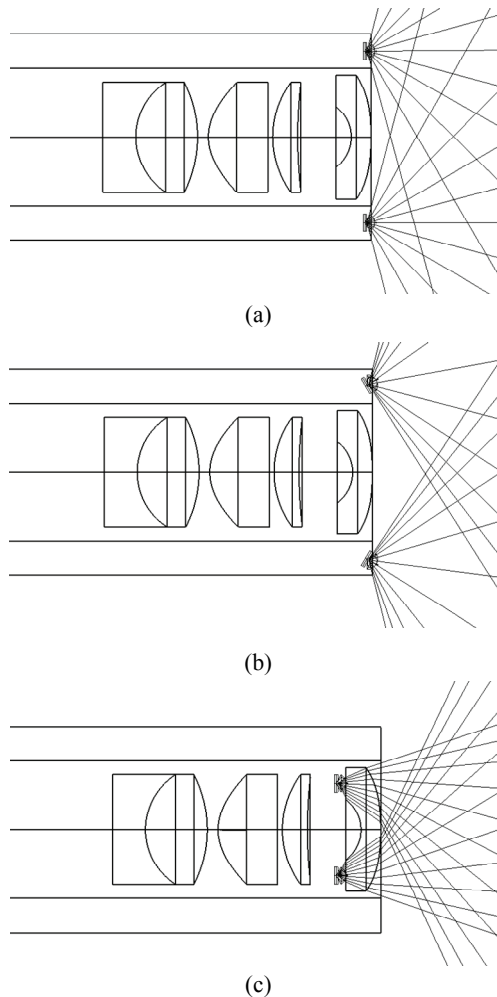


Figure 8.47 Illumination systems using LED chips at the distal end. (a) All LEDs point in the same direction, with the surfaces of all LEDs perpendicular to the optical axis, (b) LEDs point in different directions to improve FOV and uniformity, and (c) LEDs are mounted inside the objective lens.

Capsule endoscopes provide clinicians with a way to visualize the entire small bowel.³ Once the endoscope is inserted and activated, it begins to record images at a speed of a few frames per second and transmits them to a belt-pack receiver. A typical capsule endoscope examination takes approximately seven hours. Capsule endoscopy has no therapeutic capabilities. Any lesions found by the capsule endoscope must be further investigated, either through biopsy or verification by other imaging modalities.

Figure 8.48 shows a basic configuration of a capsule endoscope. Inside the cylindrical tube and dome, there are:

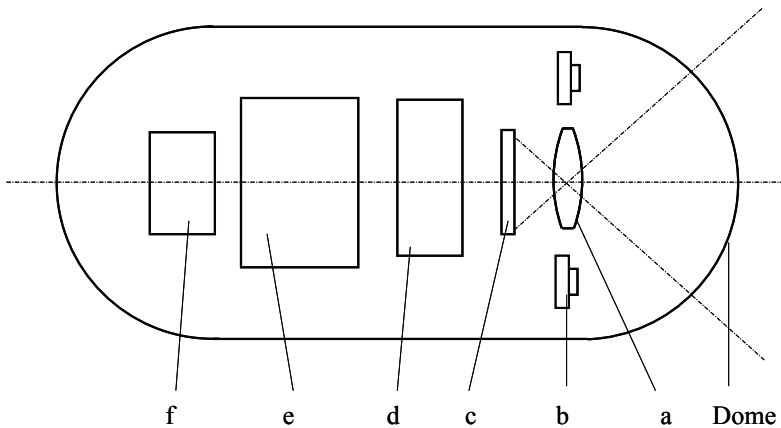


Figure 8.48 Basic configuration of a capsule endoscope: (a) imaging optics, (b) light source, (c) image sensor, (d) battery, (e) electronics, and (f) antenna/transmitter.

- Imaging optics that image the object to the sensor,
- A light source (and necessary illumination optics) that provide illumination onto the interior of the body,
- An image sensor (CCD or CMOS) that captures images,
- Electronics that operate the endoscope and digitize the image,
- A battery that supplies power to the image sensor and electronics, and
- An antenna/transmitter that receives the operation signal and transfers the image out of the patient.

The capsule is typically 1 inch long with a diameter of 0.5 in. The viewing window usually has a spherical or nearly spherical shape in order to enable smooth insertion into a body cavity. The imaging system, consisting of a lens and dome, typically has a wide-angle FOV, preferably 100 deg or more, and a large depth of field, for example, 0–30 mm, so that the pathological changes on the biological surface are not overlooked. Owing to space constraints, the distance from the dome surface to the image sensor is usually less than 10 mm.

Because of the space constraints, both the illumination light source and the imaging system are located within the same optical dome. Light reflected by the dome surface may reach the image sensor, thereby deteriorating the imaging quality. Therefore, it is necessary to optimize the shape of the viewing dome, the locations of the LEDs, and the imaging system to prevent the stray light from reaching the image sensor.

Figure 8.48 shows only the essential components necessary for conventional reflectance imaging. By adding some additional components, such as polarizers and fluorescence filters, in front of the corresponding light sources and imaging systems, a capsule endoscope can obtain polarization and fluorescence images on the same location, which helps to identify the diseased region with higher sensitivity and specificity.

8.6.2 Imaging systems

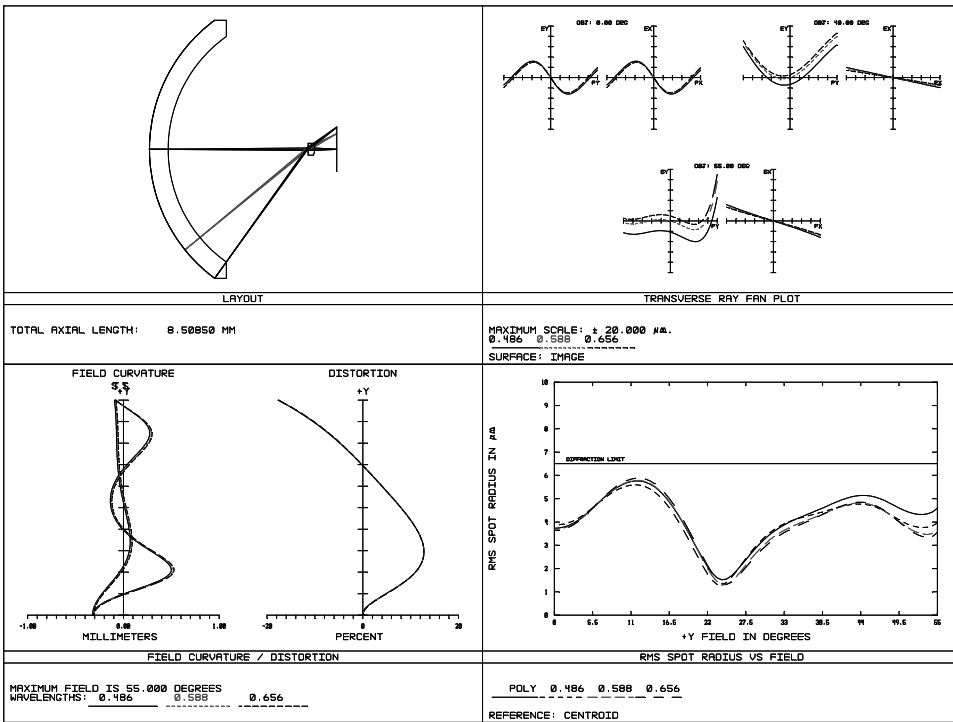
When a capsule endoscope is inserted into a body cavity, the peripheral portion of the viewing window generally is in contact with an internal wall of the body cavity, but the central part of the viewing window is not. In order to accurately observe the internal wall of the cavity, the imaging system should consistently bring the contacted region of the internal wall into focus.

The viewing dome is usually made of a plastic material. To prevent deformation with respect to humidity and degradation of optical performance, the material for the dome should have a low moisture-absorption rate, low residual metals, and adequate hardness. In addition, the material for optical domes should be able to withstand the extreme chemical environments within the stomach and intestine, where the pH value ranges from 2.0 to 9.0. The optical quality and transparency of the viewing dome need to be preserved during the entire endoscopic process. Polycarbonate and cyclo olefin polymer ZEONEX® 330R are two suitable materials. Polycarbonate has a moisture absorption rate of 0.15%, 1.36 ppm of residual metals, and a pencil hardness of HB. Cyclo olefin polymer ZEONEX® 330R has a moisture absorption rate of 0.01%, 0.06 ppm of residual metals, and a pencil hardness of 3H.

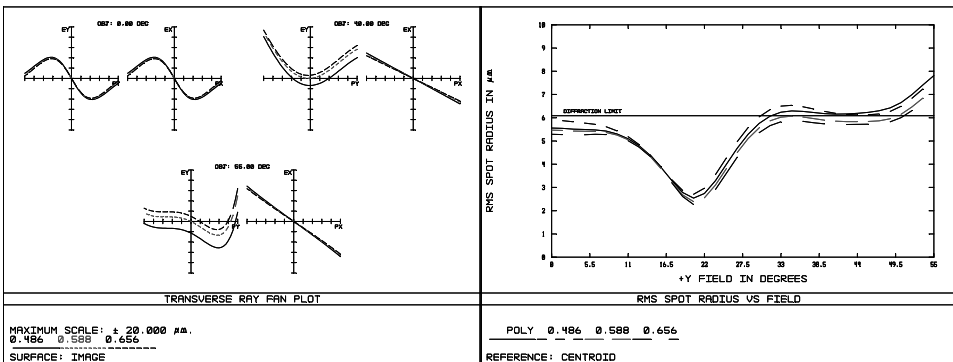
The optical system shown in Fig. 8.49 includes a spherical viewing dome, a singlet aspherical imaging lens, and a solid-state sensor. The objective is a landscape lens and has a FOV of 110 deg. The NA on the image side is 0.085. The aperture stop is placed in front of the imaging lens and is at the center of the viewing dome. The spherical viewing dome is a weak negative lens with uniform thickness and has a focal length of 115 mm. The performance of this system when the object is in contact with the spherical viewing window is plotted in Fig. 8.49(a). Given that this capsule endoscope is a single-element imaging system, there are considerable chromatic aberrations. Lateral chromatic aberration can be corrected by postimage processing. The RMS spot size is smaller than the Airy radius for the entire field. The major uncorrected aberrations include distortion and astigmatism. Both aberrations are difficult to correct because the imaging system only consists of a single lens and the aperture is in front of the lens.

When an object to be observed is flat and does not make contact with the spherical viewing window, for example, at a distance of 10 mm from the viewing window, the change of the on-axis performance is negligible, but the off-axis performance of the imaging system is degraded to some degree (Fig. 8.49(b)). Therefore, this imaging system can form a good image of the contacted internal wall in the peripheral region and the noncontacted internal wall in the central region. These are the normal working conditions of a typical capsule endoscope.

Figure 8.50 shows another design using a single aspherical lens. The aperture stop is not at the center of the window, but it is still in front of the imaging lens. The viewing dome has the smallest thickness along the optical axis and slightly increases as it moves away from the optical axis. The on-axis performance is slightly better than that of the objective lens shown in Fig. 8.49, but the off-axis performance is relatively poorer because of the field curvature and astigmatism.



(a)



(b)

Figure 8.49 Optical system for a wireless capsule endoscope: (a) the spherical object contacting the spherical viewing dome and (b) the flat object at 10 mm from the spherical viewing dome.

The RMS spot size is smaller than the Airy radius in up to 70% of the field. The advantage of this design is that the distance from the viewing dome to the sensor is shorter than the same distance in the design shown in Fig. 8.49.

In order to improve system performance, another lens can be added in front

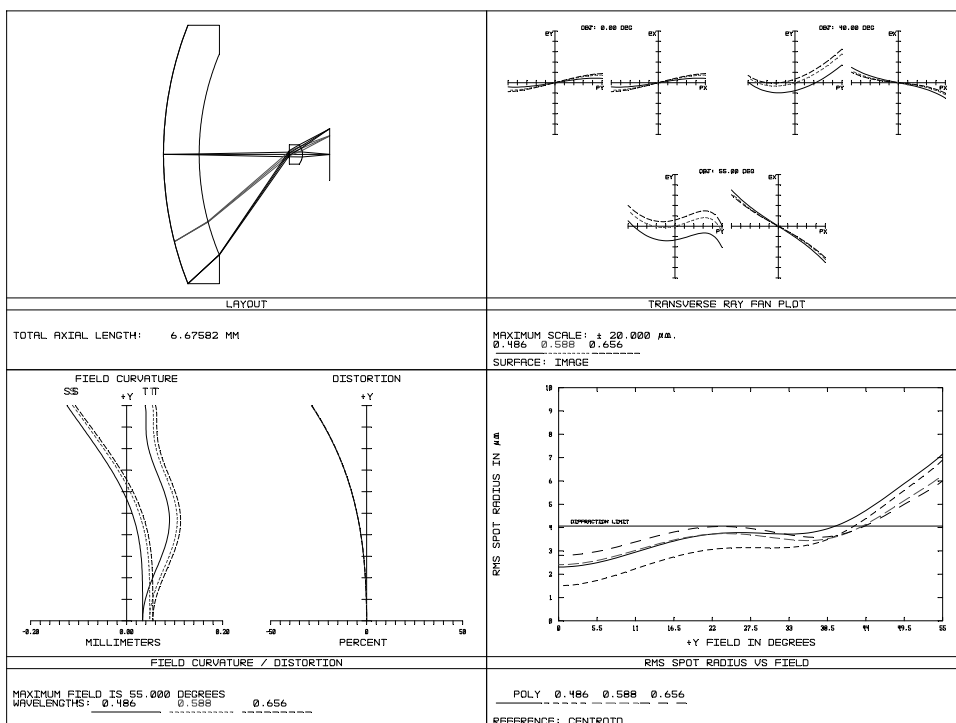


Figure 8.50 Optical system for a capsule endoscope where the aperture is not at the center of the spherical viewing dome.

of the aperture stop, making the objective lens symmetrical to the aperture. Figure 8.51 shows an objective lens consisting of two elements (not including the viewing dome). Both elements can be made from plastic, such as polystyrene, polycarbonate, or COC. The odd-order aberrations, such as coma, distortion, and lateral chromatic aberration, are reduced because of the symmetrical configuration. One or more surfaces of the objective lens can be aspherical to further reduce the aberrations.

Generally, the quantum efficiency of the sensor decreases as the ray angle increases. To reduce the incident angle of the ray on the sensor, another positive lens can be added between the aperture stop and the sensor, as shown in Fig. 8.52. This additional lens has an FOV of 120 deg and an NA of 0.085. The designed working distance is 10 mm. The first and second surfaces of the viewing dome are concentric, and the aperture stop is placed at the center of the dome. The performance of this objective lens is similar to that of the previous single-element design with an aspherical surface, as in Fig. 8.49(b). Similarly, lateral chromatic aberration is not corrected because of the asymmetrical configuration and materials used in the design.

The viewing domes discussed so far have either a uniform thickness or a thickness that slightly increases as it moves away from the optical axis. When the

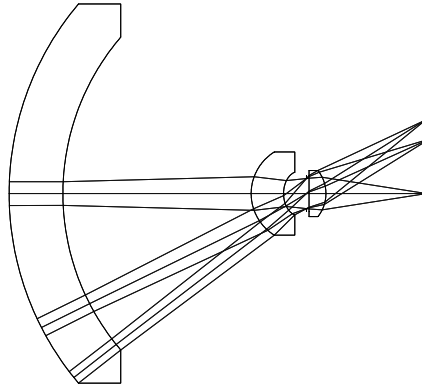


Figure 8.51 Objective lens with two elements arranged symmetrically to the aperture stop.

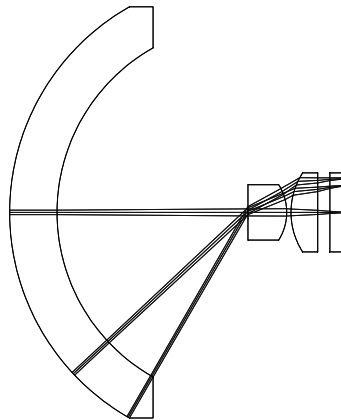


Figure 8.52 Objective lens with two positive lenses that reduce the ray angle on the sensor.

aperture stop is at the center of the viewing dome, the dome only generates a small amount of distortion. When the aperture stop moves away from the center of the viewing dome, the viewing dome introduces astigmatism, field curvature, and lateral chromatic aberration, but in amounts much smaller than those from most imaging lenses.

Instead of a concentric viewing dome, the radii of the viewing dome can be optimized to balance the aberrations generated by the imaging lens.²² Figure 8.53 shows one example. The optical parameters are the same as for the previous design. The astigmatism is less severe because the viewing window forms a thick meniscus, which helps to control the astigmatism. Compared to the previous

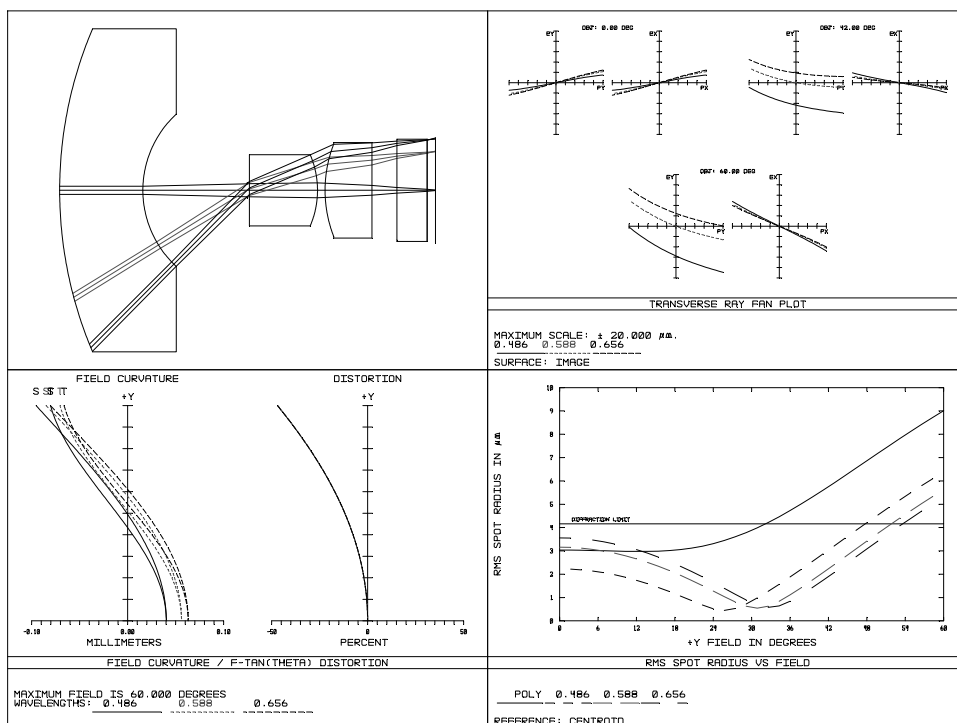


Figure 8.53 Objective lens with a viewing dome as a lens element to balance the aberrations generated by the imaging lens.

designs, this viewing dome has stronger power. The viewing dome can also be used to expand the FOV of the imaging system.

The distortion of the objective lens used in wireless endoscopes is larger than that of the objective lens used in flexible and rigid endoscopes, mainly because of an asymmetrical configuration and a large FOV. Because of the limited space within wireless endoscopes, the effective method for reducing the distortion is to use aspherical surfaces. The inner surface of the viewing dome can be designed to be aspherical to reduce the distortion more effectively because the viewing dome is the element away from the aperture stop.²³

8.6.3 Illumination systems

While the viewing dome can be optimized to improve the image quality of the objective lens, one additional consideration is the illumination because the imaging and illumination systems are in the same compartment.

The basic requirements of an illumination system in a wireless endoscope include a large FOV and good illumination uniformity. Another important consideration is whether the illumination light reflected from the viewing dome surfaces by TIR and Fresnel reflection will reach the solid-state image sensor.

Figure 8.54 shows a typical configuration of an illumination system in a wireless endoscope. A group of LEDs is arranged symmetrically around the

imaging lens. LEDs with Lambertian angular distribution are usually used to provide a large illumination field.

The irradiance distributions at different planes are plotted in Fig. 8.55. The light distribution on the external surface S1 of the viewing dome is projected on a flat surface and plotted in Fig. 8.55(a). The irradiance distribution on the flat plane S2 is shown in Fig. 8.55(b); there is significant falloff from the center to the edge. The light distribution on surface S3 is shown in Fig. 8.55(c). Given the plane's elliptical shape, Fig. 8.55(c) shows the projected light distribution on a flat surface perpendicular to the optical axis. The irradiance on plane S4, 10 mm from the viewing dome, is shown in Fig. 8.55(d). Similar to Fig. 8.55(b), the illumination falloff is obvious.

As indicated earlier, the imaging and illumination systems are located within the same viewing dome. Illumination light emitted from the light source is partially reflected by the dome surfaces and may pass through the entrance pupil of the objective lens and reach the image sensor, significantly deteriorating image contrast. Therefore, the shape of the viewing dome and the positional relationship between the illumination and imaging systems inside the dome must be optimized to prevent light from reaching the sensor.

To find out whether the light from the LEDs can reach the sensor or not, one simple approach is to trace the ray from the sensor to the LED mounting plane, as shown in Fig. 8.56. The light distribution on the LED's mounting plane determines the relative amount of light that will reach the sensor from the LED at the corresponding location. To trace the rays more efficiently, a spatially and angularly uniform light source can be placed at the entrance pupil or aperture stop instead of the sensor plane, and the subtended angle can be set at the angular

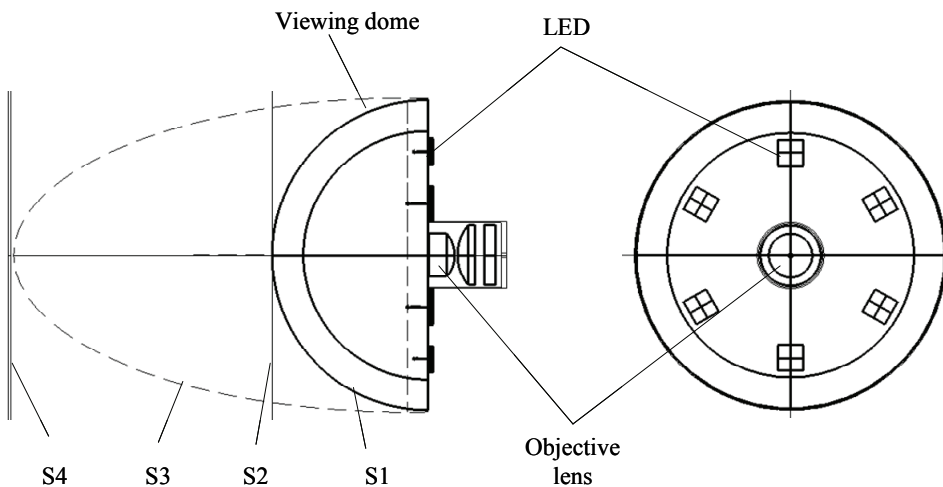


Figure 8.54 Typical configuration of an illumination system in a wireless endoscope.

FOV of the objective lens. Figure 8.56 shows that the light reflected from both surfaces of the viewing dome has a chance to reach the sensor. Besides optimizing the viewing dome and the LED's position, one direct method to reduce the light reaching the sensor is to apply an efficient AR coating on both surfaces of the viewing dome.

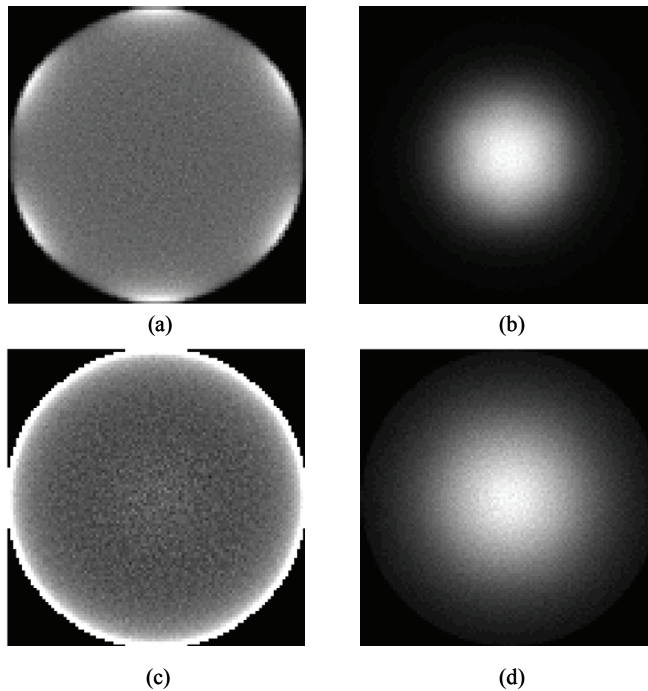


Figure 8.55 Light distributions at different planes for the illumination system in Fig. 8.54. (a)–(d) are the irradiances at surfaces S1–S4, respectively.

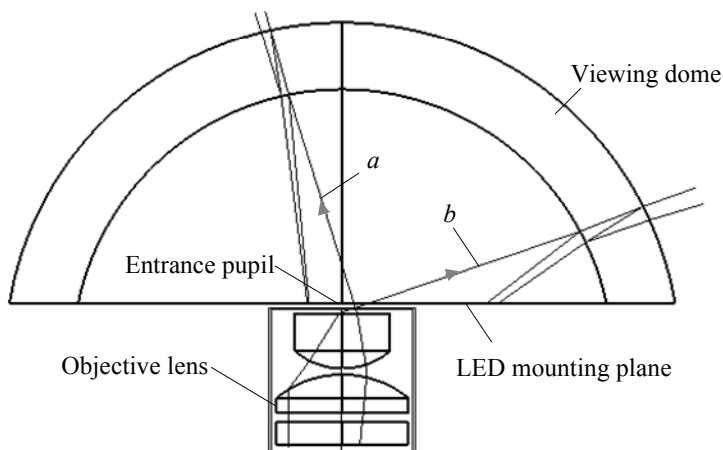


Figure 8.56 Rays are traced from the entrance pupil to the viewing dome.

Figure 8.57 shows the simulation results for the objective lens in Fig. 8.52. The viewing dome is concentric, and the aperture stop is at the center of the viewing dome. To illustrate the effect of the viewing dome, the subtended angle of the light source at the aperture stop is set to 160 deg, larger than the FOV of the objective lens. The rays are traced from the aperture stop, as shown in Fig. 8.57(a). The aperture stop is at the center of the viewing dome; therefore, the light reflected from the dome surfaces returns to the regions closest to the center of the viewing dome. The irradiance distribution at the LED mounting plane is plotted in Fig. 8.57(b). The dashed circle indicates the inner diameter of the viewing dome. Figure 8.57(c) is a magnified view of the central region of Fig. 8.57(b). The dark region in the center is the aperture stop, and the white ring shows the region where the reflected light hits the LED mounting plane. This simulation result shows that for the concentric viewing dome, when the aperture stop is in the center of the viewing dome, light from the LEDs will not reach the sensor as long as they are not mounted too closely to the aperture stop.

Figure 8.58 displays the simulation results for an objective lens whose

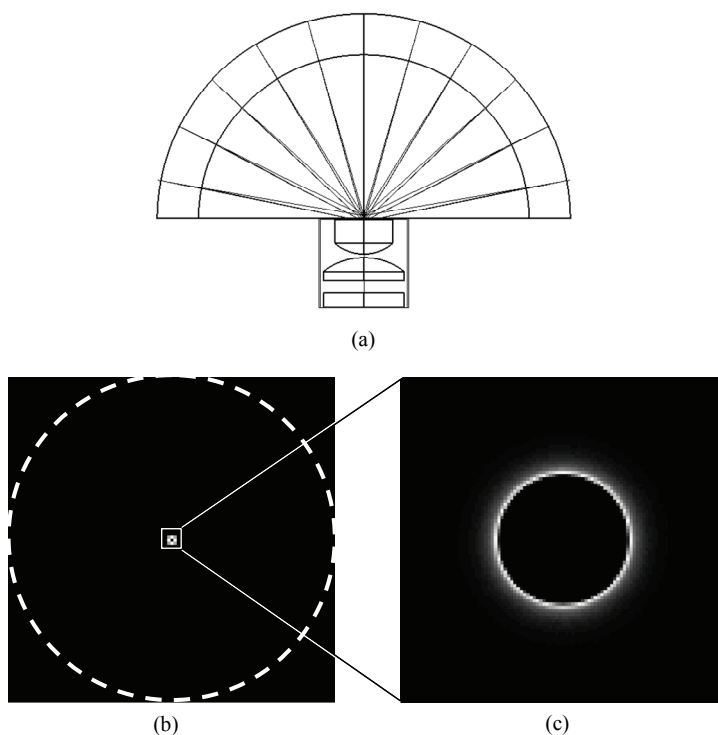


Figure 8.57 Simulation results for the objective lens in Fig. 8.52. (a) Diagram of the rays traced from the aperture stop to the LED mounting plane, (b) the irradiance on the LED mounting plane, and (c) a magnified view of the central portion of (b). The dashed circle is the edge of the region on which the LEDs can be mounted.

aperture stop is not at the center of the viewing dome. Figure 8.58(a) is the ray-tracing diagram that shows that the reflected light from the viewing dome not only returns to the region close to the aperture stop but also spreads to the other region away from the center. Figure 8.58(b) is the irradiance plot of the LED mounting plane when the subtended angle of the light source located in the aperture stop is 160 deg. Again, the central black spot is the aperture stop where the simulation light source is located, and the dashed circle is the region of the LED mounting plane. Figure 8.58(c) shows the irradiance at the LED mounting plane for a light source having a subtended angle of 100 deg. The reflected light from the viewing dome concentrates toward the central region of the LED mounting plane. Therefore, LEDs can be mounted at the edge of a plane, and no reflected light will reach the sensor. Figures 8.58(b) and 8.58(c) show that when the aperture stop is not at the center of the viewing dome, the FOV is limited.

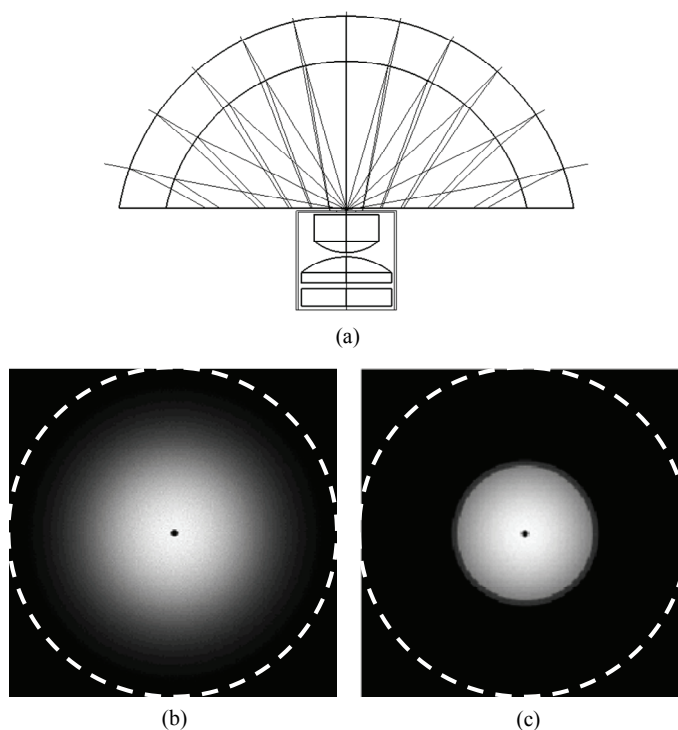


Figure 8.58 Simulation results for an objective lens whose aperture stop is not at the center of the viewing dome. (a) Diagram of the rays traced from the aperture stop to the LED mounting plane, (b) the irradiance on the LED mounting plane when the viewing angle is 160 deg, and (c) the irradiance on the LED mounting plane when the viewing angle is 100 deg. The dashed circle is the edge of the region upon which the LEDs can be mounted.

An elliptical dome is another option used for reducing stray light from dome surfaces.²⁴ An elliptical dome has two focal points; light rays from one focal point are focused at a second focal point. As shown in Fig. 8.59, a small portion of the light ray from the first focal point (dashed lines) where the LED is located is directed to the vicinity of the second focal point through the Fresnel reflection from both surfaces of the viewing dome. As the solid lines show, when the LED has a finite size, the reflected light rays still converge at the vicinity of the second focal point. Therefore, with a rotationally symmetric, elliptical viewing dome, the LEDs can be placed along the focal circle without introducing flare from the viewing dome.

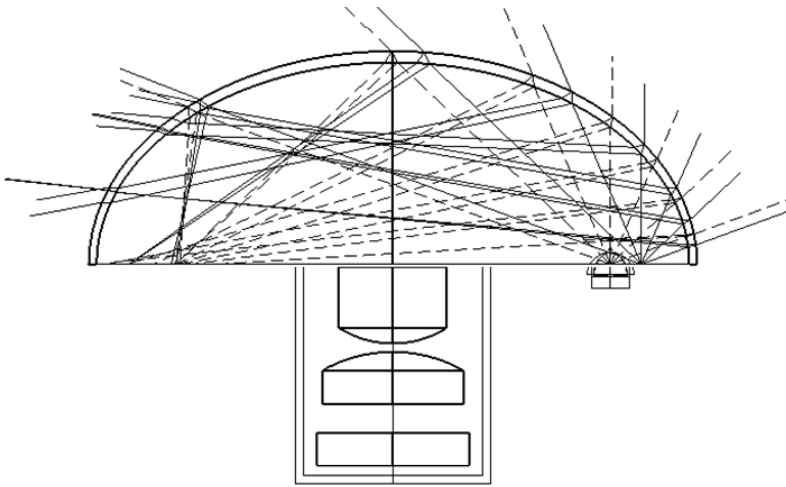


Figure 8.59 Capsule endoscope with an elliptical dome. Diagram of the rays traced from the LED at the focal point.

References

1. G. Berci and K. A. Forde, "History of endoscopy: what lessons have we learned from the past?" *Surg. Endosc.* **14**, 5–15 (2000).
2. H. H. Hopkins, "Optical principles of the endoscope," in *Endoscopy*, G. Berci, Ed., pp. 3–26, Appleton-Century-Crofts, New York (1976).
3. G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature* **405**, 417 (2000).
4. T. H. Tomkinson, J. L. Bentley, M. K. Crawford, C. J. Harkrider, D. T. Moore, and J. L. Rouke, "Rigid endoscopic relay systems: a comparative study," *Appl. Opt.* **35**, 6674–6683 (1996).
5. H. H. Hopkins, "Optical system having cylindrical rod-like lenses," U.S. Patent No. 3,257,902 (1966).

6. H. McKinley, "Endoscope relay lens," U.S. Patent No. 5,059,009 (1991).
7. S. J. Dobson and H. H. Hopkins, "A new rod-lens relay system offering improved image quality," *J. Phys. E: Sci. Instrum.* **22**, 450–455 (1989).
8. J. Hoogland, "Endoscope relay lens," U.S. Patent No. 4,946,267 (1990).
9. R. F. Horton and T. O. Bales, "New family of plastic endoscope designs," *Proc. SPIE* **2383**, 25–36 (1995) [doi: 10.1117/1.12.209032].
10. S. Fantone, P. Costa, W. Holmes, and F. Moll, "Optical viewing device," U.S. Patent No. 4,784,118 (1988).
11. D. Leiner, "Disposable rigid endoscope," U.S. Patent No. 4,964,710 (1990).
12. D. Leiner, "Optical system for endoscope," U.S. Patent No. 5,684,629 (1997).
13. K. Mitsujiro, S. Matsumoto, T. Takayama, T. Suzuki, and A. Hasegawa, "Optical system for endoscopes," U.S. Patent No. 5,793,539 (1998).
14. T. Igarashi, "Optical system for non-flexible endoscopes," U.S. Patent No. 5,933,275 (1999).
15. F. Lei, "Endoscope lens and an endoscope equipped with such a lens," International Patent Application WO/2000/037988 (1999).
16. H. Miyano, "Objective lens for variable field angle endoscope," U.S. Patent No. 5,748,385 (1998).
17. H. Miyano, "Objective lens for variable viewing depth endoscope," U.S. Patent No. 6,327,101 (2001).
18. C. Yamamoto, "Objective lens for endoscope," U.S. Patent No. 6,353,504 (2002).
19. M. Konno, "Optical imaging system with movable electronic sensor for focus control," U.S. Patent No. 6,537,208 (2003).
20. J. Hoogland, "Flat field lenses," U.S. Patent No. 4,545,662 (1985).
21. M. Farr, "Solid state illumination for endoscopy," U.S. Patent Application No. 20080045800 (2008).
22. O. Mang, S. Huang, Y. Chen, C. Lin, T. Lin, and Y. Kuo, "Distortion improvement of capsule endoscope image," *Proc. SPIE* **6430**, 643018 (2007) [doi: 10.1117/1.12.699742].
23. H. Fukuhori, "Capsule endoscope," U.S. Patent Application No. 20080045798 (2008).
24. H. Kislev, A. Glukhovskiy, G. Meron, and G. Iddan, "Optical system," U.S. Patent No. 6,934,093 (2005).

Appendix: Lens Prescriptions

Figure 4.13

EFL: 27.008

Image space NA: 0.1

Image height: 2.5

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
STO	Infinity	20.014		5.40	0
2	23.1250	7.000	N-BK7	9.22	0
3	-9.0528	2.000	N-SF2	9.13	0
4	-20.7465	23.845		9.33	0
IMA	Infinity			5.06	0

Figure 4.14

EFL: 32.0

Image space NA: 0.1

Image height: 2.5

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
STO	Infinity	24.116		6.40	0
2	22.1061	6.000	N-BK7	10.23	0
3	-10.4952	2.300	N-SF2	9.98	0
4	-24.8701	11.865		10.08	0
5	8.2885	3.000	N-SF2	7.15	0
6	5.3975	10.818		5.66	0
IMA	Infinity				0

Figure 4.15

EFL: 16.0

Image space NA: 0.25

Image height: 1.0

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
STO	Infinity	0.000		8.00	0
2	12.2413	3.000	K10	8.09	0
3	-13.3440	2.000	N-F2	7.83	0
4	341.8207	10.709		7.59	0
5	8.8888	2.500	N-BK7	6.06	0
6	-7.1212	2.000	N-SF5	5.60	0
7	-27.1433	6.000		5.07	0
IMA	Infinity				0

Figure 4.16

EFL: 4.0

Image space NA: 0.65

Image height: 0.25

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	6.6802	2.600	LLF2	6.48	0
2	-5.8897	0.900	LAF9	6.06	0
3	103.6375	0.000		5.73	0
STO	Infinity	2.040		5.73	0
5	4.6452	3.800	N-BK10	5.54	0
6	-7.3742	0.900	N-SF6	4.23	0
7	-12.1720	0.085		3.92	0
8	1.7573	1.700	N-LAK10	2.96	0
9	1.9406	0.340		1.33	0
10	Infinity	0.170	N-K5	0.71	0
IMA	Infinity			0.50	0

Figure 4.17

EFL: 2.0

Image space NA: 1.0

Image height: 0.1

Wavelengths : 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	8.8862	2.505	CAF2	4.56	0
2	-3.0048	0.624	F4	4.25	0
STO	-1164.042	0.167		4.32	0
4	5.1849	2.505	CAF2	4.37	0
5	-3.6040	0.596	N-PSK53	4.20	0
6	-5.0146	0.084		4.24	0
7	2.4628	0.835	N-LAK21	3.54	0
8	4.8250	0.084		3.16	0
9	1.1981	1.469	N-BK7	2.29	0
10	Infinity	0.140	1.52, 59.48	0.89	0
11	Infinity	0.170	N-K5	0.58	0
IMA	Infinity		N-K5		0

Figure 4.18

EFL: 17.03

Image space NA: 0.32

Image height: 0.9

Wavelengths : 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0	0
STO	20.3181	1.900	PSK52	11.59	0
2	102.2027	0.100		11.40	0
3	15.7484	4.710	CAF2	11.22	0
4	-26.5432	1.200	KZFS1	10.09	0
5	44.8967	10.488		9.57	0
6	15.6684	1.000	SF8	6.40	0
7	5.6367	3.774	CAF2	5.88	0
8	-22.0079	0.100		5.31	0
9	6.3434	4.633	SF3	4.99	0
10	4.1370	1.200		2.45	0
11	Infinity	0.170	K3	1.87	0
12	Infinity			1.81	0
IMA	Infinity				0

Figure 4.19

EFL: 4.0

Image space NA: 0.65

Image height: 0.25

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	Infinity	6.377		7.11	0
2	-4.8515	5.000	N-LAF7	6.46	0
3	-7.4062	5.898		9.41	0
STO	Infinity	11.729		9.20	0
5	38.4912	3.000	N-PK52	10.48	0
6	-8.5830	1.000	N-SF56	10.46	0
7	-23.3070	0.100		10.92	0
8	9.4242	5.000	N-PK52	11.00	0
9	-7.1377	1.000	KF6	10.64	0
10	34.2577	0.100		9.27	0
11	5.0990	2.500	N-PK52	8.42	0
12	11.8354	0.100		7.47	0
13	4.1908	3.658	N-SF11	6.28	0
14	2.5254	0.492		2.11	0
15	Infinity	0.170	N-K5	1.40	0
16	Infinity	0.253		1.19	0
IMA	Infinity				0

Figure 4.20

EFL: 4.0

Image space NA: 0.65

Image height: 0.25

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	Infinity	6.377		7.58	0
2	-5.8764	1.400	KZFS1	6.92	0
3	24.2273	4.400	N-LAF7	8.23	0
4	-9.9815	10.837		9.35	0
STO	Infinity	7.563		8.07	0
6	18.7786	3.000	N-PK52	8.46	0
7	-7.1186	1.000	N-SF56	8.32	0
8	130.3450	0.100		8.63	0
9	10.2074	4.000	N-PK52	8.94	0
10	-5.4513	1.000	KF6	8.85	0
11	-20.9998	0.100		8.64	0
12	5.0257	2.000	N-PK52	7.83	0
13	12.9973	0.100		7.24	0
14	3.6271	3.700	N-SF11	5.96	0
15	1.9978	0.450		1.82	0
16	Infinity	0.170	N-K5	1.20	0
17	Infinity	0.180		0.99	0
IMA	Infinity				0

Figure 4.21

EFL: 2.0

Image space NA: 1.2

Image height: 0.1

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	Infinity	5.000		6.37	0
2	4.4482	2.038	TPH55	5.76	0
3	-41.0841	0.599	LAH54	5.10	0
4	3.0591	2.038		4.06	0
5	-2.9671	3.237	S-NBH52	4.10	0
6	305.2876	3.596	N-PSK58	7.15	0
7	-5.7756	0.090		8.52	0
8	-16.6307	0.599	KF9	8.76	0
9	8.1740	6.000	FPL53	9.58	0
10	-7.7816	0.959	BPM4	10.57	0
11	-12.8853	0.060		11.50	0
12	14.2745	5.694	FPL53	12.35	0
13	-9.5254	0.659	BPM4	12.15	0
14	-32.8852	0.090		12.37	0
15	9.0586	5.634	FPL53	12.12	0
16	-16.3286	0.569	PBM5	10.79	0
17	24.0651	0.090		9.96	0
18	7.7459	2.338	N-PSK58	9.47	0
19	20.5779	0.060		8.56	0
STO	4.7841	2.248	N-PSK58	7.27	0
21	7.6801	0.060		5.76	0
22	2.3154	2.338	LAH75	4.25	0
23	0.5790	0.360	K4	0.99	0
24	Infinity	0.126	1.51, 51.0	0.79	0
25	Infinity	0.109	B270	0.47	0
IMA	Infinity		1.51, 51.0		0

Figure 4.22

EFL: 4.5

Image space NA: 0.55

Image height: 0.25

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	Infinity	4.000		7.44	0
2	92.6514	2.604	FK51	6.98	0
3	-12.7798	0.200		6.78	0
4	4.6185	2.638	LLF1	6.01	0
5	2.6263	2.000		3.97	0
6	-2.7122	3.904	SF5	3.92	0
7	17.4617	2.936	FK51	7.25	0
8	-5.6292	0.200		7.98	0
9	42.2154	0.700	KZFSN4	8.69	0
10	9.2080	2.842	FK54	8.91	0
STO	-14.8723	0.200		9.19	0
12	13.8695	1.958	FK54	9.69	0
13	-36.6177	0.200		9.66	0
14	10.7871	2.308	FK51	9.37	0
15	-25.2643	0.200		9.02	0
16	4.6295	2.150	SK11	7.02	0
17	6.2487	3.805		5.47	0
IMA	Infinity				0

Figure 4.24

EFL: 3.0

Image space NA: 0.55

Image height: 0.15

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	14.8237	1.832	SF3	4.75	0
2	-4.6312	1.398	LAK16A	4.54	0
3	4.3826	12.312		4.02	0
4	-59.4972	3.613	LAF3	9.03	0
5	14.0534	4.341	FK52	10.54	0
6	-8.2932	0.100		11.14	0
7	14.5105	0.750	SSK2	10.95	0
8	7.3198	6.000	CAF2	10.33	0
9	-7.6864	0.750	N-LAF7	10.04	0
STO	-120.2176	0.100		10.50	0
11	62.3822	4.505	CAF2	10.66	0
12	-9.9623	3.000	SF14	11.13	0
13	-11.2364	0.100		12.27	0
14	11.4747	2.780	FK52	11.25	0
15	48.9898	0.100		10.37	0
16	5.2242	2.367	SSK2	8.75	0
17	7.6262	5.251		7.39	0
IMA	Infinity				0

Figure 4.25

EFL: 3.3

Image space NA: 0.65

Image height: 0.21

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	-6.9715	1.500	N-SK14	6.33	0
2	5.9077	3.000	F5	7.40	0
3	-13.2583	17.009		7.73	0
4	-14.9986	3.000	KZFN1	9.50	0
STO	14.8832	5.500	N-FK56	10.96	0
6	-9.8899	0.100		12.13	0
7	17.9052	4.000	K10	12.35	0
8	-12.3198	1.500	LAKN14	12.08	0
9	14.5228	0.856		11.90	0
10	15.5522	5.000	N-FK56	12.46	0
11	-9.7393	1.000	F5	12.64	0
12	-29.3360	0.100		13.38	0
13	11.3343	6.000	N-PK52A	13.98	0
14	-11.7943	1.000	KZFN1	13.40	0
15	-230.4465	0.200		12.48	0
16	7.4350	3.000	N-PK52A	11.04	0
17	13.9209	0.200		9.64	0
18	3.9017	4.750	N-SK14	7.38	0
19	2.0156	0.850		2.19	0
20	Infinity	0.170	K4	0.64	0
IMA	Infinity				0

Figure 4.29

EFL: 200.0

Image space NA: 0.025

Image height: 12.5

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
STO	Infinity	10.00		10.00	0
2	Infinity	100.00	N-BK7	11.25	0
3	Infinity	10.00		19.51	0
4	-528.7454	5.00	N-SK10	20.74	0
5	-36.4749	2.00	N-LAF7	21.07	0
6	-89.4164	1.00		21.44	0
7	43.1762	5.00	BASF6	21.58	0
8	-84.6927	2.00	KZFSN4	21.05	0
9	37.8770	184.33		20.27	0
IMA	Infinity			25.00	0

Figure 4.32

EFL: 25.0

Image space NA: 0.1

Image height: 12.5

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
STO	Infinity	20.00		5.00	0
2	-193.3958	3.00	N-SF10	27.53	0
3	62.5837	13.00	N-PSK3	31.25	0
4	-28.8783	0.15		35.45	0
5	71.8090	10.00	N-FK5	40.19	0
6	-41.3252	0.15		40.44	0
7	28.1287	15.00	N-PSK3	36.09	0
8	-27.1268	3.00	N-SF2	34.25	0
9	27.4492	8.69		26.99	0
10	Infinity	300.00		25.82	0
11	Infinity	-300.00		60.75	0
IMA	Infinity			25.82	0

Figure 4.33

EFL: 25.0

Object space NA: 0.1

Object height: 10.0

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	-24.47			0
1	Infinity	-250.00			0
STO	Infinity	250.00			0
3	37.9447	10.00	N-SF4	22.91	0
4	-64.7142	4.70	N-LAF21	20.38	0
5	16.5587	32.96		17.74	0
6	-124.2415	10.00	N-SK16	38.64	0
7	-34.1714	1.00		41.47	0
8	352.6172	3.50	N-SF1	42.65	0
9	39.5760	10.00	N-SK16	42.91	0
10	-75.1006	1.00		42.97	0
11	63.6250	9.00	N-SK16	41.46	0
12	-235.4194	1.00		39.67	0
13	34.0613	15.00	N-SK16	34.62	0
14	-326.1752	2.50	N-SF5	25.41	0
15	27.9405	19.50		21.15	0
IMA	Infinity				0

Figure 4.36

EFL: 8.8

Image space NA: 0.9

Image height: 1.0

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	75.00		17.46	0
STO	Infinity	5.80		16.01	0
2	10.6009	7.80	N-K5	19.00	-1.524
3	-30.0928	0.10		18.40	0
4	7.5462	10.00	N-K5	13.71	0
5	Infinity	0.30		4.87	0
6	Infinity	1.00	N-K5	3.21	0
IMA	Infinity				0

Surface 2	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	1.74E-05
Coeff on r ⁶	3.74E-07
Coeff on r ⁸	-1.24E-08
Coeff on r ¹⁰	7.03E-11

Figure 4.39

EFL: 11.2

Image space NA: 0.75

Image height: 1.2

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	75.00		16.00	0
STO	Infinity	6.00		17.37	0
2	38.4762	5.00	N-K5	20.66	0
3	-44.7260	0.20		20.87	0
4	16.8020	5.00	N-K5	20.30	0
5	64.2822	0.50		19.01	0
6	9.5553	10.00	N-K5	16.06	0
7	Infinity	2.50		8.63	0
IMA	Infinity			2.84	0

Figure 5.32

EFL: 3.4

Image space NA: 0.65

Image height: 0.21

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
1	-7.4832	1.498	N-SK2	6.45	0
2	5.4547	2.995	F5	7.49	0
3	-11.4509	14.205		7.72	0
4	-11.0990	2.995	LLF1	8.00	0
STO	12.3744	5.491	N-FK56	9.32	0
6	-8.4374	0.100		10.62	0
7	88.0374	3.993	K10	10.68	0
8	-7.5915	1.498	N-LAK14	10.65	0
9	-175.7022	1.997		11.37	0
10	29.4565	4.992	N-FK56	12.18	0
11	-12.5403	1.000	F2	12.38	0
12	-43.7819	0.100		12.80	0
13	12.0474	5.990	N-PK52A	13.14	0
14	-10.2749	0.998	LLF1	12.49	0
15	54.6340	0.200		11.61	0
16	8.7534	2.995	N-PK51	11.08	0
17	78.5318	0.200		10.24	0
18	3.8770	4.742	N-SK2	7.35	0
19	2.2395	0.800		2.28	0
20	Infinity	0.170	N-BK7	0.80	0
21	Infinity	0.047		0.57	0
IMA	Infinity			0.43	0

Figure 5.34

EFL: 2.23

Image space NA: 0.9

Image height: 0.14

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
1	-26.5236	2.000	PK50	7.40	0
2	6.4903	3.000	SF13	7.44	0
3	56.9270	15.000		7.12	0
4	-2.9495	0.800	N-SF5	4.40	0
5	12.8597	3.990	CAF2	5.92	0
6	-4.1450	0.100		7.32	0
7	18.6087	4.500	CAF2	7.72	0
8	-5.1228	0.800	N-SF5	7.66	0
9	-13.1127	1.000		8.18	0
10	Infinity	12.000	N-BK7	8.23	0
11	Infinity	1.000		8.47	0
STO	16.7235	3.000	F SILICA	8.51	0
13	-13.5467	0.100		8.38	0
14	6.7161	2.000	F SILICA	7.55	0
15	18.4924	0.100		6.75	0
16	3.6170	2.200	F SILICA	5.72	0
17	3.6602	0.050		3.93	0
18	1.9068	2.000	F SILICA	3.33	0
19	1.0000	0.500	WATER	1.14	0
20	Infinity	0.170	N-BK7	0.54	0
IMA	Infinity		WATER	0.29	0

Figure 6.35

EFL: 4.14

Image space NA: 0.7

Image height: 0.34

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	Infinity	6.38		8.10	0
2	-6.2238	1.40	KZFS1	7.22	0
3	26.5023	4.40	LAFN7	8.45	0
4	-10.4202	12.23		9.50	0
STO	Infinity	7.54		7.74	0
6	19.8398	3.00	N-PK52	8.39	0
7	-7.4795	1.00	N-SF56	8.28	0
8	1272.2270	0.10		8.63	0
9	9.9276	4.00	N-PK52	8.95	0
10	-5.4899	1.00	KF6	8.84	0
11	-28.6849	0.10		8.56	0
12	4.9295	2.00	N-PK52	7.82	0
13	12.6423	0.10		7.25	0
14	3.7455	3.70	N-SF11	6.03	0
15	2.1055	0.45		1.91	0
16	Infinity	0.17	N-K5	1.35	0
17	Infinity	0.18		1.15	0
IMA	Infinity			0.69	0

Figure 7.6

EFL: 125.0

Image space NA: 0.02

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
STO	15.4009	3.00	N-BK10	5.00	0
2	-19.4441	2.00	N-SF2	4.71	0
3	-111.2265	20.91		4.55	0
4	3.7299	2.00	N-SF2	1.34	0
5	1.6461	22.09		0.87	0
IMA	Infinity				0

Figure 7.10(b)

EFL: 22.33

Image space NA: 0.063

Wavelengths: 0.65, 0.675, and 0.70

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	0.00		0.00	0
1	Infinity	2.34		0.00	0
2	11.3548	3.07	C0550	3.02	8.9793
STO	-2.7596	40.00		7.00	-0.5363
4	-11.8427	2.00	N-SF11	4.50	0
5	5.1085	4.00	N-LAK10	7.00	0
6	-8.7153	50.00		7.00	0
IMA	Infinity			4.00	0

Surface 2	VENASPH
Coeff on r ²	0
Coeff on r ⁴	-0.00907124
Coeff on r ⁶	0.001016054
Coeff on r ⁸	3.53E-05
Coeff on r ¹⁰	-1.32E-05
Surface STO	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	-0.00010444
Coeff on r ⁶	4.17E-05
Coeff on r ⁸	-1.64E-05
Coeff on r ¹⁰	3.11E-06

Figure 7.13(a)

EFL: 106.18

Image space NA: 0.076

Image height: 4.2

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	12.000		11.00	0
1	Infinity	19.800	N-SF10	11.78	0
2	Infinity	16.800		12.52	0
3	13.8390	8.860	LAK16A	13.74	0
4	-46.1097	4.400		10.82	0
5	-8.8558	9.440	SF18	6.49	0
6	15.3437	1.000		4.92	0
7	24.0748	5.830	N-FK5	4.90	0
STO	-10.7662	27.810		4.44	0
9	17.5968	4.530	LAK16A	12.00	0
10	77.2544	17.528		11.40	0
IMA	Infinity			8.42	0

Figure 7.13(b)

EFL: 252.0

Image space NA: 0.1

Image height: 4.2

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	12.000		11.00	0
1	Infinity	19.800	N-SF10	13.09	0
2	Infinity	9.000		15.07	0
3	40.3007	5.950	LAK16A	16.79	0
4	-62.5971	8.650		16.42	0
5	16.4001	5.000	N-FK5	12.76	0
6	Infinity	0.580		10.87	0
7	11.9244	5.820	N-FK5	9.62	0
8	-15.1241	5.080	SF18	5.92	0
STO	4.7357	3.820		2.53	0
10	-17.8456	4.020	KZFS8	5.81	0
11	143.1112	6.520	N-SK14	8.73	0
12	-9.0063	0.750		11.86	0
13	20.5140	5.820	LAK16A	12.74	0
14	218.4840	15.688		11.98	0
IMA	Infinity				0

Figure 7.18

EFL: 21913.16

Entrance pupil diameter: 5.0

Radial field: 5 degree

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
STO	Infinity	50.00		5.00	0
2	45.5501	9.00	N-BK7	13.84	0
3	-18.5883	2.50	N-SF2	13.91	0
4	-43.8418	50.09		14.15	0
5	Infinity	50.09		9.45	0
6	43.8418	2.50	N-SF2	14.15	0
7	18.5883	9.00	N-BK7	13.92	0
8	-45.5501	50.00		13.85	0
IMA	Infinity				0

Figure 7.21EFL: ∞

Entrance pupil diameter: 5.0

Radial field: 6 degree

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0	0
STO	Infinity	100.0		5.00	0
2	-200.0	-100.0	MIRROR	55.61	0
3	Infinity	100.0	MIRROR	21.26	0
4	-200.0	-100.0	MIRROR	55.88	0
IMA	Infinity				0

Figure 7.28

EFL: 26.0

Image space NA: 0.17

Image height: 4.5

Wavelengths: 0.64, 0.67, and 0.70

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
STO	Infinity	24.00		8.66	0
2	-9.9240	5.00	N-LAK10	15.76	0
3	-13.2945	2.20		20.39	0
4	104.8014	6.00	N-LAK10	23.83	0
5	-31.1938	1.00		24.28	0
6	46.7684	5.00	N-LAK10	22.60	0
7	-96.1701	1.00		21.37	0
8	-42.7277	4.00	N-SF6	21.12	0
9	16.8237	8.00	N-LAK10	19.19	0
10	-177.5866	4.80		18.32	0
11	23.3636	8.00	N-SF6	15.81	0
12	15.8110	8.92		11.68	0
IMA	Infinity			9.05	0

Figure 7.29

EFL: 4.75

Image space NA: 0.75

Image height: 0.35

Wavelengths: 0.78, 0.81, and 0.85

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
1	Infinity	10.00		9.90	0
2	-12.0406	2.00	F5	8.53	0
3	13.8219	6.00	K10	9.19	0
4	-20.5434	15.00		10.32	0
STO	25.4895	6.00	N-FK56	12.23	0
6	-10.0534	2.00	SF8	12.36	0
7	-38.2646	0.80		13.25	0
8	19.5988	6.00	N-FK56	13.73	0
9	-28.7257	0.80		13.34	0
10	8.6178	6.00	N-PK52A	11.91	0
11	31.5242	0.25		8.75	0
12	3.9739	4.24	N-SK11	6.73	0
13	2.7158	0.50		2.32	0
14	Infinity	0.70	N-BK7	1.78	0
15	Infinity	0.20	SEAWATER	0.98	0
IMA	Infinity		SEAWATER	0.71	0

Figure 7.30

EFL: 2.5

Image space NA: 1.0

Image height: 0.0375

Wavelengths: 0.35 and 0.633

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
1	Infinity	2.000		5.47	0
2	-7.9688	1.000	S-FPL51	5.43	0
3	11.2211	0.937		5.85	0
4	21.5986	2.444	PBL26	6.37	0
5	-8.9975	1.456		6.80	0
6	-16.0143	1.000	S-BSL7	6.75	0
7	13.1476	1.920		7.01	0
8	11.9273	1.588	S-FSL5	8.15	0
9	99.0168	31.583		8.23	0
10	-196.5839	5.000	S-FPL51	16.00	0
11	-20.3889	0.200		16.00	0
12	155.4282	4.000	CAF2	16.00	0
13	-16.0926	1.681		16.00	0
14	-11.8758	2.000	ADF10	12.73	0
STO	874.1270	0.200		16.00	0
16	18.6354	5.000	CAF2	16.00	0
17	-20.8025	2.128		16.00	0
18	-13.5250	2.000	ADF10	16.00	0
19	24.7238	0.969		12.98	0
20	28.3161	5.000	CAF2	16.00	0
21	-12.8711	0.200		16.00	0
22	14.8579	4.000	CAF2	16.00	0
23	148.8495	0.200		12.06	0
24	7.5575	4.923	CAF2	12.00	0
25	15.3298	0.200		7.62	0
26	2.8265	2.752	S-FSL5	5.46	0
27	2.9665	1.501	SEAWATER	3.12	0
28	Infinity	0.275	SEAWATER	0.71	0
IMA	Infinity		SEAWATER		0

Figure 7.33

EFL: 4.3

Image space NA: 0.34

Image height: 0.22

Wavelengths: 0.486, 0.546, and 0.587

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	-10.0000	-10.000		1.20	0
1	Infinity	10.000		2.62	0
2	-10.0000	0.545		1.20	0
3	3.1567	1.388	N-SF57	1.43	0
4	-2.8806	0.932		1.39	0
5	-1.1002	0.652	N-BK7	0.90	0
6	0.9629	0.539		0.89	0
7	-1.3718	0.733	N-LAF34	1.08	0
8	-2.0273	0.109		1.57	0
9	7.1562	0.543	N-SF57	1.77	0
10	2.9305	0.790	N-LAK10	1.90	0
11	-2.1504	0.130		2.00	0
12	2.0950	0.687	N-LAF7	1.85	0
13	-7.3523	0.143		1.61	0
14	-3.7338	0.572	N-SF57	1.47	0
15	1.0895	0.800	N-SSK2	1.20	0
STO	1.7165	0.837		1.05	0
17	-6.4488	0.722	N-LAF34	1.34	0
18	-2.5807	0.046		1.53	0
19	1.6391	0.744	N-LASF31A	1.56	0
20	7.6102	0.662		1.30	0
21	Infinity	0.687	N-BK7	0.83	0
22	Infinity	0.094		0.51	0
IMA	Infinity				0

Figure 7.39

EFL: 19.19

Image space NA: 0.32

Image height: 1.0

Wavelengths: 0.50, 0.55, and 0.60

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
STO	Infinity	15.000		11.99	0
2	12.8290	2.500	N-SF11	14.40	0
3	7.9873	6.500	N-BK7	14.40	0
4	-343.3173	6.104		14.40	0
5	-9.6544	5.000	N-SF11	11.00	0
6	-11.1996	1.500		13.00	0
7	11.0151	5.000	N-LAK10	12.00	0
8	Infinity	3.269		7.85	0
9	-11.8805	2.000	N-SF11	4.50	0
10	2.4477	3.000	N-LAK10	4.50	0
11	Infinity	0.500		2.32	0
IMA	Infinity			2.02	0

Figure 7.41

EFL: 16.69

Object space NA: 0.3

Image height: 0.15

Wavelengths: 0.85

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	0.000	1.48, 57.7	0.87	0
1	Infinity	2.474	1.48, 57.7	0.87	0
2	-1.2593	1.893	N-BK7	1.74	0
3	-1.9558	4.105		3.00	-0.3296041
STO	1.5998	1.785	N-BK7	3.00	-1.000891
5	-3.1591	0.141		3.00	0
6	0.9410	0.800	N-BK7	1.80	0
7	Infinity	0.499	SEAWATER	1.25	0
IMA	-2.5		SEAWATER		0

Surface 3	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	0
Coeff on r ⁶	0.00311376
Coeff on r ⁸	-0.00040804
Surface STO	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	0
Coeff on r ⁶	0.00218521
Coeff on r ⁸	-0.00114184
Surface 5	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	0.03438601
Coeff on r ⁶	-0.00645462
Surface IMA	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	-34.720681
Coeff on r ⁶	938.77489

Figure 7.42

EFL: 162.59

Object space NA: 0.3

Image height: 0.125

Wavelengths: 1.064

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	3.301	1.48, 57.7	0.82	0
1	-100.0000	1.400	E48R	2.19	0
2	-3.8360	6.273		2.66	0
3	-11.6200	1.400	E48R	4.13	0
4	-6.5490	0.152		4.48	-1
STO	3.1300	1.500	E48R	4.89	-1
6	100.0000	0.537		4.71	0
7	2.7930	1.400	E48R	4.09	0
8	16.6100	0.100		3.52	0
9	1.5160	1.300	E48R	2.53	0
10	-100.0000	0.450	SEAWATER	1.37	0
IMA	-1.4924		SEAWATER	0.25	0

Surface 4	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	4.62E-05
Coeff on r ⁶	2.91E-05
Coeff on r ⁸	-5.71E-06
Coeff on r ¹⁰	1.73E-06
Coeff on r ¹²	-2.26E-07
Surface STO	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	0.000247471
Coeff on r ⁶	-3.01E-05
Coeff on r ⁸	-1.60E-05
Coeff on r ¹⁰	2.07E-06
Coeff on r ¹²	-2.17E-07

Figure 7.43

EFL: 5.2

Object space NA: 0.3

Image height: 0.225

Wavelengths: 0.40, 0.587, and 0.70

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	0.300	F_SILICA	0.68	0
1	Infinity	0.266		2.00	0
2	-0.7562	1.271	SFL6	0.84	0
3	-1.4790	0.100		2.00	0
4	2.9190	2.841	SFL6	2.00	0
5	1.2367	1.500	N-PSK53	2.00	0
6	-6.2825	1.578		1.71	0
STO	Infinity	0.099		1.72	0
8	2.5623	1.800	SF6	2.00	0
9	1.4000	1.000	N-PSK53	2.00	0
10	-1.3000	1.000	SFL6	2.00	0
11	-3.6747	0.100		2.00	0
12	1.9144	0.400	SFL6	2.00	0
13	4.6111	0.096		1.58	0
14	1.6127	0.893	SFL6	1.60	0
15	1.8773	0.318		0.87	0
16	Infinity	0.150	N-BK7	1.40	0
17	Infinity	0.050	WATER	1.40	0
IMA	-4		WATER	0.45	0

Figure 8.8

EFL: 1050.45

Image space NA: 0.1

Image height: 1.0

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	2.00		2.00	0
1	Infinity	1.00	N-BK7	2.27	0
2	-11.8273	11.51		2.35	0
3	13.0651	1.00	N-SF1	2.70	0
4	4.0649	2.00	BAF51	2.70	0
5	-22.6698	4.99		2.70	0
STO	Infinity	4.99		2.68	0
7	22.6698	2.00	BAF51	2.70	0
8	-4.0649	1.00	N-SF1	2.70	0
9	-13.0651	11.51		2.70	0
10	11.8273	1.00	N-BK7	2.35	0
11	Infinity	2.00		2.27	0
IMA	Infinity				0

Figure 8.9

EFL: 3820.99

Image space NA: 0.1

Image height: 1.0

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	1.462		2.00	0
1	7.8496	19.904	N-BAK1	2.70	0
2	-3.6752	0.368	F1	2.70	0
3	-7.5710	0.767		2.70	0
STO	Infinity	0.767		2.62	0
5	7.5710	0.368	F1	2.70	0
6	3.6752	19.904	N-BAK1	2.70	0
7	-7.8496	1.462		2.70	0
8	Infinity	0.000		2.04	0
IMA	Infinity				0

Figure 8.10

EFL: 340.72

Image space NA: 0.1

Image height: 1.0

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	0.000		2.00	0
1	Infinity	1.511		2.00	0
2	7.6676	13.815	N-BAK1	2.70	0
3	-2.1080	6.174	F1	2.70	0
4	-7.5607	1.000		2.70	0
STO	Infinity	1.000		2.69	0
6	7.5607	6.174	F1	2.70	0
7	2.1080	13.815	N-BAK1	2.70	0
8	-7.6676	1.511		2.70	0
IMA	Infinity				0

Figure 8.12

EFL: 16.846

Image space NA: 0.1

Image height: 1.0

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	2.000		2.00	0
1	15.2775	15.498	N-BK7	2.70	0
2	-2.8440	1.260	N-SF5	2.70	0
3	-5.2606	0.100		2.70	0
4	4.7809	3.392	N-BK7	2.70	0
5	3.4223	0.250		2.70	0
STO	Infinity	0.250		1.91	0
7	-3.4223	3.392	N-BK7	1.91	0
8	-4.7809	0.100		2.70	0
9	5.2606	1.260	N-SF5	2.70	0
10	2.8440	15.498	N-BK7	2.70	0
11	-15.2775	2.000		2.70	0
IMA	Infinity				0

Figure 8.14

EFL: -84873.9

Image space NA: 0.1

Image height: 1.0

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	0.000		2.00	0
1	Infinity	1.000		2.00	0
2	6.9743	1.500	ACRYLIC	2.70	0
3	Infinity	18.000	POLYCARB	2.70	0
4	Infinity	1.500	ACRYLIC	2.70	0
5	-4.3792	0.500	POLYSTYR	2.70	0
6	-6.6450	0.500		2.70	0
STO	Infinity	0.500		2.69	0
8	6.6450	0.500	POLYSTYR	2.70	0
9	4.3792	1.500	ACRYLIC	2.70	0
10	Infinity	18.000	POLYCARB	2.70	0
11	Infinity	1.500	ACRYLIC	2.70	0
12	-6.9743	1.000		2.70	0
IMA	Infinity				0

Figure 8.18

EFL: 1.0

Image space NA: 0.1

Radial field: 47.5 degree

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	Infinity	0.160	N-BK7	0.65	0
2	2.2859	0.150		0.46	0
STO	Infinity	0.021		0.21	0
4	-3.2879	0.604	N-LASF31A	0.25	0
5	0.7043	1.141		0.70	0
IMA	Infinity			1.54	0

Figure 8.21

EFL: 0.5

Image space NA: 0.167

Radial field: 61.5 degree

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	Infinity	0.278	N-KZFS2	1.20	0
2	0.9815	0.329		0.69	0
STO	Infinity	0.046		0.20	0
4	-0.6804	0.510	N-LASF44	0.27	0
5	-0.5284	0.186		0.80	0
6	0.9091	1.371	N-LAK8	1.20	0
7	Infinity	0.000		1.20	0
IMA	Infinity				0

Figure 8.22

EFL: 1.0

Image space NA: 0.137

Radial field: 60.0 degree

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	10.000		31.40	0
1	Infinity	0.377	SAPPHIRE	2.11	0
2	0.8658	0.251		1.36	0
3	1.6495	0.735	S-NPH2	1.35	0
4	4.2853	0.715		1.02	0
STO	Infinity	0.000		0.46	0
6	13.7114	0.829	N-LAK8	0.46	0
7	-1.3491	0.050		0.96	0
8	2.0393	0.942	N-BK7	1.30	0
9	Infinity	1.180		1.30	0
IMA	Infinity				0

Figure 8.23

EFL: 1.0
 Image space NA: 0.137
 Radial field: 60.0 degree
 Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	10.000		35.53	0
1	-2.4703	0.297	N-BK7	0.78	0
2	0.6961	0.078		0.39	0
STO	6.7632	0.989	N-LAF34	0.31	0
4	-0.8756	0.089		1.13	0
5	6.9023	0.297	N-SF57	1.35	0
6	1.0546	1.197	N-LAK34	1.48	0
7	-1.7886	1.082		1.76	0
IMA	Infinity				0

Figure 8.24

EFL: 1.0
 Image space NA: 0.137
 Radial field: 60.0 degree
 Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity			0
1	Infinity	0.250	N-LASF31A	1.70	0
2	0.7608	0.800		1.12	0
3	1.8003	0.300	N-SF57	1.00	0
4	-4.8783	0.311		1.00	0
STO	Infinity	0.096		0.54	0
6	-9.1216	0.183	N-SF6	0.63	0
7	0.9646	0.500	N-BK7	1.10	0
8	-1.1814	0.050		1.10	0
9	1.3393	0.170	N-SF57	1.40	0
10	0.9866	0.500	N-BK7	1.40	0
11	-27.9947	1.653		1.34	0
IMA	Infinity				0

Figure 8.25

EFL: 1.0

Image space NA: 0.1

Radial field: 47.5 degree

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	Infinity	0.150	N-BK7	0.66	0
2	1.8030	0.200		0.46	0
STO	Infinity	0.000		0.21	0
4	12.1588	0.100	N-SF6	0.21	0
5	0.4736	0.250	N-LASF44	0.32	0
6	-0.7253	1.100		0.47	0
IMA	Infinity				0

Surface 2	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	3.9409151
Coeff on r ⁶	-36.306748
Coeff on r ⁸	440.13378

Figure 8.27

EFL: 1.0

Image space NA: 0.1

Radial field: 40.0 degree

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	Infinity		0.00	0
1	Infinity	0.340	N-PSK53	3.14	0
2	Infinity	0.068		2.84	0
3	11.2210	0.340	TIF3	2.60	0
4	-9.9493	0.119		2.40	0
5	Infinity	0.102	N-LAK8	1.88	0
6	1.3290	0.340		1.51	0
7	Infinity	2.278	N-SF10	1.44	0
STO	Infinity	1.479	N-SF10	1.40	0
9	Infinity	0.772	N-PSK53	1.44	0
10	-0.9978	0.238	F6	1.80	0
11	-2.3253	0.034		1.80	0
12	13.3618	0.265	LASFN31	1.80	0
13	-5.2543	0.034		1.80	0
14	2.9955	0.592	LASFN31	1.82	0
15	-3.2307	0.347	N-SF57	1.82	0
16	134.1479	0.026		1.82	0
17	-13.5734	0.445	N-SF2	1.82	0
18	1.0694	0.271		1.41	0
19	5.2122	0.258	N-SF4	1.82	0
20	1.2358	0.588	LASFN31	1.82	0
21	-3.2028	0.620		1.82	0
IMA	Infinity				0

Figure 8.30

EFL: 1.9

Image space NA: 0.1

Image height: 1.5

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	12.000		25.86	0
1	Infinity	0.400	LASFN31	3.00	0
2	1.8432	0.500		2.04	0
3	-30.8862	1.000	N-SF6	2.00	0
4	-2.8804	1.000		2.00	0
STO	Infinity	0.027		0.61	0
6	-1.7481	1.200	N-SK16	0.61	0
7	-1.4717	0.953		1.60	0
8	-103.3435	1.300	N-SK16	2.07	0
9	-1.3663	0.400	N-SF6	2.70	0
10	-2.5417	0.000		2.70	0
11	Infinity	5.338	N-KZFS2	2.71	0
12	Infinity	0.000		3.03	0
IMA	Infinity			3.03	0

Figure 8.31

EFL: 0.896

Image space NA: 0.07

Image height: 0.7

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	10.416		29.62	0
1	Infinity	0.235	N-LASF31A	1.70	0
2	0.6785	0.739		1.00	0
3	Infinity	0.235	N-LASF45	1.70	0
4	1.2194	0.793	N-FK5	1.70	0
5	-1.3403	0.869		1.70	0
6	2.0392	0.672	N-FK5	1.70	0
7	-1.8425	0.119		1.70	0
STO	Infinity	0.000		0.51	0
9	Infinity	0.202	N-LAK33A	0.80	0
10	1.3702	0.532		0.55	0
11	1.5677	0.585	N-F2	1.60	0
12	Infinity	0.067		1.60	0
13	3.5576	0.726	N-FK5	1.60	0
14	-1.0239	0.235	N-LAF33	1.60	0
15	Infinity	0.067		1.60	0
16	1.2027	0.780	LLF2	1.60	0
17	-1.1238	0.235	N-SF57	1.60	0
18	76.0416	0.603		1.60	0
19	Infinity	1.672	N-BK7	1.27	0
IMA	Infinity				0

Figure 8.32

EFL: 1.0

Image space NA: 0.123

Image height: 1.0

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	13.000		38.35	0
1	Infinity	0.150	N-BALF4	1.18	0
2	0.5255	0.170		0.78	0
3	13.0004	0.400	N-SF6	0.79	0
4	-1.2870	0.128		0.63	0
STO	Infinity	0.205		0.36	0
6	-1.2237	0.500	N-FK5	0.59	0
7	-0.6376	0.050		0.94	0
8	Infinity	0.600	N-BAK2	1.10	0
9	-0.6509	0.180	N-SF6	1.19	0
10	-1.3745	1.000		1.50	0
11	3.4334	0.450	N-LASF31A	2.20	0
12	Infinity	0.050		2.20	0
13	Infinity	0.550	N-BK7	2.08	0
14	Infinity	0.000		2.03	0
IMA	Infinity				0

Figure 8.49

EFL: 0.62
 Image space NA: 0.086
 Radial field: 55.0 degree
 Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	7.3085	0.000		9.16	0
1	7.3086	0.850	POLYCARB	9.16	0
2	6.3085	6.309		8.04	0
STO	Infinity	0.050		0.11	0
4	Infinity	0.300	N-PSK53	0.20	0
5	-0.3876	1.000		0.42	-106.8958
IMA	Infinity				0

Surface 5	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	-27.941611
Coeff on r ⁶	661.04678
Coeff on r ⁸	-7655.3358
Coeff on r ¹⁰	34103.246

Figure 8.50

EFL: 0.91
 Image space NA: 0.085
 Radial field: 55.0 degree
 Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	14.3299	0.000		10.41	0
1	14.3299	1.433	POLYCARB	10.41	0
2	10.5138	3.583		8.11	0
STO	Infinity	0.042		0.19	0
4	Infinity	0.528	N-PSK53	0.31	0
5	0.5708	1.090		0.78	0.4851432
IMA	Infinity			2.08	0

Surface 5	EVENASPH
Coeff on r ²	0
Coeff on r ⁴	1.2622975
Coeff on r ⁶	-2.3315426
Coeff on r ⁸	25.09402
Coeff on r ¹⁰	29.16119

Figure 8.53

EFL: 0.815

Image space NA: 0.086

Radial field: 55.0 degree

Wavelengths: 0.486, 0.587, and 0.656

Surf	Radius	Thickness	Glass	Diameter	Conic
OBJ	Infinity	10.000		41.69	0
1	5.9840	1.197	POLYSTYR	4.66	0
2	1.5000	1.500		2.20	0
STO	Infinity	0.033		0.19	0
4	Infinity	0.979	LAFN21	0.26	0
5	-1.3089	0.109		1.03	0
6	1.8463	0.675	LAFN21	1.34	0
7	Infinity	0.359		1.38	0
8	Infinity	0.435	N-BK7	1.44	0
9	Infinity	0.000		1.48	0
10	Infinity	0.120		1.48	0
IMA	Infinity			1.52	0

Index

- Abbe numbers, 56, 129, 225, 226
aberration theory, 13, 51, 61, 62
aberrations, 3, 103, 105, 113, 114, 115, 119, 121, 122, 128–133, 138, 141, 145, 147, 148, 150, 154, 161, 163, 207, 219, 225, 226, 227, 349, 375, 385, 387, 390, 391, 396, 401–410, 412, 414, 416, 417, 427, 433, 435, 436, 437
absorption coefficient, 2
absorption filters, 182
absorption, 1–10, 233, 243, 247, 249, 250, 252, 257, 258, 273
absorptive polarizers, 255, 258
acceptance angle, 68–72, 75, 82, 86, 98, 108, 109, 111, 152, 160, 255, 256, 257, 272, 307
achromat, 110, 113, 120, 129, 130, 131, 135
achromatization, 389
acousto-optical deflectors, 324
acousto-optical tunable filters, 182
air-guiding fibers, 68
Airy disk, 322, 323, 330, 373, 398
analyzer, 246, 247, 248, 250, 251, 252, 253, 280
angular magnification, 13
anisotropy, 4, 7, 9
antireflective (AR), 219, 220, 258, 268, 273, 281, 283, 284, 285, 293, 294, 297, 299, 300, 388, 439
aperture stop, 16, 17, 32, 36, 58, 59, 65
aplanatic conditions, 123, 125
aplanatic lens, 57
aplanatic surface, 56, 57
apodization, 274, 286
aspherical surfaces, 57, 58
astigmatism, 29, 34, 35, 36, 37, 44, 46, 48, 51, 56, 58, 113, 119, 121, 122, 126, 129, 130, 133, 134, 140, 141, 142, 144, 145, 148, 150, 151, 328, 344, 345, 346, 349
atom, 3, 7
autofluorescence, 8, 165, 169, 170, 171, 177, 183, 185, 186, 190–193, 201, 202, 203, 204, 212, 218, 221, 222, 224, 225, 226, 228, 229, 232
avalanche photodiodes (APDs), 174, 178
average passband transmission, 180, 188
axial chromatic aberration, 129, 138
axial resolution, 111

Babinet compensator, 270
back focal length (BFL), 14, 138, 399
back focal plane, 104, 120
bandpass filter, 171, 180, 181, 182, 185, 187, 191, 192, 199, 200, 216
bandwidth, 176, 178, 180, 184, 187, 188, 190, 192, 193, 194, 199
barrel distortion, 38
base curvature, 57
beam deviation, 255, 268, 270
beam scan, 324, 326, 328, 330, 331
beam splitter, 105

- beamsplitting polarizers, 255, 258
 Beer's law, 3, 4
 bi-crystalline achromatic retarder, 269
 biocompatibility, 170
 bioluminescence, 171, 197
 biomarkers, 1
 birefringence, 7, 268, 273, 288, 289, 306
 blur, 3
 Brewster's angle, 258, 275, 276, 294
 bright field, 117, 152, 153, 160
 brightness, 80, 86, 89, 98, 116, 117, 154–158
 buffer, 68, 69

 camera equation, 25
 capsule endoscopy, 430, 442
 cardinal points, 13, 14, 15
 catadioptric lenses, 145
 center wavelength, 180, 188
 central obscuration, 362, 363, 364
 charge-coupled device (CCD), 47, 48, 174, 178, 179, 180, 204, 384, 402, 432
 chief ray, 16, 26, 32, 34, 36, 40, 41, 47, 48, 58, 119, 150, 153, 154, 156, 385, 386, 387, 399, 402, 408, 411
 chromatic aberration, 26, 39, 40, 43, 44, 51, 56, 58, 59, 60, 113, 114, 119, 120, 122, 127, 129, 131, 133–138, 140, 145, 146, 147, 148, 150, 154, 156, 157, 220, 225, 228
 circular polarization, 245
 circular polarizer, 263
 cladding modes, 370
 cladding, 68–73, 76, 78, 81, 82, 84, 86, 92, 93, 210, 211, 212
 clear aperture, 188, 189
 close coupled, 339
 close-up mode, 412, 413, 414, 415
 coherence gating, 10
 cold mirror, 87

 collector, 105, 117, 153, 156, 157, 158
 color centers, 220, 221
 color fidelity, 118
 color temperature, 86
 colored filter glass, 183
 coma, 29, 32, 33, 34, 43, 44, 45, 48, 51, 56, 58, 59, 60, 121, 328, 335, 336, 343, 349, 350, 387, 391, 394, 396, 400, 403, 404, 405, 408, 409, 410, 412, 435
 complementary metal-oxide semiconductor (CMOS), 47, 48, 174, 178, 180, 197, 204, 230, 231, 384, 432
 compound parabolic concentrator (CPC), 90, 91
 compound zero-order retarder, 267
 condenser, 105, 110, 111, 112, 117, 119, 153–158, 160, 161, 162
 confocal aperture, 11
 confocal fluorescence imaging, 211
 confocality, 322
 conic constant, 57
 conjugate distance, 105
 contrast agents, 1, 2
 contrast ratio, 255, 256, 258, 261, 263, 270, 271, 272
 contrast, 1–11, 165, 172, 174, 191–196, 199, 200, 201, 202, 204, 211, 212, 218, 224, 229
 Cooke triplet, 62
 copolarization, 244, 246, 301
 core-cladding boundary, 68
 correction collar, 114
 cosine-fourth falloff law, 23, 24
 coupling lens, 358, 368, 369
 cover glass, 108, 109, 114, 118, 121, 125, 132, 134, 139, 140, 141, 142
 critical angle, 276, 277
 cross-polarization, 244
 cut-off wavelength, 180–183, 185, 199
 cypate-mono-2-deoxy-glucose, 170
 cystoscope, 379

- cytoplasm, 2
- dark field, 117, 152, 153, 160, 162
- defocus, 28–32, 43
- degree of circular polarization (DOP_C), 241
- degree of linear polarization (DOP_L), 239, 241
- degree of polarization (DOP), 239
- deoxyribonucleic acid (DNA), 3, 168, 171, 197, 231
- depolarization length, 244
- depolarization, 7
- depolarizers, 253
- depth of field, 19, 20
- diagnostic window, 3
- diattenuation, 7, 243, 271, 273–277, 281, 282, 283, 286, 287, 288, 293, 295, 297, 298, 299, 300
- dichroic beamsplitter, 172, 174, 186, 180, 181, 182, 190, 191, 193, 195, 198–200, 202, 208, 227
- dichroic mirrors, 108, 181, 214, 215
- dichroic polarizers, 256
- diffraction, 110, 112, 119, 120, 128, 129, 131, 132, 134, 136, 138, 140, 141, 145, 146, 156, 160
- diffractive optical element (DOE), 205
- diffuser, 154, 155, 156
- diffusion theory, 5
- digital micromirror devices (DMDs), 175
- distortion 29, 38, 51, 58, 59, 60, 118, 122, 148, 327, 328, 335, 336, 343, 348, 349, 351, 354, 356, 357, 360, 368, 371, 387, 390–410, 412, 416, 417, 433, 435, 436, 437, 443
- dome, 431–442
- double Gauss, 62
- double telecentric, 17, 18
- duty cycle, 337, 339, 342
- dynamic range, 118
- edge filters, 181
- effective focal length (EFL), 138, 399
- effective refractive index, 186
- eigenpolarizations, 266
- electromagnetic actuator, 209
- electron-multiplying charge-coupled device (EMCCD), 179
- ellipticity angles, 239
- ellipticity, 234, 235, 236, 239, 261, 267
- emission filters, 177, 195, 200, 207
- emission spectrum, 8, 167, 169, 178, 192, 193
- emission, 8, 117, 156, 165, 166, 167, 168, 169, 170, 172, 173, 174, 175, 176, 177, 178, 180, 181, 183, 190–202, 204–209, 211, 212, 213, 215–219, 222, 223, 225
- enamel, 2, 4
- encircled energy, 43, 47, 64
- endogenous, 169
- endoscopy, 67
- ensquared energy, 47
- entrance pupil, 16, 17, 44
- entrance window, 17
- epifluorescence, 313
- epi-illumination, 105, 117, 119, 152, 153, 158, 162
- étendue, 21, 23
- excitation filters, 213
- excitation, 3, 8, 11, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 180, 181, 190–214, 216–219, 222, 223, 224, 225, 226, 227, 311, 312, 315, 319, 321, 329, 351, 352, 355, 364, 373
- excitation-emission matrices (EEM), 84, 212, 213
- exit pupil, 16, 17, 19, 28, 32, 48, 103, 104, 115, 116, 120, 145, 146, 148, 149, 150, 151
- exit window, 17
- exitance, 22
- exogenous fluorophores, 165, 169, 170, 199

- eye relief, 115, 149, 151
 eyepiece matching distance, 108
 eyepiece, 103, 105, 108, 109, 110, 113, 115, 116, 122, 131, 147–151, 163, 379, 380, 381, 383, 384, 385, 387, 388, 398, 419
- F*-number (*f*/#), 51, 55, 61
- faceplates, 81
- fiber bundles, 67, 68, 79, 81–87, 91–93, 95, 98, 101, 379, 381, 382, 385, 386, 400, 401, 403, 410, 420, 421, 422, 424–430
- fiber light guide, 79, 82, 83, 175, 219
- fiber tips, 77, 80
- field curvature, 29, 36, 37, 43, 48, 49, 51, 58, 61, 65, 113, 119–121, 125–131, 133–137, 145, 147, 148, 151, 154, 207, 208, 228, 327, 328, 332, 335, 336, 343, 344, 346, 348, 349, 351, 353, 356–371, 390–397, 399, 402, 403, 404, 409, 410, 414–416, 433, 436
- field lens, 61
- field of view (FOV), 14, 17, 24, 32, 37, 40, 49–52, 61, 62, 63, 113, 118–122, 128, 129, 130, 133, 134, 138, 141, 145, 146, 149, 151, 153, 160, 175, 196, 203, 204, 207, 217, 218, 219, 225, 228, 246, 300, 312, 329, 335, 337, 340, 345, 346, 353, 355, 356, 357, 360, 363, 364, 365, 368, 370–373, 382, 383, 391, 395, 396, 398, 402–421, 427, 431, 432, 433, 435, 437–441
- field stop, 17, 153, 154
- finite-corrected, 103, 104
- flavin adenine dinucleotide (FAD), 167, 169
- fluence rate, 78
- fluorescein isothiocyanate (FITC), 169, 176
- fluorescence decay time, 173
- fluorescence filters, 174, 177, 180
- fluorescence imaging, 165, 167, 169–180, 182, 183, 186, 190–193, 196–221, 229, 231
- fluorescence lifetime imaging, 173
- fluorescence lifetime, 8
- fluorescence process, 165, 166
- fluorescence, 1, 7, 8, 10, 11, 165–182, 183, 186, 189–226, 229–232
- fluorescent dyes, 166
- fluorescent light, 171
- fluorescent probes, 166
- fluorochromes, 166, 172, 179
- fluorophores, 8, 167–170, 174, 175, 176, 177, 191, 192, 194, 200, 201, 213, 217
- focal planes, 14
- focal points, 13, 14, 15
- focusing lens, 381, 398
- Fourier transform spectroscopic imaging, 173
- frequency domain, 173, 230
- frequency-domain imaging, 10, 11
- Fresnel loss, 82
- Fresnel number, 323
- Fresnel reflection, 82, 410, 430, 437, 442
- Fresnel rhomb, 269, 272
- front focal length (FFL), 14
- full width at half maximum (FWHM), 9, 180, 192, 245
- full-wave retarder, 266
- galvanometer scanners, 324
- galvanometer, 324–326, 339, 356
- gastroscope, 379
- Gaussian equations, 15
- Gaussian optics, 13
- geometrical extent. *See* throughput
- geometrical optics, 13
- graded-index fiber, 71
- graticule, 115
- green fluorescent protein (GFP), 167, 169–171
- GRIN lens, 385, 389, 397, 410, 411

- hard coating, 186
hard tissues, 2
hard-coated filters, 186
helicity, 245
hemoglobin, 3, 247, 252
heterogeneity, 9
hexagonal packing, 81
higher-order aberrations, 43
hollow waveguides, 67, 101
holographic filters, 182
Hopkins rod-lens relay, 391, 393
hot mirror, 86, 87
hybrid imaging, 216
- illumination, 67, 68, 77, 79, 80, 82, 84, 86, 87, 89, 91, 92
image formation, 13
image quality, 13, 17, 29, 31, 43, 47, 48, 49, 61, 65
imaging fiber bundles, 79, 82
immersion liquid, 109, 125, 132, 139, 140
incoherent fiber bundle, 79
index-guiding fibers, 68
indocyanine dye Cy5.5, 170
indocyanine green dye (ICG), 170
infinity-corrected, 104, 105, 106, 108, 110, 114, 115, 122, 145
infrared (IR), 3, 170, 175, 178, 182, 420
intensified CCD (ICCD), 179
intermediate image plane, 103, 104, 105, 108, 114, 145, 146, 149, 151
intermediate images, 381
intrinsic absorption, 72
intrinsic birefringence, 288, 291
inverse-square law, 23, 24
irradiance, 9, 10, 22–25
- Jablonski energy diagram, 165, 166
Jones calculus, 236, 237
Jones matrix, 236, 237, 303
Jones vector, 236, 243
- kidney bean effect, 149
- Köhler illumination, 105, 153, 156–159, 191, 204, 218
- Lagrange invariant, 20, 21, 23
Lambert's cosine law, 23
Lambertian source, 23
laser damage threshold, 255, 270
lateral chromatic aberration, 122, 131, 134, 140, 146, 147, 327, 328, 343, 349–352, 356, 357, 370, 373, 390, 391, 394, 396, 398–406, 408, 409, 417, 435, 436
lateral magnification, 15
lateral resolution, 110, 111
lens bending, 51, 54, 122
lens combination, 55
lens compounding, 55
lens design, 13, 51, 61, 65
lenslet array, 204, 205, 219
light absorber, 200
light collection efficiency, 108, 118
light gathering power, 108
linear birefringence, 243
liquid crystal devices (LCDs), 175
liquid crystal tunable filters, 182
liquid light guides, 67, 87
Lister objective lens, 129, 130
longitudinal aberration, 26, 27, 35, 37
longitudinal magnification, 15, 16
longpass filter, 171, 180, 181, 183, 192, 199
- macrofluorescence, 203
macroimaging, 196, 203, 205, 219
magnetic resonance imaging (MRI), 1
magnification, 104, 109–113, 115, 119, 120, 121, 128, 129, 131, 133, 135, 136, 140, 149, 162, 320, 323, 330, 343, 351, 353, 367, 370
Maréchal criterion, 51
marginal focus, 31
marginal ray bundle, 136
marginal ray, 16, 20, 31, 40, 41, 61

- mechanical tube length, 107
melanin, 2, 3
meridional rays, 70
merit functions, 62, 64
metal halide, 175
micro-electromechanical system (MEMS), 209, 210, 324, 326, 339, 359, 360, 362, 363, 365, 376
microimaging, 196, 199
Mie scattering, 3
mitochondria, 2
modal dispersion, 71
mode mixing, 72
mode stripping, 72
modulation transfer function (MTF), 43, 48, 49, 62, 64, 323, 324, 343, 345, 346, 347, 364, 370, 413, 416
molecular imaging, 1
molecule, 3, 6, 7, 167, 170
monochromatic aberration, 26, 31
Mueller matrix, 241, 253, 303, 305, 306
multicolor, 172, 213
multimode fibers, 70, 71, 78, 92, 209
multiphoton imaging, 11
multiple-order retarder, 267, 268
- natural birefringence, 288
Newtonian equations, 15
nicotinamide adenine dinucleotide (NADH), 8, 167
Nipkow disk, 330, 332–337, 375
noncoherent light sources, 174
noninvasive, 2
normalized frequency, 70, 71
nuclei, 2
Nyquist frequency, 48
- object distance, 123, 124
objective lens, 17, 47–52, 62, 105–110, 114, 119–123, 126–130, 132, 134, 135, 136, 139, 140, 145, 153, 192, 199, 201, 202, 206, 218–220, 225–228, 232
- oblique aberrations, 58
odd aberrations, 32, 59, 387
off-axis aberration, 31
oil immersion, 124, 125, 132, 134
optical activity, 243
optical activity, 7
optical coherence tomography (OCT), 17, 216, 217, 231
optical contrasts, 1
optical fibers, 67, 70
optical path difference (OPD), 19, 289
optical properties, 1, 2, 12
optical sectioning, 313, 321, 322
optical sections, 311, 313
optical spectroscopy, 67
optical tube length, 103, 107, 108
organelles, 2
orientation, 233–236, 239, 247, 264, 291, 295
orientational birefringence, 289
orientational coefficient, 289
- packing fraction, 81, 93
paddle scanner, 340, 341
Pancharatnam achromatic retarder, 269
parameters, 70, 86, 105, 108, 115, 116, 118, 123, 139, 141, 144, 147, 149, 160
paraxial approximation, 25
paraxial focus, 31
paraxial image plane, 26–31, 38, 40, 45
parfocal distance, 105, 106
patterned polarizers, 264
penetration depth, 5, 249
Petzval curvature, 36, 37, 61, 113, 126, 129, 135
Petzval curvature
Petzval sum, 61, 394, 415
Petzval surface, 36, 37
phase contrast, 103, 114
photobleaching, 166, 168, 192, 207
photoelastic coefficient, 291

- photomultiplier tube (PMT), 174, 178, 191, 196, 205, 209, 211
- photonic crystal fibers (PCFs), 67
- photonic crystals, 67
- photons, 3, 6, 7, 8, 11, 173, 178, 179, 183, 250
- phototoxicity, 207
- phycoerythrin (PE), 169
- piezoelectric modulators (PEMs), 270
- piezoelectric driven actuator, 209
- pincushion distortion, 38
- pinhole lens, 322, 323, 324, 330
- pinhole, 188, 189, 191, 311, 313–318, 321–324, 330, 333, 335, 337, 339, 349, 351, 358, 359, 361, 366, 367
- pivot point, 316, 337, 339, 340, 342, 360, 361, 365
- plan apochromat, 49–52, 110, 120, 135, 137, 148
- plan apochromatic
- plan fluorite, 110, 120, 133
- plate polarizer, 258
- Pockels cells, 270
- Poincaré sphere, 236
- point spread function (PSF), 47, 49, 51
- polarization aberrations, 273, 288, 293, 295, 299, 300
- polarization beamsplitter, 246, 258, 259, 261, 263, 288
- polarization converter, 265
- polarization ellipse, 236, 239, 295
- polarization gating, 7
- polarization ratio image, 248
- polarization splitting, 187
- polarization state, 5
- polarization, 1, 5–11, 186, 187, 216, 233–261, 263, 264, 265, 266, 267, 270, 271, 272, 273, 274, 276, 279, 280, 281, 283, 286, 287, 288, 291, 292, 293, 294, 295, 296, 297, 299–308
- polarizers, 236, 239, 241, 253–259, 261–265, 271, 272, 280, 281, 288, 295, 430, 432
- polygonal scanners, 324, 325
- porphyrins, 167, 169, 228
- power density, 78
- principal planes, 14, 15
- principal points, 13, 14, 51
- principal rays, 14
- prismatic scanner, 325
- projected area, 21–23
- projected, 22
- protected-graft copolymer (PGC), 170
- protein, 3
- pupil spherical aberration, 148
- pushbroom scanning, 173
- pyramidal error, 325
- pyramidal scanner, 325
- quantum efficiency (QE), 118, 168
- quarter-wave retarder, 263
- radiance, 23, 25
- radiant intensity, 23
- Raman scattering, 1, 3, 6
- ray aberration, 26–30, 32, 35, 37, 38, 43
- Rayleigh criterion, 110
- Rayleigh scattering, 3
- red fluorescence protein (RFP), 170
- reflectance, 10, 11, 12, 22, 171, 172, 184, 201, 216, 217, 229, 230
- reflection, 1, 2, 5, 11
- refraction, 125, 160
- refractive index, 2, 3, 5, 8, 9, 108, 109, 113, 123–127, 132, 133, 139, 140, 383, 387, 388, 391, 394, 403
- relay lens, 316, 317, 325, 335, 336, 340, 342–346, 357, 368
- resolving power, 108, 110, 111, 160
- resonant scanner, 325
- retardance, 266–277, 281–287, 293, 295, 297–300

- retarders, 236, 241, 253–270, 272, 273, 295
- rms spot radius, 44, 49, 62, 390, 391
- rms wavefront error, 48, 62
- rod lenses, 379, 380, 392, 394, 396, 398
- rotation matrix, 243
- rotators, 253
- sagittal coma, 34
- scatterers, 233
- scattering coefficient, 2, 5, 9
- scattering function, 4
- scattering, 1–11, 160, 170, 173, 185, 186, 196, 201, 216, 233, 243–245, 247–253, 258, 270, 304–306
- Schwarzschild objective lens, 141, 144, 145, 146, 300–303
- scratches and digs, 188
- second harmonic generation (SHG) imaging, 11
- secondary chromatic aberration, 40
- Seidel aberrations, 40
- sheet polarizer, 255–257
- shortpass filter, 180–182, 192
- side view, 360
- signal-to-noise ratio (SNR), 93, 312, 322
- single-mode fiber
- single-mode fibers, 70, 71, 92, 209
- skew rays, 70
- slope, 181, 185, 192
- Snell's law, 25, 53
- soft coatings, 184
- solarization, 220
- solid angle, 4, 21, 22
- solid-core fibers, 67, 68
- solid-state detector, 118
- spatial filtering, 10
- spatial light modulators, 175
- spectra, 1, 8, 11, 167, 168, 170, 173, 176, 193, 194, 199, 222, 223
- spectral imaging, 11
- spectral polarization difference imaging (SPDI), 249–251
- specular reflection, 246–250, 253
- spherical aberration, 3, 9, 29–32, 43, 44, 46, 51, 53, 55–59, 113, 119–122, 125, 129–136, 138, 140–145, 148, 149, 151, 154, 157, 313, 320, 328, 335, 349, 350, 360, 363, 365, 371
- spot diagram, 43, 44, 45, 46
- square packing, 81
- stage scan, 324, 328, 329, 330, 352
- step-index fiber, 68, 71, 73
- Stokes shift, 167, 193, 199
- Stokes vector, 238, 239, 241, 243
- stop shift, 58, 59
- straight view, 360
- Strehl ratio, 43, 49, 51, 52, 64, 119, 128, 130, 131, 140, 145
- stress-induced birefringence, 270, 273, 289, 291–294
- stress-optical coefficient, 289, 293
- surface flatness, 188, 189
- tandem-scanning reflected-light microscope (TSRLM), 332, 333
- tangential coma, 34
- Telan system, 104, 108
- telecentric, 17, 18, 60, 103, 119, 129, 153, 384, 385, 389, 401–404, 413, 415
- telecentricity, 17, 119, 131, 153, 384, 385, 386, 398, 400
- therapeutic window, 3, 5, 6
- thin-film filters, 182, 184, 186, 188, 203
- third-order aberrations, 29, 51, 64
- through-focus, 44, 46
- throughput, 19, 21, 23
- time domain, 173
- time gating, 10
- total internal reflection (TIR), 67, 68, 73, 78, 82, 276, 283
- transillumination, 10, 117, 119, 152, 153, 158, 162, 247
- transmission, light, 1

- transmitted wavefront distortion, 188
- transverse ray aberrations, 27
- true zero-order retarder, 267
- tube lens, 104, 105, 108, 110, 114, 115, 118, 119, 121, 122, 131, 145, 146, 147
- ultrasound imaging, 1, 11
- ultraviolet (UV), 3, 165, 168, 175, 176, 182, 183, 188, 192, 201, 216, 220, 221, 222, 225, 226, 228, 420, 421
- unit magnification, 14, 60, 381
- unpolarized light, 236
- variable retarders, 270
- video endoscope, 379, 386, 387, 419
- viewing window, 432, 433, 436
- vignetting, 17, 119, 149, 389, 390, 391, 394, 395, 415
- wave plate, 265, 266, 268
- wavefront aberration, 26–29, 31, 40, 42
- wavefront deformation, 9
- wavefront quality, 255, 257, 268, 270, 271
- wedge error, 186
- wedge, 185, 186, 188, 189, 191
- wide-angle mode, 411–414
- wire-grid polarizer, 261–263, 272, 307
- wireless endoscope, 379, 430, 437, 438
- working distance, 14, 25, 62, 106, 108, 113, 114, 121, 128–130, 135, 136, 138, 139, 141
- x-ray imaging, 1
- zero-order retarder, 267, 268



Rongguang (Ron) Liang is currently a Senior Principal Research Scientist at Carestream Health, Inc. (formerly Health Group of Eastman Kodak Company), where he develops and designs biomedical optical imaging systems. Previously, he was a Principle Research Scientist at Eastman Kodak Company, where he developed imaging technologies for medical/biomedical systems, display, digital imaging, and printing. Dr. Liang currently serves as a topical editor for *Applied Optics*. He received his Ph.D. in Optical Sciences from the College of Optical Sciences, University of Arizona.