

Fundamentals of Micro-optics

Technology, Devices and Applications

A comprehensive overview of the micro-optics field, from the relevant fundamental physics to advanced applications, the book covers the essentials of optics, optical materials, micro-optical devices and systems. Micromirrors, refractive microlenses, diffractive optics, waveguide optics, and active micro-optical components are covered in detail, as are fabrication and assembly technologies and advanced topics in micro- and nano-optics. Designed for use in the classroom, the text includes a large number of worked examples, a spectrum of problems for each topic and an extensive bibliography.

Hans Zappe is Professor of Micro-optics in the Department of Microsystems Engineering at the University of Freiburg, Germany. Born in France and raised in New York, he earned his Bachelor's and Master's degrees from MIT and his PhD from the University of California, Berkeley, all in Electrical Engineering. After pursuing research activities in electronics, integrated optics and semiconductor lasers at IBM, the Fraunhofer Institute for Applied Solid State Physics, the Paul Scherrer Institute and the Centre Suisse d'Electronique et de Microtechnique, he joined the University of Freiburg in 2000. His present research interests focus on tunable micro-optics, optical microsystems for medical applications and novel nano-optics, at least for those moments when he isn't distracted by Pierre Hermé apricot-pistachio macarons.

Fundamentals of Micro-optics

Technology, devices and applications

Hans Zappe

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521XXXXXX

© Cambridge University Press 2010

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2010

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

ISBN-13 978-0-521-XXXXX-X hardback

ISBN-10 0-521-XXXXX-X hardback

**On ne peut me connaître
Mieux que tu me connais...**

Contents

<i>Preface</i>	<i>page</i>	xi
<i>Acknowledgements</i>		xiii
<i>Notation</i>		xv
Part I Essential optics		1
1 Introduction		3
1.1 Optics: history in a nutshell		4
1.2 Micro-optics: a smaller nutshell		10
1.3 Micro-optics for the home: the DMD		14
2 The physics of light		17
2.1 Basic electromagnetics		17
2.2 Wave propagation		25
2.3 Electromagnetic waves		30
2.4 Polarization		36
2.5 Wavefronts		40
2.6 Gaussian beams		46
3 Optical materials		55
3.1 Refractive index		56
3.2 Dispersion		68
3.3 Attenuation		77
3.4 Glass		83
3.5 Semiconductors		87
3.6 Other materials		93
4 Optical interfaces		101
4.1 Reflection and refraction		101
4.2 Reflected and transmitted fields		106
4.3 Power transmission and reflection		112
4.4 Internal reflection		115
4.5 Evanescent field		118

5	Interferometry	125
	5.1 Coherence and interference	125
	5.2 Interferometers	136
	5.3 Optical multilayers	150
	Part II Micro-optics	167
6	Reflective micro-optics	169
	6.1 Reflection	169
	6.2 Planar mirrors	172
	6.3 Non-planar mirrors	175
	6.4 Micromirrors	180
	6.5 Adaptive micro-optics	189
	6.6 Micromirrors: case studies	193
7	Refractive micro-optics	203
	7.1 Lens fundamentals	204
	7.2 Imaging	212
	7.3 Advanced lenses	225
	7.4 Primary aberrations	235
	7.5 Chromatic aberrations	243
	7.6 Microlenses: case studies	247
8	Diffraction micro-optics	265
	8.1 Diffraction	266
	8.2 Practical apertures	270
	8.3 Gratings	283
	8.4 Diffraction microlenses	300
	8.5 Diffraction micro-optics: case studies	311
9	Guided-wave micro-optics	321
	9.1 Waveguides: ray optic model	322
	9.2 Waveguides: electromagnetic model	337
	9.3 Channel waveguides	350
	9.4 Waveguide characterization	355
	9.5 Waveguide components	364
	9.6 Optical fibers	369
	9.7 Waveguide micro-optics: case studies	380
10	Active micro-optics	395
	10.1 Physics of light emission	395
	10.2 Light-emitting diodes	405
	10.3 Laser diodes	413

10.4	Photodetectors	430
10.5	Phase and intensity modulators	439
10.6	Other active components	448
10.7	Active micro-optics: case studies	453
11	Micro-optical fabrication	465
11.1	Basic semiconductor processing	466
11.2	Self-assembly lenses	468
11.3	Lithography for microlenses	474
11.4	Replication	482
11.5	Specialized processes	487
11.6	Assembly	491
Part III	Neoteric optics	501
12	Tunable micro-optics	503
12.1	Liquid microlenses	504
12.2	Membrane microlenses	510
12.3	LCD-based tunable micro-optics	517
12.4	Tunable diffractive micro-optics	518
12.5	A tunable micromanagerie	522
13	Fluidic micro-optics	525
13.1	Optofluidics	526
13.2	Fluidic waveguides	526
13.3	Optofluidic lasers	527
13.4	Fluidic microlenses	531
13.5	Filters and displays	533
13.6	Other optofluidic concepts	534
14	Nano-optics	537
14.1	Nanophotonics	537
14.2	Photonic crystals	539
14.3	Plasmonics	554
14.4	Metamaterials	558
	<i>References</i>	561
	<i>Index</i>	609

Preface

*Certains hommes parlent durant leur sommeil. Il n'y a guère que les conférenciers pour parler pendant le sommeil des autres.*¹

Alfred Capus

Microsystems have evolved from laboratory curiosities to pervasive constituents of a colorful spectrum of technical products. We can shake our MP3 players to change the song order; print on square meters of paper at 600 dot per inch resolution; have our doctors perform on-the-spot diagnostics; interact with computer games through our body motions; or have our lives saved by an inflated airbag: all these are possible because a microsystem is embedded somewhere, usually as invisible as it is indispensable.

The micro-electro-mechanical systems which first emerged from university laboratories in the 1980s combined mechanics with the technologies of electronics, leading to the moniker MEMS. But the technology has expanded enormously since then, so that the microelectronics and micromechanics of the germinal MEMS field now includes microfluidics, microacoustics, micromagnetics, microchemistry, microbiology and, not least, micro-optics. The integration of these disparate disciplines into highly functional microsystems with myriad applications is a key reason for the explosive development of these technologies, and as a result, the essence of much microsystems research is now often highly interdisciplinary.

Of these diverse disciplines, optical microsystems have seen particularly strong development, primarily due to the insatiable demand for communications bandwidth. Long and medium range telecommunications networks are almost completely optical, and it is the availability of micro-optical components, such as laser diodes, microlenses, fiber couplers, photodetectors, modulators and optical switches, and their integration into complex communications subsystems, that allows us to convey hundreds of petabytes about the planet. . . daily.

Micro-optics developed as a branch of classical optics in the latter half of the twentieth century, and initially focussed on miniaturized refractive lenses and novel forms of diffractive optics. In recent years it has begun to meld with the capabilities of microsystems such that the fields of micro-optics, optical microsystems, integrated optics, and optical MEMS share many technologies, components and envisaged applications. Thus

¹ “Some people talk in their sleep. Lecturers talk while other people sleep.” Alfred Capus (1857-1922), French journalist and playwright.

if we consider “micro-optics” in this text, we do so in the context of numerous other overlapping disciplines, and realize that the field is now much more than microscopic lenses and clever holograms.

As a result, whereas optical telecommunications continues to provide considerable impetus for further development of optical microsystems, the range of micro-optical applications has expanded significantly. Beyond the laser-based reader heads used in conventional optical data storage and the micromirror arrays on which many beamers rely, micro-optics is found in systems ranging from clinical instrumentation, including intra-corporal imaging and advanced optical medical diagnostics, to advanced scientific imaging concepts, with which astronomers use MEMS-based adaptive optical mirrors for wavefront correction to compensate for atmospheric distortion. As an increasing spectrum of industrial manufacturers becomes aware of the capabilities of micro-optics, the applications spectrum will likewise grow.

We therefore address the varied aspects of the micro-optics field in the text which follows, and run the gamut from the basics of classical optics to advances in nanophotonics. The scope is wide, but the focus is on the engineering of micro-optics: the reader should feel confident in the analysis or design of a micro-optical component or system after ploughing through the text. If, in addition, a fascination with photons ensues, I will have the same feeling of success as after holding an optics lecture in which no one audibly snored.

Freiburg and Zurich, January 2010

Acknowledgements

That this text saw completion is due in no small part to contributions from a great many people, all with better things to do than provide me with materials, pictures, data or information. It is with sincere gratitude that I acknowledge the selfless support of the many students, senior scientists, colleagues and random victims who happened to be in the lab at the wrong time who all cheerfully lent me a hand.

I am particularly grateful to many present and past members of the Laboratory for Micro-optics in the Department of Microsystems Engineering at the University of Freiburg for their unwavering support and also to many of my past students for their contributions. Thanks are due to Bernd Aatz for the fringes of Chapters 5, 7 and 8, and all the ZEMAX simulations in Chapter 7; Stefan Reichelt, now at SeeReal Technologies, Dresden (D), for the lens profiles in Chapter 5; Khaled Aljasem and Andreas Fischer, the latter now at Chalmers University of Technology, Göteborg (S), for the micromirror photos of Chapter 6; David Kallweit, now at the Fraunhofer Institute for Photonic Microsystems, Dresden (D), for his mirror pictures in Chapter 6 and grating structures of Chapter 8; Armin Werber, now at Carl Zeiss SMT, Oberkochen (D), for all the pneumatics and balloon pictures in Chapter 6, the pneumatic microlens process of Chapter 11 and the membrane microlens diagrams and photos of Chapter 12; Christoph Friese, now at Robert Bosch in Reutlingen (D), for his SU-8 adaptive optical mirrors, presented in Chapter 6; Fabian Zimmer, now at the Fraunhofer Institute for Photonic Microsystems, Dresden (D) for the polymer lens profile in Chapter 7; Andreas Mohr, now at the Fraunhofer Institute for Solar Energy Systems, Freiburg (D) for the embossed Fresnel lens in Chapter 7; Daniel Hofstetter, now at the Université de Neuchâtel (CH) for the DBR grating photograph of Chapter 8, the Michelson interferometer displacement sensor presented in Chapter 9 and the photograph of the quantum cascade laser of Chapter 10; Eva Geerlings, now at Sick AG, Waldkirch (D) and her colleagues at the Fraunhofer Institute for Applied Solid State Physics, Freiburg (D) for the photos of the external cavity structure of Chapter 8; Dennis Hohlfeld, now at the Holst Center of IMEC, Eindhoven (NL) for the thermo-optic filter photograph of Chapter 8 and the silicon optical bench with fibers in Chapter 11; Bernd Maisenhölder, now at Oerlikon Balzers (CH) for the Mach-Zehnder based refractive index sensor of Chapter 9; Robert Gehrke for the LED photograph and spectra of Chapter 10; Jens Fiala and Niklas Weber for the photograph of the implantable pulse oxymeter in Chapter 10; Daniel Mader for the photos of the molded PDMS micro-optical/microfluidic structures in Chapter 11 and the multi-chamber microlens picture of Chapter 12; Holger Krause, now working on oil

platforms off the coast of Nigeria, for the sol-gel lens photos of Chapter 11; Wolfgang Mönch for the photoresist reflow lens pictures of Chapter 11; Michael Engler, who now has his own optical design company, for the mold and injection-molded Fresnel lens of Chapter 11; Florian Krogmann, now at IST AG, Wattwil (CH) for the silicon optical bench of Chapter 11 and the liquid microlens pictures of Chapter 12; Thorsten Faber, now at the Fraunhofer Institute for the Mechanics of Materials, Freiburg (D), for the actuated liquid lenses of Chapter 12; Ryaz Pasha Shaik for the two-dimensional repositionable liquid lens of Chapter 12; Khaled Aljaseem and Wei Zhang for the membrane lens data of Chapter 12; and Christoph Schlögl and Yaxiu Sun for the bandgap calculations and 2D photonic crystal photos of Chapter 14.

In addition, I am most indebted to Wolfgang Mönch, Andreas Seifert, Daniel Mader, Philipp Waibel, Philipp Müller and Yaxiu Sun for reading through portions of the manuscript and providing corrections and useful comments. Additional and repeated thanks go to Daniel Mader for putting up with inane \LaTeX questions; Simon Dreher for writing the \LaTeX routine which allows the citations at the start of each chapter; Philipp Waibel for compensating for my inadequacies in *Mathematica*; Claudia Duppé for taming my propensity for excessive hyphenation; and Nadja Katthagen for keeping the wolves at bay when I needed a few hours without someone desperately clamoring for the Dean's opinion or signature on something.

I also extend a warm thanks to friends and colleagues distributed about the globe for their generous contributions of photos and figures representing some cutting-edge work in micro-optics. Thank you to Markus Rossi of Heptagon Oy (SF and CH), for the Fresnel-like diffractive lens and the DOE of Chapter 8; Oliver Ambacher and Wolfgang Bronner of the Fraunhofer Institute for Applied Solid State Physics (D), for the photograph of the OEIC chip in Chapter 9; Kerry Vahala of Caltech (USA) for the disk laser photograph in Chapter 10; James Harris of Stanford University (USA) for the SEM of the tunable VCSEL in Chapter 10; George Whitesides of Harvard University (USA) for the diagrams of his optofluidic dye laser of Chapter 13; Demetri Psaltis and Wuzhou Song of the EPFL (CH) for the schematics and optical characteristics of their opto-fluidic DFB lasers of Chapter 13; Jason Heikenfeld and Kaichang Zhou of the University of Cincinnati (USA) for the photos of their electrowetting display of Chapter 13; Susumu Noda of Kyoto University (JP) for the characteristics and photos of his photonic crystal resonant cavities in Chapter 14; Martin Wegener of the Karlsruhe Institute of Technology (D) for the photo of his split ring metamaterials in Chapter 14; Mark Reed of Yale University (USA) for his nanowires in Chapter 14; Shawn-Yu Lin of the Rensselaer Polytechnic Institute (USA) for his bent photonic crystal waveguides of Chapter 14; and Philip Russell of the Max Planck Institute for the Science of Light (D) for his sampler of photonic crystal cross sections of Chapter 14.

Finally, I extend my appreciation to all the Microsystems Engineering students who provided suggestions and corrections based on preliminary versions of the text used in the *Mikrooptik* and *Optical Microsystems* courses.

Notation

List of symbols

Notes:

- Vector quantities are given in **boldface** type.
- SI units are given except for those cases where non-SI units are traditional.

a	aperture dimension	m
a	spacing between cells (period) of a photonic crystal	m
A	area	m ²
A	aperture function	–
b	aperture dimension	m
\mathbf{B}	magnetic flux density	Wb/m ²
BW	bandwidth	Hz
c	speed of light in vacuum	m/s
d	spacing or displacement	m
C	contrast	–
d	penetration depth of evanescent field	m
D	aperture diameter	m
D	dispersion parameter	ps/km · nm
D_M	fiber dispersion due to the material	ps/km · nm
D_T	total fiber dispersion	ps/km · nm
D_{WG}	fiber dispersion due to the waveguide	ps/km · nm
\mathbf{D}	electric flux density	C/m ²
E	electric field (scalar)	V/m
\mathbf{E}	electric field (vector)	V/m
E	energy	J
E_a	electric field at a given point in an aperture	V/m
E_B	electric field of backward propagating wave	V/m
E_c	conduction band energy	eV
E_C	electric field due to diffraction from a circular aperture	V/m
E_D	electric field due to diffraction from dual slits	V/m
E_F	electric field of forward propagating wave	V/m

E_g	energy gap	eV
E_i	electric field at a given point in an image	V/m
E_i	electric field incident on an interface	V/m
E_N	electric field due to diffraction from N slits	V/m
E_r	electric field reflected from an interface	V/m
E_R	electric field due to diffraction from a rectangular aperture	V/m
E_S	electric field due to diffraction from a 1D slit	V/m
E_t	electric field transmitted through an interface	V/m
E_v	valence band energy	eV
$E_x(y)$	x-directed electric field in a waveguide	V/m
E_0	magnitude of electric field	V/m
E'_0	electric field pre-factor for diffraction integrals	V/m ²
f	focal length	m
f_C	focal length for red wavelengths	m
f_d	focal length of the diffractive part of a hybrid lens	m
f_F	focal length for blue wavelengths	m
f_i	focal length inside a medium, or on image side of lens	m
f_o	focal length on object side of lens	m
f_0	focal length outside a medium	m
f_r	focal length of the refractive part of a hybrid lens	m
F	force	N
F	finesse	–
F_{FK}	Franz-Keldysh coefficient	m ² /V ²
F_i	transmission matrix for thin film i	–
$f/\#$	f-number (of a lens)	–
g	circular aperture radius	m
$g(\nu)$	lineshape function	Hz ⁻¹
h	Planck's constant	Js
h	height of the source of a ray above the optical axis	m
h	half of the waveguide core thickness	m
\mathbf{H}	magnetic field (vector)	A/m
H_i	magnetic field incident on an interface	A/m
H_r	magnetic field reflected from an interface	A/m
H_t	magnetic field transmitted through an interface	A/m
I	current	A
I	optical intensity	W/m ²
I_d	photodetector dark current	A
I_{max}	maximum optical intensity	W/m ²
I_{min}	minimum optical intensity	W/m ²
I_s	photodetector signal current	A
$I(z)$	optical intensity as function of position	W/m ²
$I(\nu)$	optical intensity as function of frequency	W/m ²
I'	optical power	W

J	current density	A/m^2
k	propagation constant	m^{-1}
k	wave vector	m^{-1}
k_B	Boltzmann constant	J/K
k_I	imaginary part of the wave vector	m^{-1}
k_r	radial propagation constant (for diffraction)	m^{-1}
k_R	real part of the wave vector	m^{-1}
k_x	x-directed propagation constant	m^{-1}
k_{xg}	x-directed propagation constant inside waveguide core	m^{-1}
k_{xL}	x-directed propagation constant in lateral evanescent field	m^{-1}
k_y	y-directed propagation constant	m^{-1}
k_{yc}	y-directed propagation constant in waveguide cap	m^{-1}
k_{yg}	y-directed propagation constant in waveguide core	m^{-1}
k_{ys}	y-directed propagation constant in waveguide substrate	m^{-1}
k_z	z-directed propagation constant	m^{-1}
k₀	wave vector in free space	m^{-1}
k_0	propagation constant in free space	m^{-1}
K	spring constant	N/m
L	length	m
L_c	coherence length	m
L_C	overlap length in a waveguide coupler	m
L_{opt}	optical path length	m
L_x	lateral waveguide curve transition distance	m
L_z	axial waveguide curve transition distance	m
L_π	coupling length for waveguide proximity coupler	m
m	mass	kg
m	azimuthal index of Zernike polynomial	–
m	diffraction order	–
m	zone number in a diffractive Fresnel-like lens	–
m	waveguide mode number	–
m_e	electron mass	kg
m'	waveguide mode number for a mirror waveguide	–
m^*	electron effective mass	kg
M	magnification (for a Gaussian beam)	–
M_T	transverse magnification	–
$M_{T(C)}$	transverse magnification for red wavelengths	–
$M_{T(F)}$	transverse magnification for blue wavelengths	–
M	magnetization	Wb/m^2
n	refractive index	–
n	maximum order of Zernike polynomial	–
n	carrier concentration	cm^{-3}
n_A	refractive index of ambient in liquid lens systems	–
n_c	refractive index of waveguide cap material	–

n_c	refractive index of fiber cladding material	–
n_c	electron density in the conduction band	cm^{-3}
n_C	refractive index at Fraunhofer C-line	–
$n_{cladding}$	refractive index of the waveguide cladding material	–
n_{core}	refractive index of the waveguide core material	–
n_d	refractive index at Fraunhofer d-line	–
n_e	extraordinary refractive index	–
n_e	refractive index at Fraunhofer e-line	–
n_e	electron density	cm^{-3}
n_F	refractive index at Fraunhofer F-line	–
n_g	refractive index of waveguide or fiber core material	–
n_i	refractive index, incident side of interface	–
n_L	refractive index inside a lens	–
n_o	ordinary refractive index	–
n_s	refractive index of waveguide substrate material	–
n_t	refractive index, transmitted side of interface	–
n_v	electron density in the valence band	cm^{-3}
n_0	refractive index outside a lens	–
N	waveguide effective index	–
N	number (of atoms, spots, ...)	–
NA	numerical aperture	–
NA_{wg}	numerical aperture of a waveguide facet	–
NEP	noise-equivalent-power	W
N_i	number of illuminated grating periods	–
N_n^m	normalization factor for Zernike polynomial	–
P	polarization	C/m^2
P	number of phase levels in binary optics	–
P_d	dark power (in a photodetector)	W
P_{opt}	input optical power (in a photodetector)	W
Q	charge	C
Q	quality factor	–
r	distance	m
r	reflectivity	–
r	linear electro-optic (Pockels) coefficient	m/V
r	3D spatial vector	m
r	contour vector (for a contour integral)	m
r_a	aperture radial dimension	m
r_{Airy}	radius of Airy disk	m
r_i	image radial dimension	m
r_g	fiber core radius	m
r_{TE}	TE reflection coefficient	–
r_{TM}	TM reflection coefficient	–
r_0	radius	m

R	reflectance	–
R	radius of curvature	m
R	aperture / image spacing, for diffraction	m
R	spectrometer resolution	–
R	responsivity	A/W
R_L	Gaussian beam radius of curvature, left (input)	m
R_n^m	radial function of Zernike polynomial	–
R_R	Gaussian beam radius of curvature, right (output)	m
$R(z)$	Gaussian beam radius of curvature	m
R_1	radius of curvature of left lens surface	m
R_2	radius of curvature of right lens surface	m
s	sag height (of a lens)	m
s	quadratic electro-optical (Kerr) coefficient	m^2/V^2
S	scalar power density	W/m^2
\mathbf{S}	surface vector (for a surface integral)	m^2
\mathbf{S}	Poynting vector (for energy transfer)	W/m^2
S	scale factor for microlens size	–
S_{GH}	Goos-Hänchen shift	m
S_i	spacing between lens and image	m
S/N	signal-to-noise ratio	–
S_o	spacing between lens and object	m
t	time	s
t	transmittivity	–
t	waveguide core thickness	m
t_c	coherence time	s
t_{max}	maximum thickness of a Fresnel-like diffractive lens	m
t_{PR}	initial photoresist thickness for reflow microlens	m
t_{step}	height of a phase step for a Fresnel-like diffractive lens	m
t_{TE}	TE transmission coefficient	–
t_{TM}	TM transmission coefficient	–
T	period	s
T	temperature	K
T	transmittance	–
v	velocity	m/s
v_g	group velocity	m/s
v_p	phase velocity	m/s
V	Abbe number	–
V	volume	m^{-3}
V	visibility	–
V	Verdet constant (magneto-optics)	rad/Tm
V_π	modulator voltage required for phase shift of π	V
V_{cyl}	volume of a photoresist cylinder before microlens reflow	m^3
V_{hemi}	volume of a hemispherical microlens after reflow	m^3

w	width	m
w_x	width in the x direction	m
w_y	width in the y direction	m
W_E	electric energy density	J/m ³
W_H	magnetic energy density	J/m ³
$W(z)$	Gaussian beam (half-) width	m
W_0	Gaussian beam waist (half-) width	m
W_{0L}	Gaussian beam waist (half-) width at left (input)	m
W_{0R}	Gaussian beam waist (half-) width at right (output)	m
x_a	aperture x dimension (for diffraction)	m
x_i	image x dimension (for diffraction)	m
y	height of the intersection of a ray with the lens surface	m
y_a	aperture y dimension (for diffraction)	m
y_i	image y dimension (for diffraction)	m
y_i	image size	m
y_o	object size	m
z_i	distance between image and focal point	m
z_L	spacing between Gaussian beam waist and lens, left (input)	m
z_o	distance between object and focal point	m
z_0	Rayleigh range (half of depth of focus)	m
z_{0L}	Rayleigh range (half of depth of focus) at left (input)	m
z_{0R}	Rayleigh range (half of depth of focus) at right (output)	m
z_R	spacing between Gaussian beam waist and lens, right (output)	m
Z	impedance	Ω
Z_0	impedance of free space	Ω
Z_n^m	Zernike polynomial	–
α	absorption coefficient	m ⁻¹
α	evanescent decay constant	m ⁻¹
α	blaze angle (for blazed gratings)	rad
α	thermal coefficient of expansion	K ⁻¹
α_{dB}	logarithmic absorption coefficient	dB/m
α_{wg}	waveguide loss	m ⁻¹
α_0	residual absorption losses in a laser material	m ⁻¹
β	z-directed propagation constant in a waveguide	m ⁻¹
β_0	Bragg (resonance) condition in a 1D periodic structure	m ⁻¹
β_{1D}	propagation constant for coupled waves in a periodic structure	m ⁻¹
γ	damping coefficient	s ⁻¹
γ	optical gain	m ⁻¹
γ_{th}	threshold gain	m ⁻¹
Γ	phase retardation	rad
Γ	confinement factor, for waveguides	–
δ	phase shift between two electric fields	rad

Δr_{min}	optical resolution	m
Δx	spacing between peak and first minimum in an Airy disk	m
$\Delta\beta$	detuning from resonance in a periodic structure	m^{-1}
$\Delta\lambda$	linewidth	m
$\Delta\lambda_{FP}$	free spectral range	m
$\Delta\nu$	linewidth	Hz
$\Delta\nu_{FP}$	free spectral range	Hz
ϵ	dielectric constant	–
ϵ_d	permittivity of a dielectric layer	F/m
ϵ_m	permittivity of a material	F/m
ϵ_0	permittivity of free space	F/m
ζ	dummy variable	–
η	efficiency	–
η	electrowetting parameter	–
η_{abs}	absolute efficiency (of a grating)	–
η_{abs}	relative efficiency (of a grating)	–
θ_c	critical angle	rad
θ_{div}	divergence angle due to diffraction	rad
θ_i	angle of incidence	rad
θ_o	angle of diffraction	rad
θ_{pe}	external polarization (Brewster) angle	rad
θ_{pi}	internal polarization (Brewster) angle	rad
θ_r	angle of reflection	rad
θ_t	angle of transmission	rad
θ_{tilt}	maximum tilt angle of a micromirror	rad
θ_V	contact angle with applied bias	rad
θ_0	contact angle with no applied bias	rad
Θ	polarization rotation angle	rad
κ	coupling coefficient for waveguide proximity coupler	m^{-1}
κ_m	coupling coefficient in coupled wave analysis	m^{-1}
λ	wavelength	m
λ_B	blaze wavelength	m
λ_C	wavelength of Fraunhofer C-line	m
λ_d	wavelength of Fraunhofer d-line	m
λ_e	wavelength of Fraunhofer e-line	m
λ_F	wavelength of Fraunhofer F-line	m
λ_0	wavelength in free space	m
Λ	spatial period (e.g., of a grating)	m
μ	magnetic constant	–
μ_m	permeability of a material	H/m
μ_0	permeability of free space	H/m
ν	frequency	Hz
ν_d	Abbe number based on Fraunhofer d-line	–

ν_D	Abbe number for diffractive lens	–
ν_e	Abbe number based on Fraunhofer e-line	–
ξ	opto-thermal expansion coefficient	K^{-1}
ξ_d	opto-thermal expansion coefficient of a diffractive lens	K^{-1}
ξ_r	opto-thermal expansion coefficient of a refractive lens	K^{-1}
ρ	charge density	C/m^3
σ	conductivity	$\Omega^{-1}m^{-1}$
σ_{lv}	interfacial energy between liquid and vapor	J/m^2
σ_{sl}	interfacial energy between substrate and liquid	J/m^2
σ_{sv}	interfacial energy between substrate and vapor	J/m^2
τ	lifetime	s
$v(x)$	step function	–
ϕ	phase	rad
ϕ	tilt of polarization ellipse	rad
$\phi(z)$	Gouy phase shift	rad
$\Phi(x, y)$	phase function of a diffractive lens	rad
$\Phi(\nu)$	optical energy density per volume per frequency	J/m^3Hz
Φ_ν	optical energy density per volume	J/m^3
χ	susceptibility	–
$\psi(x)$	envelope function of Gaussian beam	V/m
$\Psi(x, y)$	phase function of a DOE wrapped to 2π	rad
ω	angular frequency	rad/s
ω_p	plasma frequency	rad/s
ω_0	resonance frequency	rad/s

List of acronyms and abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
AFM	atomic force microscope
APS	active pixel sensor
AR	anti-reflection
AWG	arrayed waveguide grating
BBO	barium borate
BD	Blue-ray disc
CA	chromatic aberration
CCD	charge coupled detector
CD	compact disc
CMOS	complementary metal oxide semiconductor
CRT	cathode ray tube
CVD	chemical vapor deposition
COC	cyclo-olefin copolymer
DBR	distributed Bragg reflector (laser)
DI	de-ionized (water)
DOE	diffractive optical element
DVD	digital versatile disc
DFB	distributed feedback (laser)
DOF	depth of focus
EDFA	erbium-doped fiber amplifier
EAP	electroactive polymer
EWOD	electrowetting-on-dielectrics
FBG	fiber Bragg grating
fcc	face-centered cubic
FOG	fiber-optic gyro
FTTH	fiber-to-the-home
FSR	free spectral range
FT	Fourier transform
FWHM	full-width at half-maximum
GaAs	gallium arsenide
GGG	gadolinium gallium garnet
HEBS	high-energy beam sensitive (glass)
HeCd	Helium-Cadmium
HEMT	high electron mobility transistor
HeNe	Helium-Neon
HR	high-reflection
ITU-T	International Telecommunication Union
LBO	lithium triborate

LED	light emitting diode
LiNbO ₃	lithium niobate
IBM	ion beam milling
InP	indium phosphide
IR	infrared
ITO	indium tin oxide
KDP	KH ₂ PO ₄
KTP	KTiOPO ₄
LCD	liquid crystal display
LPCVD	low pressure chemical vapor deposition
LSF	line spread function
LTCC	low-temperature co-fired ceramic
MEH-PPV	poly(2-methoxy-5-(2'-ethyl-hexoxy)-1,4-phenylene-vinylene)
MEMS	microelectromechanical systems
μTAS	micro-total-analysis systems
MMI	multimode interference coupler
MOEMS	micro-opto-electromechanical systems
MOS	metal-oxide-semiconductor
MOSFET	metal-oxide-semiconductor field-effect transistor
MOMS	micro-opto-mechanical systems
MUMPS	multi user MEMS processes
MSM	metal-semiconductor-metal (photodetector)
NA	numerical aperture
NEP	noise-equivalent-power
NIR	near infrared
OCT	optical coherence tomography
OEIC	opto-electronic integrated circuit
OMEMS	opto-micro-electromechanical systems
OPD	optical path difference
OPL	optical path length
OXC	optical cross-connect
PC	polycarbonate
PCF	photonic crystal fiber
PDMS	polydimethylsiloxane
PDOT	poly(3, 4-ethylene-dioxythiophene)
PECVD	plasma-enhanced chemical vapor deposition
PEDOT	alternative for poly(3, 4-ethylene-dioxythiophene)
PFPE	perfluoropolyether
PIC	photonic integrated circuit
PLC	planar lightwave circuit
PLZT	lead-lanthanum zirconate-titanate
PMMA	poly methyl methacrylate
PPV	poly-(para-phenylene vinylene)

POF	plastic optical fiber
PS	polystyrene
PSD	position-sensitive detector
PSF	point spread function
PTFE	polytetrafluoroethylene (Teflon [®])
QCL	quantum cascade laser
QCSE	quantum-confined Stark effect
RGB	red, green, blue
RIBE	reactive ion beam etching
RIE	reactive ion etching
rms	root mean square
SA	spherical aberration
SEM	scanning electron microscope
SERS	surface-enhanced Raman spectroscopy
SHG	second harmonic generation
SI	système international
SiOB	silicon optical bench
SLED	superluminescent light emitting diode
SMD	surface mount device
SMSR	side-mode suppression ratio
SOA	semiconductor optical amplifier
SOI	silicon-on-insulator
SPR	surface plasmon resonance
SRAM	static random access memory
SUMMiT	Sandia ultra-planar multilevel MEMS technology
SXGA	super extended graphics array
TDLAS	tunable diode laser absorption spectroscopy
TE	transverse electric
TEM	transverse electromagnetic
TIR	total internal reflection
TM	transverse magnetic
UV	ultraviolet
VCSEL	vertical-cavity surface-emitting laser
VOA	variable optical attenuator
WDM	wavelength division multiplexing
YAG	yttrium aluminum garnet
YIG	yttrium iron garnet
YVO	yttrium aluminum vanadate

Part I

Essential optics

1 Introduction

*Nature and Nature's laws lay hid in night: God said, "Let Newton be!" and all was light*¹

Alexander Pope

Optics is one of the oldest fields of human enquiry, the results of which are literally visible to a sizable fraction of the world's population on a daily basis. Micro-optics is a sub-discipline which has been more the domain of scientists and engineers, yet is increasingly, if sometimes almost invisibly, encroaching on the public domain.

The reader may justifiably ask at the outset: how do we distinguish between the two? The physics for optics and micro-optics is identical: image formation is the same for a set of 200 μm diameter microlenses as it is for 20 cm diameter projection lenses, even if secondary effects, such as diffraction, may play more of a role in one than in the other.

Nevertheless, micro-optics has two distinguishing characteristics:

- The feature sizes of micro-optics, and in many cases the sizes of the components themselves, are small, typically in the micrometer to nanometer range; and
- Micro-optical components are typically manufactured using highly-parallel mass fabrication techniques derived from those used by the semiconductor industry.

It is primarily for the latter reason that the history of micro-optics is a relatively recent one, as it is strongly coupled to developments in semiconductor technology.

Historical outline

Present day applications of micro-optics range from medicine to entertainment and it is interesting to see how we have arrived at this point. We therefore defer the rigor of the science of optics for one chapter and, in this introductory survey, spend a few moments looking at the origins of optics. The history of physics, of which optics is a part, is well documented in the literature, so we will take only a cursory look at the development of a few key technologies and applications in optics, as well as a glance at how our understanding of light has developed.

Whereas optics stretches into prehistory, micro-optics is at best two centuries old, and most of this field developed significantly only in the latter part of the twentieth century. Having seen how optics evolved through the centuries, we will then focus our

¹ The poet Alexander Pope (1688-1744) intended these lines as an epitaph for Isaac Newton.

attention on the short history of micro-optics, and will look at some of the significant predecessors to the technologies and devices which form the heart of this textbook.

Finally, so as not to leave the reader with the feeling that all we are about to study is abstract or of merely theoretical interest, we will conclude the chapter with a look at a micro-optical component which has undergone an almost unparalleled development, the digital micromirror device or DMD. A commercial product which proudly sits unrecognized in many living rooms, the DMD is hidden in many video beamers but represents micro-optical technology at its finest.

1.1 Optics: history in a nutshell

There is no shortage of texts available to the historically minded optical engineer which provide a detailed look at many aspects of the history of optics; eminently readable, compact histories may be found, for example, in (Römer, 2006, Chapter 1) or (Hecht, 2002, Chapter 1), concerning medieval optics in (Lindberg, 1983), and in the somewhat dated (Ronchi, 1970). An excellent history of fiber optics is the subject of (Hecht, 1999). Although history of course unfolds chronologically, we will take a somewhat thematic approach here, considering three fields of enquiry which occupied many of our optical ancestors: What is light? What are its properties? And how can I engineer optics to manipulate it?

1.1.1 The nature of light

Optics is the engineering of light. But what is light, this ethereal luminosity? As obvious as the answer seems to the scientifically trained public today, it took a few thousand years to get here. The Greeks thought a great deal about the subject, and the philosophers well known to us, Plato, Aristotle and Pythagoras, developed various theories of light. In Alexandria, the center of learning for the western world following the fading of Greece's Golden Age, Euclid² (about 300 BCE) was one of the first to realize that light travels in straight lines. He was convinced, however, that vision is due to light rays which emanate from our eyes and strike the object seen.

Optically, very little happened in the ensuing millennium, until Ibn-al-Haitham, known in the west as Alhazen (966–1020), who came from Basra in modern day Iraq but lived in Egypt, wrote numerous books on the study of optics; these, translated into Latin, became the source of virtually all knowledge on the subject for Europeans as they slowly awakened from the dark ages. In contrast to Euclid, Alhazen was certain that light is reflected *from* an object and that the rays enter the eyes, giving rise to vision.

Despite the volume of effort expended in the development of optics in the following centuries, it was well into the Renaissance before the discussion on the nature of light

² We will introduce a pantheon of optical individuals in this chapter; in later chapters, we will add a little footnote describing the person for which an equation or a concept is named, thereby adding a face to the effect.

took a new turn, based on efforts to understand the measurements made of light traversing optical components. René Descartes (1596–1650), a Frenchman who lived most of his life in The Netherlands, imagined that light must be due to pressure in an unseen medium. His contemporary Pierre de Fermat (1601–1665), also French, developed the theory of “least time”, which states that light will move from source to destination along the path which takes the smallest time; this principle was useful as an aid to understanding refraction.

Of what did these light beams consist? The Englishman Robert Hook (1635–1792), who first observed the details of diffraction and investigated the interference which give rise to colored films, thought that light must have a wave-like nature. His eminent compatriot, Isaac Newton (1642–1727) was convinced of the opposite: light must consist of particles. Newton was one of the first thinkers to rely on experiment rather than conjecture, and his extensive optical studies (described in a manner which still manages to captivate the modern reader in his text *Opticks* (Newton, 1740)) led him to assert that only a corpuscular nature of light could explain the rectilinear propagation that was by this time well known. The stage was thus set for an intellectual debate that was to be carried out, often acrimoniously, for several hundred years.

The weight of Newton’s considered opinion was such that proponents of a wave theory for light were often ridiculed or had their qualifications placed into doubt. Prominent corpuscularists, including most of the British scientific community as well as eminent French thinkers such as Jean Baptiste Joseph Fourier (1768–1830) and Pierre-Simon Laplace (1749–1827), stood in opposition to the Dutchman Christiaan Huygens (1629–1695), who developed a wave theory of light involving “wavelets”: each point on a propagating light wave emits in turn a secondary wavelet, and the next instant of propagation is defined by the sum of these wavelets which form a new wave. Alas, he required a medium for this to be explicable, an aether, which posed a problem to which we will return. The wave model was also supported by Francesco Maria Grimaldi (1618–1663), an Italian from Bologna, who used it to explain diffraction phenomena, and, most significantly, by the Englishman Thomas Young (1793–1829) and his French contemporary Augustin Jean Fresnel (1788–1827), who both studied interference and diffraction phenomena (Young, 1804). Young first deduced that light must be wavelike with transverse (as opposed to longitudinal, as in sound pressure waves) undulations; Fresnel extended Huygens’ approach to explain rectilinear wave propagation, thereby removing Newton’s primary objection to wave propagation.

The nineteenth century witnessed considerable development in the understanding of electric and magnetic phenomena, and these soon became relevant for a further understanding of light. In 1845, Michael Faraday (1791–1869) showed that the polarization of light can be rotated by the application of a magnetic field to a material in which it propagates (Faraday, 1846), hinting for the first time that optics and electromagnetics might be intertwined. The Scotsman James Clerk Maxwell (1831–1879) first unified the known relationships between electric and magnetic fields in his eponymous equations in 1873, from which it appeared that electromagnetic fields could propagate as waves with a velocity suspiciously identical to the known speed of light (Maxwell, 1861). Maxwell’s ideas were singularly unpopular, particularly in Britain, and it was a promi-

nent German physicist, Hermann von Helmholtz (1821–1894) who indirectly supported Maxwell’s conclusions by offering a prize to anyone who could verify his predictions. The verification did come from another German, the Karlsruher engineering professor Heinrich Hertz (1847–1894), whose experiments showed that electromagnetic fields could be reflected and refracted, and made to form standing waves, just as one had come to expect of light. Hertz’s experiments, published in 1888 (Süsskind, 1988), nine years after Maxwell’s death, seemed to show unequivocally that light was an electromagnetic wave.

There was one remaining problem. All “waves” known to nineteenth century scientists propagated in a medium: air, water, and the aether. The difficulty with the latter was that it could not be detected, and various experimental (particularly astronomical) observations only made sense if this aether were fixed as objects (such as the earth and the rest of the universe) moved through it. Since the speed of light was known to be constant (in the aether), measurements of its velocity should thus vary if taken parallel or normal to the motion of the earth. The famous experiments of the Americans Albert Michelson (1852–1931) and Edward Morley (1838–1923) showed that there was no detectable motion of the earth with respect to the aether (Michelson and Morley, 1887), shedding serious doubt on its existence (the aether, not the earth). One of the first to interpret the results of Michelson and Morley as implying that there actually was no such thing as the aether was the eminent French scientist Jules Henri Poincaré (1854–1912) and the issue was seemingly put to rest by the special theory of relativity published by Albert Einstein (1879–1955) (Einstein, 1905): light waves always propagate at the same speed in vacuum, regardless of the motion of source or observer. The aether had evaporated.

Thus light was clearly a wave. Or was it? Seemingly unrelated developments in the earlier nineteenth century were unknowingly providing the seeds for the growth of an entirely new field, atomic physics. William Wollaston (1766–1828) first noted that the solar spectrum showed dark absorption lines (Wollaston, 1802), a fact re-discovered and extensively studied by the Munich glass factory owner and precision optical equipment manufacturer Joseph von Fraunhofer (1787–1826), whose name today graces these optical features. In Heidelberg, Robert Bunsen (1811–1899) and Gustav Kirchhoff (1824–1887) studied the emission spectra of a variety of media, concluding that these spectra are characteristic for any material (Kirchhoff and Bunsen, 1860). The Swiss Johann Balmer (1825–1898) derived a set of empirical expressions for the emission lines of hydrogen in 1885.

Based on these studies, the optical properties of the atom, of hitherto undiscovered structure, came more into focus. Max Planck (1858–1947) proposed that radiation emitted from a body does so in discrete quanta and Einstein, building on these ideas, suggested that light could consist of particles of distinct, discrete energy (Einstein, 1906). These were among the first building blocks of the field of quantum mechanics, which saw rapid development in the early part of the twentieth century, with contributions by the most illustrious physicists of the age. Niels Bohr (1885–1962) developed the quantum theory of transitions in the atom, which requires discrete optical quanta, and in England in 1927, Paul Dirac (1902–1984) quantized the electromagnetic field (Dirac, 1927). The optical quanta, these “corpuscles of light,” were dubbed *photons* by the

Berkeley chemist Gilbert Lewis in 1926 (Lewis, 1926) and it seems subsequently to have become clear that light is both particle and wave and we now know what light is. Consider, however, that Einstein, toward the end of his life, wrote to his best friend “All these fifty years of conscious brooding have brought me no nearer to the answer to the question ‘What are light quanta?’ Nowadays every Tom, Dick and Harry thinks he knows it, but he is mistaken.³” Some of us may thus hesitate a bit in affirming that the nature of light is really definitively understood.

1.1.2 The properties of light

While trying to understand its nature, researchers from the time of antiquity have been fascinated by the many aspects of light, many of which were only slowly understood over the centuries. Refraction and reflection were known to exist by the Greeks, even if the effects were not explainable. Euclid and Ptolemy (approximately 130), both working in Alexandria, wrote about reflection and refraction, respectively, as did Alhazen, almost one thousand years later. Very little happened in between.

The first modern text which addressed these topics was the *Perspectiva* written by Witelo (b. 1230– d. between 1280 and 1314), from Silesia, in modern day Poland, although some historians argue that the work is merely a translation of Alhazen. The book, which appeared in about 1270, became the standard optics textbook until the 17th century. Witelo considered refraction and was aware that the angle of transmission of light into a medium did not vary in direct proportion to the angle of incidence, but could neither derive the relationship nor explain the phenomenon. Almost 400 years later, the German Johannes Kepler (1571-1630) discovered total internal refraction and noted that the intensity of light decreases with the square of distance from the source, but was also unable to describe refraction adequately. He was also one of the first to consider the speed of light, thinking it infinite.

Willebrord Snell (1591–1626), a Dutchman from Leiden, finally explained the behavior of refracted light, leading to the law of refraction which now bears his name; it was Descartes who first formulated the relationship in terms of sines, the form we are familiar with today. Fermat showed in 1657 that the Snell’s relationship was consistent with his dictum that light propagates along the path which requires the least time, if the speed of light is slower in the denser medium. The latter point was still a bone of contention at the time.

Shortly thereafter, Robert Hook observed the effects of diffraction, as did Grimaldi for clear apertures and in the shadows of opaque objects. Hook concluded that the colors present in thin films, such as in flakes of mica or in thin soap films, were due to interference phenomena, but could not find a relationship between color and film thickness. The understanding of this effect was aided by Newton, who decomposed white sunlight into its constituent colors using a prism in 1666, causing the poet John Keats (1795–1821) to lament the unweaving of the rainbow, “Philosophy will clip an Angel’s wings, Conquer

³ Albert Einstein to Michele Besso, December 12, 1951 (Klein, 1979)

all mysteries by rule and line, Empty the haunted air, and gnomed mine—, Unweave a rainbow...”⁴ Those of us who find beauty in science could hardly disagree more.

The Dane Erasmus Bartholinus (1625–1698) first described double-refraction (birefringence) in Iceland spar (calcite) in 1669. That this effect is due to polarization of light was known to Newton and ultimately adequately explained by Christiaan Huygens. It was shown by the Frenchman Etienne Louis Malus (1775–1812) in 1808 that polarization could be induced not only by specific crystal materials, but also through reflection. In his famous (accidental) experiment, Malus looked at the windows of the Palais du Luxembourg in Paris through a crystal of calcite and observed that the reflected light was polarized. The cantankerous Scottish minister David Brewster (1781–1868) experimented extensively on the properties of light, studying reflection, polarization and absorption; he discovered the polarization angle (now often bearing his name) in 1811 but benefitted most (financially and in posterity) from his invention of the kaleidoscope (Brewster, 1831).

Brewster was one of the last undying corpuscularists but many of his contemporaries were wrestling with the phenomena of polarization, diffraction and interference, trying to reconcile the effects with the assumed wave nature of light. In France, Jean Augustin Fresnel and Dominique Francois Arago (1786–1853), concomitantly with Thomas Young in England, worked extensively on this problem, coming to the crucial conclusion that light must consist of *transverse* waves. This result paved the way for significant further work on diffraction, notably by Joseph von Fraunhofer, who made the first grating-like structures, using closely-spaced arrays of tightly spanned wires, and issued a first comprehensive theory of diffraction in 1823. In England, George Airy (1801–1892) first calculated the form of the diffraction intensity distribution resulting from transmission through a circular aperture (Airy, 1835), a pattern which still bears his name.

Reflection, refraction and diffraction were effects whose understanding proved to be essential for designing optical components. A further characteristic of light, namely its speed, was also the subject of increasing interest in the seventeenth century. In Denmark, Ole Rømer (1644–1710) used measurements of the time required by Io, one of the moons of Jupiter, to transit behind the planet, taken for two positions of the Earth in its orbit, to estimate the speed of light. Based on imprecise knowledge of the radius of our mother planet’s orbit, a value of $c \approx 2 \times 10^8$ m/s was determined.

The first terrestrial measurement of c was performed in 1849 by the Frenchman Armand Hippolyte Louis Fizeau (1819–1896) who used light transmission through a rapidly rotating gear and its reflection from an 8.6 km distant mirror to obtain a more accurate value, $c \approx 3.133 \times 10^8$ m/s (Fizeau, 1850). Also in France, Jean Foucault (1819–1868) demonstrated conclusively that light in a material moves more slowly than in vacuum and in 1850 employed a rotating mirror arrangement to measure its speed, arriving at $c \approx 2.98 \times 10^8$ m/s. Albert Michelson also used rotating mirrors with a 35 km optical path length and determined $c \approx 2.99796 \times 10^8$ m/s in 1926 (Michelson,

⁴ John Keats *Lamia*, Part II

1927). The value for c converged as the century continued; the defined⁵ value today is $c = 2.99792458 \times 10^8$ m/s.

1.1.3 The essential optics

The engineering of optical components is likely to be considerably older than structured deliberation about the properties of light. Numerous optical devices are known from the dawn of history: Egyptian mirrors date from 1200 BCE; it is known that the Babylonians had lenses made from quartz or crystal; and similar archeological finds have been made in Assyria, dating from 700 BCE and in Crete, from about 500 BCE.

It was again Alhazen to whom we are indebted for the first more or less modern description of basic optical components. In his 11th century work, he described spherical and parabolic mirrors, explained the magnifying properties of lenses, and described spherical aberration as well as refraction. One of the first known practical applications for optical components, namely the use of lenses for spectacles, was described in England by Roger Bacon (ca. 1214– ca. 1292), who proposed that these would be useful for correcting vision. Spectacles in general use appeared soon thereafter, notably in Florence, Italy, around 1280; these used solid glass lenses, portions of glass spheres, as opposed to the water-filled spherical containers which had been used as lenses for optical experiments up to that time.

In the sixteenth century, Johannes Kepler was one of the first to describe the structure and workings of the eye, realizing correctly that the image is projected onto the retina. He explained how spectacles work and developed the first-order theory of thin lenses. In 1608, a German-born Dutch spectacle maker named Hans Lippersley (1570–1619) first described in writing the combination of a convex and a concave lens, joined by a tube: the telescope. Hearing of this, Galileo Galilei (1564–1642) built one merely a year later, observed the mountains of the moon and the four moons of Jupiter, and changed forever the way we look at the universe. Kepler improved on this original design, incorporating a convex rather than a concave eyepiece, and established the basis for the design on which modern refractors are still based.

In the same era, another Dutch lens maker, Hans Jansen (or, by some accounts, his son Zacharias Jansen), is believed to have proposed the the first practicable compound microscope. Still in Holland, Anton van Leeuwenhoek (1632–1723), a Delft haberdasher who took up microscopy as a hobby, ground his own lenses and made hundreds of observations using a single-lens magnifier; he thereby avoided the strong chromatic aberration which limited the performance of compound lens systems. Robert Hook designed a three-lens microscope, with a converging “telescope” eyepiece, as designed by Christiaan Huygens, which represented a considerable improvement over the original systems.

⁵ The speed of light is no longer measured; since 1983, it has been defined based on the definition of the meter, which is in turn the distance light travels in vacuum in $\frac{1}{299792458}$ s.

Chromatic aberration proved to be a serious problem for telescopes as well, and Isaac Newton, believing the problem unsolvable, designed what is known today as the Newtonian reflector in 1668. Despite the quality and future applicability of this telescope design, chromatic aberration was not an unsolvable problem. The English jurist and amateur mathematician Chester Moor Hall (1703–1771), first proposed a combination of crown and flint glass as an achromatic lens; he had it fabricated and made the first achromatic telescope, but failed to have the lens patented. That latter mistake was not made by the London silk weaver turned lens maker John Dollond (1706–1761), who patented and produced achromatic lenses using this principle starting in 1757, subsequently being appointed as optician to the king.

In the eighteenth and nineteenth centuries, lenses, and the optical equipment made from them, improved rapidly. The American polymath, politician and bon vivant Benjamin Franklin (1706–1790) is credited with having invented bifocals. The astigmatic George Airy used cylindrical lenses to make spectacles which corrected astigmatism and in 1888 the German Adolf Eugen Fick (1829–1901) first proposed and demonstrated contact lenses (Fick, 1988).

The design and improvement of lenses and lens systems was supported by developments in theoretical understanding of optical systems. Instrumental in these advances was the German Johann Carl Friedrich Gauss, who proposed a comprehensive mathematical description of imaging, forming the basis for what is today known as “Gaussian optics.” In England, Joseph Jackson Lister (1786–1869), a wine merchant and amateur microscopist, was the first to propose and fabricate achromatic lens systems for microscopes, and also produced designs which corrected spherical aberration; his son, also Joseph, trained by the father in microscopy, is arguably better known today, for promoting antiseptic procedures in surgery. The German optician Ernst Abbe (1840–1905), professor of physics in Jena and a director of the Carl Zeiss company, finally, developed a more detailed theory of image formation and invented approaches for lens design which corrected many aberrations, including coma; many of these techniques are unparalleled today and are still in use.

With that we arrive in the twentieth century, the first half of which saw optics established as a considerable industrial enterprise and become the subject of a wide spectrum of research endeavors. Perhaps the two events which did most to irrevocably expand the field were the invention of the light bulb by the American Thomas Edison (1847–1931) in 1879 (Edison, 1881) and the invention of the laser, also by an American, Theodore Maiman (1927–2007) in 1960 (Maiman, 1960). These light sources have become as established in optics technology as much as the lenses first regarded by the Assyrians three millennia ago.

1.2 Micro-optics: a smaller nutshell

The history of micro-optics is considerably younger than that of macroscopic optics. The expression *micro-optics* first appears in the literature only in the 1980s (Iga et al., 1984), but some of the optical components which today comprise our field are consid-

erably older. The diffraction grating is likely the senior member of the micro-optical family, and is probably the first optical device for which micrometer size features are essential; one of the classic modern texts in diffractive optics is that of Hutley, and his historical introduction (Hutley, 1982, Chapter 1) inspires us here.

The first mention of fabricated periodic structures with an optical effect refers to the studies of the American astronomer and public official David Rittenhouse (1732–1796). Apparently inspired by the transmission of light through a fine silk handkerchief, he fabricated closely-spaced arrays of fifty hairs and noted wavelength-dependent diffraction as well as the separation of an image into its various color components. Rittenhouse, however, made no quantitative analysis of what he saw, so it is Joseph von Fraunhofer, a half century later, who is commonly credited with having invented the diffraction grating. Working with periodic arrays of fine wires and ruled grooves on a mirror in 1821, von Fraunhofer derived the grating equation, quantified the wavelength dependence of diffraction and for the first time used these advances to measure the wavelength of light.

In England, Lord Rayleigh (1842–1919) with the civil name John William Strutt, first used gratings for spectroscopy (Rayleigh, Lord, 1874), showing that the resolution was superior to that achievable using prisms. However, gratings did not replace prisms for this application until the 1950s, due primarily to their high cost, which was based on the way they were manufactured. The American Henry Augustus Rowland (1848–1901) at Johns Hopkins University in Baltimore, invented the grating ruling engine in 1882 and Robert W. Wood (1868–1955), of the same institution, devised the concept of blazing to improve grating efficiency. In 1915, Albert Michelson employed interferometric measurement techniques coupled with servo mechanisms on a ruling engine for precise positioning and spacing of the grating periods. The precision of the manufacturing process made Johns Hopkins the prime source for ruled diffraction gratings world-wide in the early twentieth century.

The popularity of diffraction gratings increased markedly with a decrease in their cost brought about by the use of replication techniques for mass producing the components. Patented by John U. White and Walter A. Fraser of Perkin-Elmer Corporation in 1949 (White and Fraser, 1949), the approach, still the basis for low cost grating manufacture today, allowed the reproduction of large volumes of high-precision gratings from a single expensive master. As a result, the spectrum of potential applications expanded rapidly and replication is still one of the essential techniques employed in the fabrication of modern diffractive optical components.

Working in England in 1948, the Hungarian Dennis Gabor (1900–1979) invented holography (Gabor, 1948). Using a coherent electron beam (the laser was still 12 years away), Gabor showed that an interference pattern generated and written into a film by reflection from an object and a coherent reference beam could be used to store image information about the object, and this image could be reconstructed upon coherent illumination. Although the technique is now done with photons rather than electrons, it forms the basis for the generation of general diffractive structures which, upon illumination with a coherent light source, can yield almost arbitrary intensity patterns.

The invention of the laser in 1960 greatly expanded the capabilities of holography and related techniques. In 1968, it was shown that one-dimensional interference pat-

terns could be used to generate gratings by exposure of photosensitive material (Burch and Tokarski, 1968), thereby obviating the need for mechanical ruling of diffraction gratings; this so-called interference lithography is the method of choice for defining 1-, 2- and 3-dimensional periodic structures today. In the late 1960s, the generation of holograms by computer was demonstrated (Lohmann and Paris, 1967), allowing the definition of arbitrary diffraction patterns (no longer necessarily periodic) in a film and their associated intensity patterns generated upon illumination. In the same era, kinoforms, phase-only holograms, were conceived, these reducing the parasitic diffraction orders and improving the efficiency of image generation (Lesem et al., 1969).

The advent of lithography technologies, developed primarily by the burgeoning semiconductor industry, in the early 1970s, provided new opportunities for diffractive optics and ultimately paved the way for micro-optics as we know it today. Newly developed abilities to photolithographically define arbitrary two-dimensional patterns with micrometer accuracy on a semiconductor or glass substrate allowed definition and fabrication of microlenses (Iga et al., 1982), Fresnel zone plates or Fresnel lenses defined by multiple phase levels (d'Auria et al., 1972; Fujita et al., 1981). These were circularly-symmetric structures with micrometer-sized features both laterally and in depth and represent among the earliest microlens structures of the form we often employ now, the wide, relatively flat components fabricated on a transparent substrate. By the mid 1980s, these diffractive lenses had advanced in vertical resolution using the concepts of binary optics, in which N masks can be used to fabricate 2^N phase levels (Stern, 1997); for large numbers of phase levels, the discrete surface steps in one of these Fresnel-like lenses approach the continuous form of a refractive lens.

The wish to make truly refractive lenses, with a continuous surface profile, became reality in the 1980s through the use photolithographically patterned circular posts which were subject to reflow at high temperatures; surface tension forms a spherical cap of the molten resist and thus, upon hardening, a spherical lens (Popovic et al., 1988; Daly et al., 1990; Daly, 2001). The photoresist itself could be used as the optical element, or the pattern could be transferred into the substrate by etching; the technique allows fabrication of accurately defined single lenses or large arrays and is still a standard technology.

Micro-optics, however, is more than gratings and microlenses, even if it once may have seemed that way. The field finds itself today at the juncture of numerous technologies, which in combination have considerably broadened the scope of possibilities for miniaturized optics (Herzig et al., 2005; Borrelli, 2005; Sinzinger and Jahns, 2003; Kress and Meyrueis, 2009). One of these essential technologies is that of light sources; for micro-optical systems, these are most often laser diodes or LEDs. The history of semiconductor light emitters begins with the invention of the semiconductor laser diode, demonstrated and published by four American groups almost simultaneously in 1962. The rapid development of diode lasers, spurred primarily by the goal of viable optical telecommunications, has resulted in a field that has advanced to the point that ultra-miniaturized, high-performance semiconductor light emitters are now available at reasonable cost and at high volumes at wavelengths ranging from the ultraviolet to the far infrared; for a compact historical overview, see (Zappe, 2004, Chapter 1.1).

Closely coupled to the development of the laser diode is that of the optical fiber, a second technology with relevance for micro-optical systems. Fiber optics is also a child of the 1960s, whose history is engagingly presented in (Hecht, 1999). Whereas optical fibers were originally developed as a means to transmit optical signals over hundreds of kilometers, again primarily for telecommunications applications, the maturity of the technology implies that micro-opticians can now make do with considerably shorter pieces of fiber for short range interconnect in micro-optical systems. Micro-optics and fiber optics have also evolved symbiotically on another level: for the millions of kilometers of optical fiber deployed for telecommunications worldwide, optical signal generation and detection at the fiber ends is almost invariably performed by a micro-optical system.

Related to both fibers and laser diodes is field of integrated optics, the technology in which light is guided by waveguides, microlithographically fabricated components which also rely on the capabilities of semiconductor processing for their manufacture (Zappe, 1995). Originally proposed as a form of “laser beam circuitry” (Miller, 1969b) in the late 1960s, integrated optical circuits rapidly began to combine passive waveguides with lasers, detectors, modulators and other optical components, leading to complete photonic integrated systems (Koch and Koren, 1991). Whereas micro-optics often tacitly assumes miniaturized free-space optics (as opposed to guided wave optics, in fibers or waveguides), many optical microsystems today combine technologies from both areas.

An umbrella technology with relevance for all of those mentioned in the previous paragraphs is that of MEMS, micro-electromechanical systems (Fujita, 1997; Petersen, 2005). Starting in the 1960s, engineers began using the materials and processes being rapidly developed for integrated electronics for the fabrication of micromechanical structures, initially pressure sensors. In the 1980s, the field expanded rapidly and the spectrum of micromechanical components and structuring processes grew to the point that one could veritably witness the birth of an entirely new engineering discipline. Optical functionality began to appear in MEMS in the early 1990s, leading to acronyms such as MOEMS (micro-opto-electromechanical systems) or MOMS (micro-opto-mechanical systems), colorful but increasingly obscure.

MEMS has now grown to include much more than electromechanics, and the field is now better characterized by the designation “microsystems”. Fluidic, acoustic, biological, magnetic, chemical and of course optical microsystems are presently the subject of considerable academic research and are already found in a wide variety of commercial products (Gianchandani et al., 2007). The interdisciplinary nature of the microsystems domain, which includes a large portion of the micro-optics field, makes it now, in the first decade of the 21st century, one of the most exciting engineering areas in which to work. The realm of micro-optics encompasses (and is encompassed by) many of these disciplines, such that it shares the rich history and the bright future of all of them.

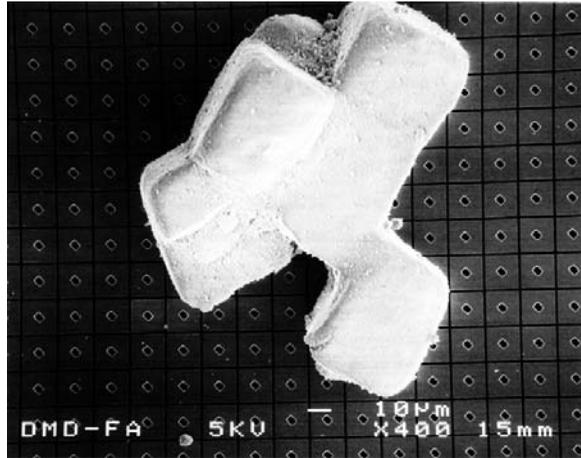


Figure 1.1 A DMD chip showing the individual mirrors, with a salt grain providing an indication of its scale. The little rectangles at the center of each mirror are the mounts to the mechanics hidden below, and can be seen if one looks at the enlarged projected image from a DMD beamer carefully. Photo ©Texas Instruments; reprinted courtesy of Texas Instruments.

1.3 Micro-optics for the home: the DMD

We conclude this introductory section with a glance at a micro-optical component that is probably not only the most commercially successful of those that have survived beyond the bounds of the laboratory, but also one found in products which some readers may actually have in their home. The *Digital Micromirror Device*, or DMDTM, is an array of electrostatically-movable micromirrors fabricated on a single substrate and is used for digital display technology. We will look at the DMD in more detail in the chapter on reflecting micro-optics, Section 6.6.1, but mention it here in the introduction to dispel any lingering notion the reader may have as to the technical, commercial and social relevance of micro-optics.

The DMD was invented by Texas Instruments and first presented, after about ten years of development, in 1985 (Hornbeck, 1993; Van Kessel et al., 1998). The first digital projector based on the micromirrors was launched in 1996, and ten years later, the DMD had become one of the stars of the MEMS industry, providing in excess of \$ 1 billion in revenues for the manufacturer (Petersen, 2005). The prime example of the product in which the DMD is found is in a family of video beamers⁶, whose price in recent years has decreased to the point that there is a large home market in which the CRT or plasma television has been replaced by a projector. It is likely that the fewest consumers know that that there are hundreds of thousands to millions of micromirrors lurking in their beamer.

⁶ Not all beamers are based on the DMD; many use LCD (liquid crystal display) technology. DMD-based beamers usually have a DLPTM (Digital Light Processing) logo somewhere on it.

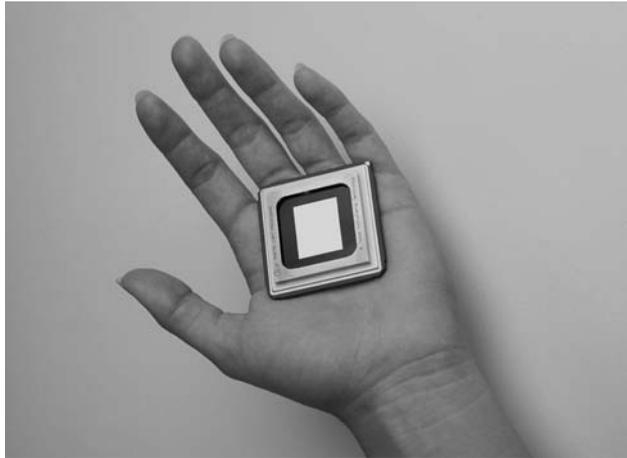


Figure 1.2 An SXGA DMD chip with more than 1.3 million micromirrors. Photo ©Texas Instruments; reprinted courtesy of Texas Instruments.

The DMD is based on an optically simple device: a $16\ \mu\text{m} \times 16\ \mu\text{m}$ aluminum tilting micromirror. A DMD chip consists of a large two-dimensional array of these mirrors, as seen in Figure 1.1, where each individually controllable reflector is used to define a single pixel in a projection display. The largest DMD chip currently available, with SXGA resolution (1280×1024 pixels, or mirrors), shown in Figure 1.2, implies that one can hold a chip with over 1.3 million individually addressable micromirrors in the palm of one's hand.

The exceptional color saturation and brightness achievable with DMD projection optics has been one of the keys to its success. Another is certainly its robustness and reliability: mirrors have been cycled continuously for 10^5 hours, implying 10^{12} cycles, without failure. As a result, the technology is being applied to a new spectrum of applications, including digital photo-finishing and most significantly digital cinema. Tests of the latter have generated public acclaim, due to the quality of the projected image, and the DMD is poised to play a significant role in the replacement of the century-old projection of cinema images through a moving film with fully digital technology.

The DMD is just one example of successful implementation micro-optics technology. In the chapters which follow, we will see a continuously broadening spectrum of such success stories.

2 The physics of light

We can scarcely avoid the conclusion that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena¹

James Clerk Maxwell, 1861

To understand optics, we must understand light. A topic which alone would fill a large textbook, the physics of light forms the basis for understanding how optical fields propagate through materials and interfaces, leading to the effects on which optics is based. In this chapter, we will review some of the basic physics required for understanding light propagation. The wave equation is derived from Maxwell's equations and the types of wave solutions which satisfy it are discussed. We review energy and power flow, introduce polarization, characterize the wavefront and finally discuss Gaussian beams. Numerous applications-oriented optics textbooks (Lipson et al., 1995; Smith and King, 2000; Iizuka, 2002a,b; Brooker, 2003) discuss some of these topics in greater detail.

2.1 Basic electromagnetics

Our discussion here will focus on rudimentary electromagnetic field theory. Many readers will be familiar with some of the fundamentals of electromagnetics and a wide selection of didactically excellent textbooks is available on the subject (Kraus and Carver, 1992; Jackson, 1998; Sadiku, 2001). Nonetheless, this first chapter provides an overview from which all students may benefit; even those readers for whom this is completely old hat may wish to peruse this section to familiarize themselves with the notation we employ in later sections.

2.1.1 Notation

We digress for a moment and briefly present the notational conventions we will employ in the text. Scalar variables are generally given in an *italicized* font, thus for example E for the scalar electric field. Vector quantities, having magnitude and direction, are denoted using **boldface** type, thus \mathbf{E} for the electric field vector.

¹ Whereas Maxwell could scarcely avoid drawing this conclusion (Maxwell, 1861), it took experimental verification by Hertz in 1888 to convince the rest of the scientific community.

Useful for defining spatial directions is the unit vector, defined as a vector of length unity with a defined direction given by its subscript. Thus \mathbf{u}_x corresponds to the x -direction in the Cartesian coordinate system. Given an x -directed electric field \mathbf{E}_x , the unit vector for the x -direction would then be found from

$$\mathbf{u}_x = \frac{\mathbf{E}_x}{|\mathbf{E}_x|}. \quad (2.1)$$

Tensorial quantities will also appear from time to time, but no distinct notation is employed for these. We will explicitly make it clear in context when a parameter is a tensor. Likewise, complex numbers are not subject to a distinct notation.

Finally, most physical quantities have dimensions and the units are defined in [square brackets] when the quantity is first defined. Thus the electric field vector, \mathbf{E} [V/m], has units of volt per meter. SI units are employed semi-rigourously, meaning that we bow to tradition in certain special cases (such as cm^{-3} for carrier concentration). Dimensionless quantities are defined with empty brackets, such as refractive index, n [].

2.1.2 Field quantities

Electric and magnetic fields

Light is an electromagnetic wave and is thus composed of propagating electric and magnetic fields. The electric field, \mathbf{E} [V/m] is related to the electric flux density \mathbf{D} [C/m²] by the simple relationship

$$\mathbf{D} = \epsilon_m \mathbf{E} \quad (2.2)$$

where ϵ_m [F/m] is the permittivity of the medium in which the field is found. We can alternatively include the polarizability, \mathbf{P} [C/m²], to yield

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}. \quad (2.3)$$

Analogously, the magnetic field, \mathbf{H} [A/m] and the magnetic flux density, \mathbf{B} [Wb/m²], are related as

$$\mathbf{B} = \mu_m \mathbf{H} \quad (2.4)$$

for which μ_m [H/m] is the permeability of the medium, or, when considering the magnetization \mathbf{M} [Wb/m²], as

$$\mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M}. \quad (2.5)$$

Permittivity and permeability

The permittivity and permeability are constants which define the characteristics of the medium through which the fields extend and we will see in subsequent chapters that these parameters are of fundamental importance in defining the optical characteristics of materials. In free space (or vacuum), the permittivity and permeability have values

$$\epsilon_0 = 8.8542 \cdot 10^{-12} \text{ F/m} \quad (2.6)$$

and

$$\mu_0 = 4\pi \cdot 10^{-7} \text{ H/m} \quad (2.7)$$

respectively. Outside vacuum, these free-space values increase in magnitude and are given by

$$\epsilon_m = \epsilon \epsilon_0 \quad (2.8)$$

$$\mu_m = \mu \mu_0 \quad (2.9)$$

where the subscript m denotes the value in a material. The dielectric constant $\epsilon []$ and the magnetic constant $\mu []$ are dimensionless numbers which define how the material differs from vacuum. In most (but not all) materials relevant for optics, $\mu = 1$ but ϵ typically differs significantly from unity and is one of the most important material parameters in optics. As we shall see in later sections, ϵ is typically a tensorial quantity and in general complex.

2.1.3 Maxwell's equations

A number of relationships have been derived for describing the variation of electric and magnetic fields in time and space and as a function of each other.

Gauss' law

Gauss' law² relates the electric field to the charge density in a given volume, namely

$$\nabla \cdot \mathbf{D} = \nabla \cdot \epsilon_m \mathbf{E} = \rho \quad (2.10)$$

for charge density $\rho [C/m^3]$. In free space, $\rho = 0$, and we are left with $\nabla \cdot \mathbf{D} = 0$. Written in integral form, Gauss' law may be expressed as

$$\oint_S \mathbf{D} \cdot d\mathbf{S} = \int_V \rho \cdot dV = Q \quad (2.11)$$

for integration over the surface \mathbf{S} of a volume V . $Q [C]$ is the total charge in the volume.

Gauss' law states that electric field lines originate and terminate on charges. As we see schematically in Figure 2.1, any difference between the total electric field entering a given volume and that leaving the volume has to be accounted for by charges inside that volume.

Divergence of magnetic fields

An analogous relationship holds for the magnetic field, namely

$$\nabla \cdot \mathbf{B} = \nabla \cdot \mu_m \mathbf{H} = 0 \quad (2.12)$$

² Johann Carl Friedrich Gauss (1777–1855), the German polymath who made fundamental contributions in mathematics, astronomy, magnetism and optics. The breadth of his activities was legendary, ranging from theoretical contributions in geometry to astronomical observations which he continued into his 70s.

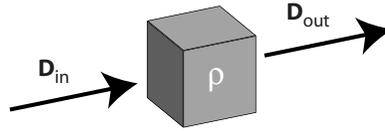


Figure 2.1 Gauss' law states that any difference between the electric field entering a volume and that exiting must be accounted for by electric charges in that volume.

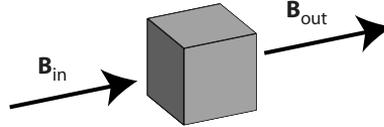


Figure 2.2 Since there are no magnetic charges, all magnetic field lines entering a volume must also exit that volume.

or, in integral form,

$$\oint_S \mathbf{B} \cdot d\mathbf{S} = 0. \quad (2.13)$$

Sometimes referred to as a variant of the Biot-Savart law,³ or occasionally informally as “Gauss’ law for magnetic fields”, this relationship is actually nameless. The relationship states that the divergence of the magnetic field is zero, due to the fact that there are no magnetic charges analogous to electric charges. Thus, for a given volume, as we see in Figure 2.2, all magnetic field lines which enter, must exit.

Ampère’s law

A first relationship between electric and magnetic fields is given by Ampère’s law⁴ which states

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (2.14)$$

for current density \mathbf{J} [A/cm²]. The integral representation takes the form of a contour integral of the magnetic field around a closed contour surrounding the flow of current, or

$$\oint_r \mathbf{H} \cdot d\mathbf{r} = \int_S \left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \right) \cdot d\mathbf{S} = I \quad (2.15)$$

for total current I [A] enclosed by the contour vector \mathbf{r} , as we see in Figure 2.3. The two terms on the right hand side of Equation 2.14 are both current densities. \mathbf{J} is current

³ Jean-Baptist Biot (1774–1862) and Felix Savart (1791–1841), both French, and both active in diverse fields, including electricity, magnetism, and acoustics, leading the latter to invent, among other things, a trapezoidal violin. The Biot-Savart law relates the static magnetic field to the distribution of current flow.

⁴ André Marie Ampère (1775–1836), one of the great French contributors to science, survived the trauma of seeing his father guillotined in the revolution to become the scientist whom Maxwell referred to as “the Newton of electricity.”

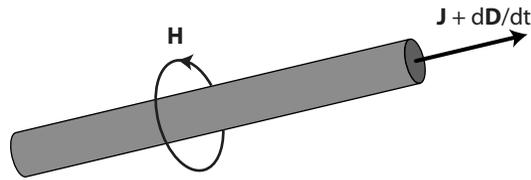


Figure 2.3 Ampère's law relates the magnetic field around a closed contour to the total current flowing through that contour.

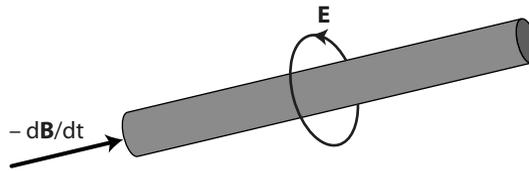


Figure 2.4 Faraday's law relates the electric field around a closed contour to the time-varying magnetic field flowing through that contour.

flow due to electric field, given by

$$\mathbf{J} = \sigma \mathbf{E} \quad (2.16)$$

for conductivity σ [$\Omega^{-1}\text{m}^{-1}$]. The differential, $\partial\mathbf{D}/\partial t$, defines a displacement current, which results from the time variation of an electric field; this latter component arises, for example, for current flow through a capacitor.

Faraday's law

Faraday's law,⁵ finally, relates the electric field to the time change of the magnetic field. In differential form,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}; \quad (2.17)$$

we see thus that an electric field may be generated not only by fixed electric charges but also by a time-varying magnetic field. The geometry may be clearer when Faraday's law is written in integral form, namely

$$\oint_r \mathbf{E} \cdot d\mathbf{r} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot d\mathbf{S} \quad (2.18)$$

from which we see, as shown schematically in Figure 2.4, that a time changing magnetic field will generate an electric field around a closed contour surrounding the magnetic field line.

⁵ Michael Faraday (1791–1867), an Englishman who made it from apprentice bookbinder to superintendent of the Royal Institution in London and subsequently discovered electromagnetic induction, diamagnetism, the magneto-optical effect and much more.

Example 2.1: Let us verify that the electric and magnetic fields sketched schematically in Figure 2.4 are indeed related in the manner shown. In this sketch, the (time varying) magnetic flux density vector \mathbf{B} is surrounded by a circulating electric field \mathbf{E} . Recalling that $\mathbf{B} = \mu\mathbf{H}$, let us take the axis in which the magnetic field points ($-\mu\frac{\partial\mathbf{H}}{\partial t}$) as the \mathbf{z} axis with unit vector \mathbf{u}_z . \mathbf{B} therefore only has a \mathbf{z} component, such that

$$\mathbf{B} = 0 \cdot \mathbf{u}_x + 0 \cdot \mathbf{u}_y + B_z \cdot \mathbf{u}_z.$$

The question we have to answer is thus: what form of \mathbf{E} will lead to \mathbf{B} with only B_z ? We can apply Faraday's law, Equation 2.17, to determine the electric field. Recalling the definition of the curl,

$$\nabla \times \mathbf{E} = \begin{vmatrix} \mathbf{u}_x & \mathbf{u}_y & \mathbf{u}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix}$$

we see that the case for which only \mathbf{u}_z is non-zero results when

$$\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= 0 \\ \frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} &= 0 \end{aligned}$$

which are satisfied if both

$$E_z = 0 \quad \text{and} \quad \frac{\partial}{\partial z} = 0.$$

We thus need an electric field which has only \mathbf{x} and \mathbf{y} components which do not vary with \mathbf{z} . By pondering a bit, we can propose that a form for \mathbf{E} which satisfies these conditions is

$$\mathbf{E} = \frac{V_0}{x^2 + y^2} (-y\mathbf{u}_x + x\mathbf{u}_y)$$

where V_0 is a numeric constant with units [V]; we see that the dimensions of \mathbf{E} are thus correct ([V/m]). The normalization in the denominator, $x^2 + y^2 = r^2$, is included so that $\mathbf{E} \propto 1/r$, decreasing as radius from the \mathbf{z} axis increases.

If we take the curl of the previous expression, we find that

$$\nabla \times \mathbf{E} = 0 \cdot \mathbf{u}_x + 0 \cdot \mathbf{u}_y + \frac{2V_0}{x^2 + y^2} \cdot \mathbf{u}_z$$

showing that, since $\nabla \times \mathbf{E}$ has only a \mathbf{z} component, \mathbf{B} also has only a \mathbf{z} component, as desired. By plotting \mathbf{E} as defined above in a vector plot (see the Problems at the end of the chapter), we find that indeed it is a rotational field centered on the time-varying magnetic field line, just as shown in Figure 2.4. We can see this roughly without plotting it by looking at the direction of \mathbf{E} for $x = \pm 1$ and $y = \pm 1$.

Table 2.1. Maxwell's equations

differential form	integral form
$\nabla \cdot \epsilon_m \mathbf{E} = \rho$	$\oint_S \epsilon_m \mathbf{E} \cdot d\mathbf{S} = \int_V \rho \cdot dV$
$\nabla \cdot \mu_m \mathbf{H} = 0$	$\oint_S \mu_m \mathbf{H} \cdot d\mathbf{S} = 0$
$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$	$\oint_r \mathbf{H} \cdot d\mathbf{r} = \int_S (\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}) \cdot d\mathbf{S}$
$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$	$\oint_r \mathbf{E} \cdot d\mathbf{r} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot d\mathbf{S}$

Table 2.2. Maxwell's equations in free space

differential form	integral form
$\nabla \cdot \epsilon_0 \mathbf{E} = 0$	$\oint_S \epsilon_0 \mathbf{E} \cdot d\mathbf{S} = 0$
$\nabla \cdot \mu_0 \mathbf{H} = 0$	$\oint_S \mu_0 \mathbf{H} \cdot d\mathbf{S} = 0$
$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}$	$\oint_r \mathbf{H} \cdot d\mathbf{r} = \frac{\partial}{\partial t} \int_S \mathbf{D} \cdot d\mathbf{S}$
$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$	$\oint_r \mathbf{E} \cdot d\mathbf{r} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot d\mathbf{S}$

Maxwell's equations

The four relationships we have just discussed are referred to collectively as Maxwell's equations⁶ and represent the basic interrelation between electric and magnetic fields as well as their spatial and time variation. Table 2.1 summarizes the two mathematical forms for these equations. The Maxwell equations form the mathematical and physical basis for all of classical electromagnetics.

In free space, meaning in the absence of charges, Maxwell's equations simplify somewhat and are given as shown in Table 2.2. As a significant subset of optical calculations are for wave propagation in free space, these simplified forms often apply.

2.1.4 The wave equation

For a description of the propagation of electromagnetic waves, we need a single relationship between the time and spatial variation of either the electric or the magnetic field; this relationship is given by the wave equation.

⁶ James Clerk Maxwell (1831–1879), Scottish physicist, was the first Cavendish Professor of Experimental Physics at Cambridge. His research as a young man allowed him to predict theoretically that light is an electromagnetic phenomenon, prompting him to laconically pen the lines which introduce this chapter.

We can easily derive the wave equation from Maxwell's equations and a bit of vector mathematics. Taking the curl of Faraday's law, Equation 2.17, we obtain

$$\nabla \times \nabla \times \mathbf{E} = -\mu_m \frac{d}{dt} (\nabla \times \mathbf{H}). \quad (2.19)$$

We can replace the $\nabla \times \mathbf{H}$ in the right hand side with Ampère's law, Equation 2.14, resulting in

$$\nabla \times \nabla \times \mathbf{E} = -\mu_m \left[\frac{d\mathbf{J}}{dt} + \frac{d^2\mathbf{D}}{dt^2} \right]. \quad (2.20)$$

Using the definitions of electric flux density and current density, as given in Equations 2.3 and 2.16, this expression becomes

$$\nabla \times \nabla \times \mathbf{E} = -\mu_m \left[\sigma \frac{d\mathbf{E}}{dt} + \epsilon_m \frac{d^2\mathbf{E}}{dt^2} \right]. \quad (2.21)$$

We appeal to an identity known from vector mathematics (Råde and Westergren, 1999), namely that for any vector \mathbf{A}

$$\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \quad (2.22)$$

such that we have

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu_m \left[\sigma \frac{d\mathbf{E}}{dt} + \epsilon_m \frac{d^2\mathbf{E}}{dt^2} \right]. \quad (2.23)$$

If we now consider the case for wave propagation in a volume with no free charges, such that $\rho = 0$ and $\sigma = 0$, resulting in $\nabla \cdot \mathbf{E} = 0$, Equation 2.23 simplifies to

$$-\nabla^2 \mathbf{E} = -\mu_m \epsilon_m \frac{d^2\mathbf{E}}{dt^2}. \quad (2.24)$$

Canceling the minus signs and realizing that we could have done this derivation for \mathbf{H} as easily as we did it for \mathbf{E} , we are left with the wave equation, which we can express in terms of the electric field

$$\nabla^2 \mathbf{E} = \mu_m \epsilon_m \frac{d^2\mathbf{E}}{dt^2} \quad (2.25)$$

or equivalently in terms of the magnetic field,

$$\nabla^2 \mathbf{H} = \mu_m \epsilon_m \frac{d^2\mathbf{H}}{dt^2}. \quad (2.26)$$

Both of the previous two expressions relate the the spatial to the time variation of the field quantities.

Examining the factor $\mu_m \epsilon_m$ in these equations more closely, we see that this product has the dimensions $1/v^2$ for velocity v [m/s], such that Equation 2.25 may be re-written as

$$\nabla^2 \mathbf{E} = \frac{1}{v^2} \frac{d^2\mathbf{E}}{dt^2}. \quad (2.27)$$

In free space, $\epsilon_m \rightarrow \epsilon_0$ and $\mu_m \rightarrow \mu_0$, and by entering the numerical values of these natural constants, it appears that

$$\sqrt{\frac{1}{\mu_0 \epsilon_0}} = c \quad (2.28)$$

for the speed of light in vacuum, $c = 299\,792\,458$ m/s. Maxwell realized this in 1873, causing him to postulate that light is an electromagnetic wave. This assertion was treated with considerable skepticism until 1888, when it was experimentally verified by Heinrich Hertz.⁷

2.2 Wave propagation

The wave equation relates the time and spatial variation of a field. Solutions to the wave equation thus represent the functions which define a propagating wave.

2.2.1 General solution to the wave equation

Examining the wave equation, Equation 2.25, we see that a second spatial derivative and a second time derivative must sum to zero. We thus expect harmonic solutions in space and time for the electric and magnetic fields. For the time-dependent electric field vector, $\mathbf{E}(\mathbf{r}, t)$ and a general three dimensional spatial vector \mathbf{r} [m], the most general solution will take the form

$$\mathbf{E}(\mathbf{r}, t) = E_0 [\cos(\pm \mathbf{k} \cdot \mathbf{r} \pm \omega t) \pm \sin(\pm \mathbf{k} \cdot \mathbf{r} \pm \omega t)] \quad (2.29)$$

or, in the more convenient exponential notation,

$$\mathbf{E}(\mathbf{r}, t) = E_0 \exp[\pm j \mathbf{k} \cdot \mathbf{r} \pm j \omega t]. \quad (2.30)$$

Since Equation 2.30 is in general complex, it is the real part which then defines a measurable wave, namely

$$\mathbf{E}(\mathbf{r}, t) = \text{Re}\{E_0 \exp[\pm j \mathbf{k} \cdot \mathbf{r} \pm j \omega t]\}. \quad (2.31)$$

In the last three expressions, E_0 [V/m] is a scalar magnitude for the electric field, ω [rad/s] is the angular frequency and \mathbf{k} [m^{-1}] is the wave vector (sometimes referred to as the propagation vector), defined as

$$\mathbf{k} = \frac{2\pi}{\lambda} \mathbf{u}_r \quad (2.32)$$

⁷ Heinrich Hertz (1857–1894), a German who spurned engineering to become professor of physics, first generated and detected radio waves in the laboratory, conclusively demonstrating that light is an electromagnetic field.

for propagation in the \mathbf{r} direction. λ [m] is the wavelength of the electromagnetic field, related to its frequency ν [Hz] by⁸

$$\lambda = \frac{c}{\nu} = \frac{2\pi c}{\omega} \quad (2.33)$$

since we remember that frequency and angular frequency are not identical but rather related as

$$\omega = 2\pi\nu. \quad (2.34)$$

The magnitude of the wave vector is then the scalar k , and is typically referred to as the propagation constant, which may then be written as

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{v} \quad (2.35)$$

for velocity v .

We thus see that the electric (and by extension the magnetic) fields propagate as sinusoidal waves in space and time, the most general solution being the sum of the individual sinusoidal solutions. The wave vector defines the propagation direction. We will discuss the relevance of this parameter in Section 2.2.3 below.

Example 2.2: Let us apply the definitions of frequency and propagation constant to the optical field emitted by a helium-neon (HeNe) laser. This laser is standard laboratory equipment in any optics lab and emits in the red, at a wavelength of $\lambda = 633$ nm. If we assume propagation in air⁹ for which $n \approx 1$, we may calculate

$$\nu = \frac{c}{\lambda} \approx 474 \text{ THz},$$

$$\omega = 2\pi\nu \approx 3 \cdot 10^{15} \text{ rad/s},$$

and

$$k = \frac{2\pi}{\lambda} \approx 9.9 \cdot 10^6 \text{ m}^{-1} = 9.9 \mu\text{m}^{-1}.$$

These are values with the typical order of magnitude for optical fields in the visible wavelength range.

⁸ We use ν rather than f for frequency, since the latter is most commonly used in optics for focal length. One should not confuse ν (frequency) with v (velocity).

⁹ For most applications, the approximation $n_{air} \approx 1$ is sufficient; a more exact value for dry air at room temperature and standard pressure is $n_{air} = 1.0002718$. The exact refractive index for air varies with pressure, temperature, humidity and composition, and comprehensive models for the variation of n_{air} have been developed (Edlén, 1966; Ciddor, 1996). For high-resolution interferometry, as will be considered in Chapter 5, the change of n_{air} as temperature varies, for example, may no longer be negligible.

2.2.2 Transverse electromagnetic waves

As presented above, the wave equation is for an arbitrary three-dimensional space. While retaining all three dimensions, we can simplify our lives by using the Cartesian¹⁰ coordinate system and taking a choice of axes for which the electric field only varies along one axis. Let us take $\mathbf{E} = \mathbf{E}(z)$ such that \mathbf{E} only varies in the z direction; \mathbf{E} is thus constant in the (x, y) plane.

If we consider Gauss' law, Equation 2.10, in free space, where $\rho = 0$, then

$$\nabla \cdot \epsilon_m \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0. \quad (2.36)$$

With the assumption that \mathbf{E} is constant in the (x, y) plane, we have $\partial E_x / \partial x = 0$ and $\partial E_y / \partial y = 0$. Since Equation 2.36 sums to zero, we are thus left with $\partial E_z / \partial z = 0$.

For this last equality to hold, two conditions are possible: 1) E_z is a constant in z , which would imply the wave does not propagate; or 2) $E_z = 0$, meaning there is no component of E in the z direction. Since we have a propagating wave, it is the latter condition which holds.

$E_z = 0$ implies that only E_x and E_y are non-zero; the only non-zero components of the electric field are thus normal (or transverse) to the propagation direction. This type of a wave, which characterizes electromagnetic propagation in a wide variety of cases, is known as a *transverse electromagnetic*, or TEM, wave. TEM waves are useful for characterizing propagation in cases of homogeneous, non-conducting materials. Longitudinal components of the electric field may be found in cases where propagation is guided, such as in fibers or waveguides.

One-dimensional wave equation

For a TEM wave, the wave equation and its solutions simplify and allow us to define a useful special case applicable in many situations. Let us take an electric field propagating in the z direction with the electric field vector pointing only in the x direction (i.e., x -polarized; see Section 2.4), $E_x(z)$. From Faraday's law, we can convince ourselves that the corresponding magnetic field also propagates in the z direction and has the magnetic field vector in the y direction, thus $H_y(z)$. The wave equation then simplifies to

$$\nabla^2 \mathbf{E}_x(z, t) = \frac{\partial^2 \mathbf{E}_x(z, t)}{\partial z^2} = \epsilon_m \mu_m \frac{\partial^2 \mathbf{E}_x(z, t)}{\partial t^2} \quad (2.37)$$

since all other spatial second derivatives are zero and we have briefly included the time-dependence of the electric field explicitly. This one dimensional wave equation has the solutions

$$\mathbf{E}_x(z, t) = E_{x0} [\cos(\pm kz \pm \omega t) \pm \sin(\pm kz \pm \omega t)] \quad (2.38)$$

¹⁰ René Descartes (1596–1650), French founder of much of modern mathematics and not an inconsiderable branch of philosophical thought, who, a life-long late riser, decided late in his career to teach the Swedish Queen mathematics in her dank, cold castle at 5 am and promptly died of pneumonia. *Cuiusvis hominis est errare.*

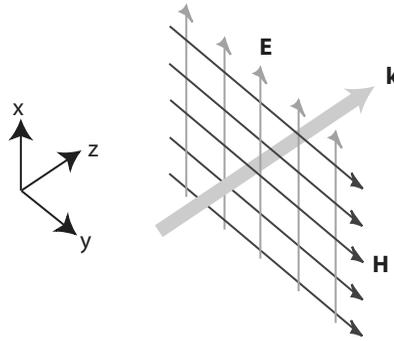


Figure 2.5 A transverse electromagnetic (TEM) wave. The electric and magnetic field components, E and H are always normal to the propagation direction, indicated by the wave vector \mathbf{k} . In this example, we have E polarized in the x direction and H polarized in the y direction.

or, in exponential notation,

$$\mathbf{E}_x(z, t) = E_{x0} \exp[\pm jkz \pm j\omega t]. \quad (2.39)$$

where we again take the real part to describe a measurable wave. The electric and magnetic field components of a one-dimensional TEM wave are shown schematically in the sketch of Figure 2.5.

In many cases, a proper choice of axes may allow the description of a propagating wave with the solutions to this one-dimensional wave equation; we will use this practical case in the discussion below.

2.2.3 Temporal and spatial frequency

The argument of the harmonic function describing a TEM wave represents a phase, ϕ [rad], thus

$$\pm kz \pm \omega t = \phi \quad (2.40)$$

with a spatial and a time variation. Taking the time derivative, we obtain

$$\frac{\partial \phi}{\partial t} = \omega \quad (2.41)$$

so that ω represents the change in phase as a function of time, thus defining a temporal frequency ω (in rad/s) or ν (in Hz = s⁻¹).

Alternatively, taking the spatial derivative of the phase, we have

$$\frac{\partial \phi}{\partial z} = k \quad (2.42)$$

implying that k gives the change in phase as a function of position, thus defining a spatial frequency (with units m⁻¹).

Whereas temporal frequency is a concept quite familiar to engineering students, spatial frequency may be new. Spatial frequency may become more clear upon examination

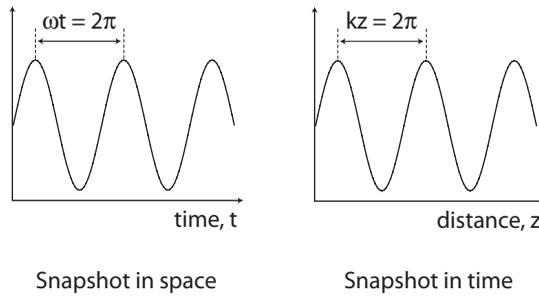


Figure 2.6 Two “snapshots” of a sinusoidally varying electromagnetic wave. By examining the variation of the wave at a fixed point in space (the snapshot in space, *left*) we see a temporal variation with a period $\omega t = 2\pi$. If we look at the wave at a fixed point in time (the snapshot in time, *right*), the spatial variation has a period $kz = 2\pi$.

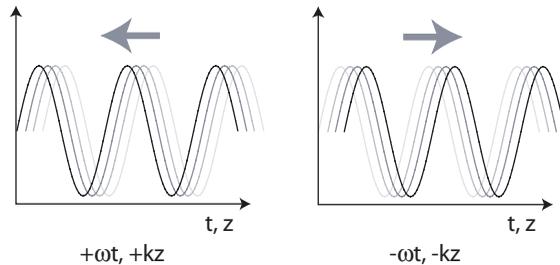


Figure 2.7 The direction of movement of an electromagnetic wave in time or space as a function of the polarity of the ωt or kz argument. The positive t or z axis is toward the right. A $+$ sign for the argument implies movement to the left and a $-$ sign, movement to the right.

of Figure 2.6, in which we see two “snapshots” of an electromagnetic wave. The snapshot in space looks at the wave at a fixed point in space, and yields a time variation with a period ωt . On the other hand, taking a snapshot at a fixed point in time shows the spatial variation of the wave, with a period kz . Thus k defines the spatial change in phase.

2.2.4 Sign conventions

The most general form of the solution to the one-dimensional wave equation is given in trigonometric form by Equation 2.38 or in exponential form by Equation 2.39; both of these include both polarities for the time and space variables. By convention, we can choose one or the other polarity and select our axes to correspond to the movement of the wave. As we see in Figure 2.7, by defining the positive time (t) or space (z) axis as pointing to the right, a $+$ sign for the ωt or kz argument implies movement to the left and a $-$ sign, movement to the right.

By convention, we will employ in this text $+\omega t$ and $-kz$; that way, waves propagate from left to right along the z axis. This definition is arbitrary, but it de-clutters our

notation a bit. Thus we will describe wave propagation with the expression, for either electric or magnetic field,

$$E, H = (E_0, H_0) e^{-jkz} e^{j\omega t}. \quad (2.43)$$

from this point forward.

2.3 Electromagnetic waves

The analysis and solution of the wave equation we undertook in Section 2.2 was a general approach, which yielded a mathematical description of a propagating wave. We now look at a few detailed aspects of such electromagnetic waves.

2.3.1 Wave velocity

From the wave phase, $\phi = \omega t - kz$, we may easily derive its velocity. The position of a maximum (or minimum) of the wave is at the point where the derivative vanishes, or

$$d\phi = \omega dt - k dz = 0. \quad (2.44)$$

The movement of this peak may then be found from this expression by taking

$$\omega dt = k dz \quad (2.45)$$

so that

$$\frac{dz}{dt} = \frac{\omega}{k} = \frac{2\pi\nu}{\frac{2\pi}{\lambda}} = \lambda\nu = v_p \quad (2.46)$$

for velocity v_p [m/s] equal to the phase velocity. The phase velocity defines the movement of a point on the wave describing the electromagnetic field, implying a continuous sinusoid of a single wavelength and of infinite extent.

For an envelope of wavelengths, we may more usefully describe the group velocity, v_g [m/s], as

$$v_g = \frac{\partial\omega}{\partial k}. \quad (2.47)$$

Only in a non-dispersive medium (vacuum, strictly) is $v_g = v_p$; in general, the two are not identical. This is of relevance when considering that information transfer takes place at group velocity, not phase velocity.

2.3.2 Refractive index

We have seen two definitions for wave velocity: in the section above, based on the nature of the harmonic function describing the wave (Equation 2.46) and in Section 2.1.4 (Equation 2.28) as a factor relating the spatial and time derivatives of the fields. Implicit in these expressions was that the wave propagates in free space in which $\epsilon_m \rightarrow \epsilon_0$ and $\mu_m \rightarrow \mu_0$.

In a medium (i.e., not in vacuum), with $\epsilon_m \neq \epsilon_0$ and $\mu_m \neq \mu_0$, the velocity of the wave is smaller than c and is given by

$$v = \sqrt{\frac{1}{\mu_m \epsilon_m}} = \sqrt{\frac{1}{\mu \mu_0 \epsilon \epsilon_0}} = \frac{c}{\sqrt{\mu \epsilon}} \quad (2.48)$$

recalling that μ and ϵ are dimensionless constants. Thus the velocity of the electromagnetic wave is reduced with respect to c by a factor $1/\sqrt{\mu \epsilon}$ and propagation of light in any material is thus slower than in vacuum.

For most optical materials, with the exception of those magnetically active and used, for example, in magneto-optics, we have $\mu \approx 1$. Using this generally applicable approximation, the velocity of an electromagnetic wave may be given by

$$v \approx \frac{c}{\sqrt{\epsilon}} = \frac{c}{n} \quad (2.49)$$

for the dimensionless constant n [] which is referred to as the refractive index of the material. The constant n may most simply be related to the dielectric constant ϵ as

$$n = \sqrt{\epsilon}. \quad (2.50)$$

As we shall see in Chapter 3, ϵ and n are in general complex, strongly frequency and direction dependent and are often defined as a tensor. However, in many well-defined special cases, refractive index may be given as a single dimensionless number, as we have done here. As we will see in the rest of this book, the refractive index of a material is of primary importance in characterizing its effect on the propagation of an optical field.

If we compare Equations 2.33 and 2.46, we see that a refractive index $n \neq 1$ has implications for the other constants we have defined. From

$$v = \frac{c}{n} = \frac{\lambda \nu}{n}, \quad (2.51)$$

and realizing that frequency ν is a fundamental quantity, we see that the wavelength in a material is reduced from its vacuum value λ_0 by

$$\lambda = \frac{\lambda_0}{n}. \quad (2.52)$$

If the wavelength is reduced, the propagation constant is augmented from its vacuum value k_0 as

$$k = \frac{2\pi}{\lambda} = \frac{2\pi n}{\lambda_0} = nk_0. \quad (2.53)$$

As we implicitly defined in the relationships above, the vacuum values for parameters are typically indicated by the subscript 0. We summarize the distinction between propagation in vacuum and in a material for the most important parameters in Table 2.3 and note parenthetically that forgetting to adapt the value for λ or k from its vacuum value when calculating in a material is one of the most common errors students are prone to make in simple calculations.

Table 2.3. Propagation in vacuum and in a medium

parameter	in vacuum	in a medium
permittivity	ϵ_0	ϵ_m
permeability	μ_0	μ_m
refractive index	1	n
speed of light	c	$v = c/n$
wavelength	λ_0	$\lambda = \lambda_0/n$
propagation constant	k_0	$k = nk_0$

2.3.3 Energy

An electromagnetic field is a form of energy. The electric and magnetic energy densities, W_E [J/m³] and W_H [J/m³] respectively, define the energy per unit volume of the field and are given by

$$W_E = \frac{1}{2}\epsilon_m E^2 \quad (2.54)$$

$$W_H = \frac{1}{2}\mu_m H^2 \quad (2.55)$$

where we note that $W_E = W_H$. Thus the electric field component and the magnetic field component both contribute equally to the total energy density of an electromagnetic wave.

Poynting vector

Electromagnetic waves propagate through space and the rate of energy transfer, or power flow, is given by the Poynting¹¹ vector \mathbf{S} [W/m²] defined as

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}. \quad (2.56)$$

Notice that \mathbf{S} is a vector quantity with direction normal to \mathbf{E} and \mathbf{H} and thus parallel to the wave vector \mathbf{k} . The Poynting vector therefore defines the direction and magnitude of power flow of an electromagnetic wave.

By definition of the impedance of a medium, Z [Ω], as

$$Z = \sqrt{\frac{\mu_m}{\epsilon_m}} \quad (2.57)$$

we can further define a scalar power density S [W/m²] of the form

$$S = \frac{1}{Z} E^2 = Z H^2. \quad (2.58)$$

¹¹ John Henry Poynting (1852–1914), an English physicist who was one of the first to consider energy transfer through space by means of fields and not the (non-existent) aether.

Since the impedance of vacuum, Z_0 [Ω] is given by

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \approx 377 \Omega \quad (2.59)$$

and is thus a natural constant, we can then write the scalar electric and magnetic power densities as

$$S = \frac{n}{Z_0} E^2 = \frac{Z_0}{n} H^2 \quad (2.60)$$

for propagation in a material with refractive index n .

Time averages

Any real measurement of the energy or power of an electromagnetic wave takes place over a certain finite time. As a result, the measured value of S is a time average of the power density, which is given by an integral over one or more periods T [s] of the wave, or

$$\langle S \rangle = \frac{1}{T} \int_0^T S(t') dt' = \frac{1}{2} (EH). \quad (2.61)$$

Thus a measurement of the power density of an electromagnetic wave ultimately yields the value

$$\langle S \rangle = \frac{n}{2Z_0} E^2 = \frac{Z_0}{2n} H^2, \quad (2.62)$$

again for propagation in a material with refractive index n .

Example 2.3: Let us consider the powers and power densities of two popular light sources: the incandescent lightbulb and the semiconductor laser.

The light bulb is an infrared-emitting black-body heater which happens to have a small portion of its spectrum in the visible wavelength range; let us optimistically take an efficiency (i.e., that portion of the spectrum in the visible) of $\eta = 5\%$. For a 100 W light bulb with a bulb diameter of 10 cm, we can calculate the power density (summing the fields of the visible part of the spectrum) over the surface of the bulb (assuming it is a sphere) to be $\langle S \rangle = 159 \text{ W/m}^2$.

From Equation 2.62, assuming a wave in free space (i.e., $n = 1$), we can find the peak electric field on the surface of the bulb to be

$$E = \sqrt{2Z_0 \langle S \rangle} \approx 346 \text{ V/m}$$

and thus the electric energy density of the integrated optical field is

$$W_E = \frac{1}{2} \epsilon_0 E^2 \approx 0.53 \text{ pJ/m}^3.$$

It is instructive to compare this value with that generated by a laser beam. Taking a small laser diode emitting 1 mW, perhaps of the type used in a laser pointer, we can easily focus this emission (all of which is at a visible red wavelength) to a spot of

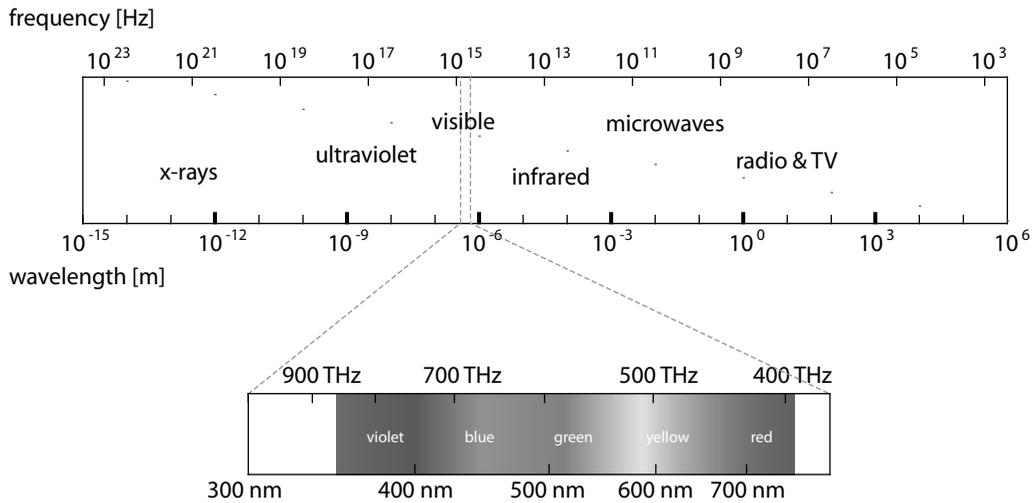


Figure 2.8 The electromagnetic spectrum given in wavelength and frequency. The visible wavelengths range from about $\lambda = 380$ nm to about $\lambda = 750$ nm; most micro-optical components employ wavelengths close to this regime, from about 300 nm to 1 700 nm

1 mm² area. The time average power density is thus $\langle S \rangle = 10^3$ W/m². Using the same calculation as above, the peak electric field and electric energy density are then

$$E = \sqrt{2Z_0 \langle S \rangle} \approx 868 \text{ V/m}$$

and

$$W_E = \frac{1}{2} \epsilon_0 E^2 \approx 3.34 \text{ pJ/m}^3.$$

We thus see that the laser diode, with a factor 10^5 less power than the light bulb, generates double the peak electric field and six times the energy density in the visible wavelength range than the light bulb. One of the many reasons that lasers are so attractive is this ability to generate high power densities.

2.3.4 Spectrum

The title of this chapter is *The physics of light*, but the analysis we have undertaken holds for electromagnetic waves in general. As we can see in Figure 2.8, the electromagnetic spectrum goes far beyond what is considered colloquially as “light”, from x-rays with wavelengths in the picometer range to long radio waves with wavelengths measured in kilometers. The mathematical description of electromagnetic waves we have developed apply for this entire wavelength range.

Historically, the field of optics was limited to the visible wavelength range, which, as we can see from the figure, extends from about $\lambda = 380$ nm in the blue to about $\lambda = 750$ nm in the red. The human eye is sensitive in this range. The relevant range of wavelengths in contemporary optics developments is now considerably larger than this, ranging from the “extreme ultraviolet”, a euphemism for x-rays with wavelengths in the 10s of nanometers range, to the mid- and far-infrared, where wavelengths reach 10s of micrometers. Beyond these ranges, one generally doesn’t refer to “optics”.

Most micro-optical components and applications are found in a more limited wavelength range, roughly from the ultraviolet (UV) around 300 nm to the near-infrared (IR), about 1 700 nm; specialized applications extend into the mid-IR, from 2 to 10 μm .

Notation

The author thinks in meters; for that reason, the position of an optical signal in the electromagnetic spectrum will be given in this text, with few exceptions, in terms of wavelength, λ . Depending on the background or field in which the reader is working, other forms of notation may be more popular; transforming on paper is usually trivial, even though thinking in other units often presents more of a psychological challenge.

From the wavelength λ , we can easily calculate the frequency ν from

$$\nu = \frac{c}{\lambda} \quad (2.63)$$

for propagation in vacuum. Thus the visible wavelengths, as can be seen in Figure 2.8, have frequencies roughly in the range $400 \text{ THz} \leq \nu \leq 800 \text{ THz}$.

When considering the energy of individual photons or electronic transitions across the bandgap of a semiconductor, for example, it is convenient to think in terms of energy. From frequency or wavelength, we may determine the energy E [J] from¹²

$$E = h\nu = \hbar\omega = \frac{hc}{\lambda} \quad (2.64)$$

for Planck’s constant¹³ $h = 6.626 \cdot 10^{-34}$ Js and $\hbar = h/2\pi$. More practical (and traditional) for electronic transitions is the energy unit electron-Volt [eV], for which $h = 4.136 \cdot 10^{-15}$ eV s. A handy relationship to remember relating wavelength in micrometers to energy in electron-Volts is

$$E [\text{eV}] = \frac{1.23985}{\lambda[\mu\text{m}]} \quad (2.65)$$

Visible wavelengths are therefore found in the energy range $3.1 \text{ eV} \geq E \geq 1.8 \text{ eV}$.

¹² The author regrets the plethora of E’s: we use the roman E for energy, the italic E for electric field and the bold roman E for electric field vector.

¹³ Max Planck (1858–1947), German physicist and, with Albert Einstein, one of the founders of quantum theory. Conservative by nature, he accepted the reality of quantized radiation that his own theory predicted during a stroll with his young son on New Year’s Day 1900, averring that classical physics had now passed away.

Some opticians, physicists in particular, like to think in inverse centimeters [cm^{-1}]. This inverse relationship with wavelength is easily calculated from

$$\lambda[\text{cm}^{-1}] = \frac{10^7}{\lambda[\text{nm}]} \quad (2.66)$$

The visible wavelength range in inverse centimeters thus encompasses about $25\,000\text{ cm}^{-1} \geq \lambda \geq 14\,300\text{ cm}^{-1}$.

Example 2.4: We can express the emission of an argon (Ar) ion laser in a number of equivalent units: as wavelength, $\lambda = 514.5\text{ nm}$; as frequency $\nu = 583\text{ THz}$; in terms of energy $3.9 \cdot 10^{-19}\text{ J}$ or 2.41 eV ; and finally as $19\,436\text{ cm}^{-1}$.

2.4 Polarization

As we saw in Section 2.2.2 above, the electromagnetic field propagating as a TEM wave has electric and magnetic fields normal to the propagation direction. The direction in which the *electric* field E points is known as the polarization direction of the field and this parameter has important implications for wave propagation.

2.4.1 Forms of polarization

The definition of the polarization state is important for many optical applications; very often, a linear polarization in a known direction is desirable. As we will see in the chapters below, some light sources (particularly lasers) are inherently polarized; others are completely unpolarized. We discuss here the three fundamental polarization states.

Linear polarization

The TEM wave shown in Figure 2.5 represents a special case in which the electric field points in the x direction (and the magnetic field in the y direction). This case corresponds to linear polarization, as is schematically depicted in Figure 2.9; the electric field points in a single, fixed direction, normal to the wave vector \mathbf{k} . The direction and length of E remain constant as the wave propagates.

Linear polarization is the simplest and in many cases most desirable case, and a superposition of two orthogonally linearly polarized waves can often be used to describe more complex polarization situations. We can explicitly include the polarization direction in the description of the electric field by including a subscript; thus $E_x(z)$ implies an x -polarized field propagating in the z direction.

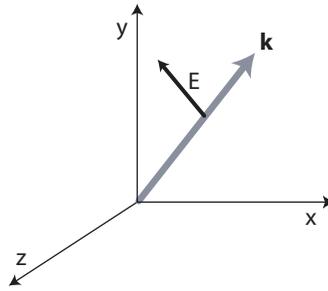


Figure 2.9 The electric field vector E for a linearly polarized electromagnetic wave points in a single fixed direction normal to the wave vector k .

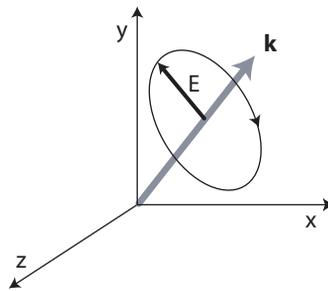


Figure 2.10 The electric field vector E for a circularly polarized electromagnetic wave is of fixed length and rotates about the axis of the wave vector k .

Circular polarization

If the electric field rotates about the propagation axis as the wave propagates, circular polarization, as shown in Figure 2.10, results. The length of E remains constant but its orientation rotates, describing a circle with k at its center.

The electric field vector can rotate in either direction. If E rotates clockwise when viewed with the wave approaching (i.e., k could poke the viewer in the eye), this condition is referred to as *left circularly polarized*; counterclockwise rotation is then by extension *right circularly polarized*. To help memorize these definitions, consider your hands: with the thumb pointing in the direction of k , i.e., toward you, the fingers curl in the rotation direction of the E -field vector. Thus the right hand curls counter-clockwise; the left hand clockwise.

Elliptical polarization

The most general case for polarization is elliptical polarization. In this case, as shown in Figure 2.11, the electric field vector rotates about k and changes its length as it does so. Thus magnitude and orientation of E vary with propagation and the tip of the E -vector therefore describes an ellipse.

Linear and circular polarization are both special cases of elliptical polarization. An ellipse with a minor axis of length zero describes a line and an ellipse with equal lengths for both axes describes a circle. We will see this more clearly mathematically below.

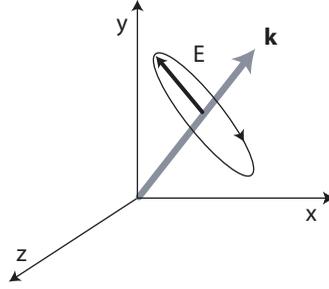


Figure 2.11 The electric field vector E for an elliptically polarized electromagnetic wave varies in length as it rotates about the axis of the wave vector \mathbf{k} .

2.4.2 The polarization ellipse

A general mathematical description of polarization is useful for describing propagation in media, particularly those which have polarization-dependent properties, such as, for example, birefringent crystals. The most general form of polarization is elliptical, and we can describe an elliptically polarized field as the superposition of two, phase shifted linearly polarized fields, namely

$$E_x = E_{x0} \sin(\omega t - kz) \mathbf{u}_x \quad (2.67)$$

$$E_y = E_{y0} \sin(\omega t - kz + \delta) \mathbf{u}_y. \quad (2.68)$$

E_x is x polarized whereas E_y is y polarized and leads E_x by a phase angle δ [rad] and the total electric field \mathbf{E}_T is then the sum of these two

$$\mathbf{E}_T = E_{x0} \sin(\omega t - kz) \mathbf{u}_x + E_{y0} \sin(\omega t - kz + \delta) \mathbf{u}_y. \quad (2.69)$$

If we take an arbitrary position in space, which we can conveniently choose to be $z = 0$, the two fields become

$$E_x = E_{x0} \sin(\omega t) \mathbf{u}_x \quad (2.70)$$

$$E_y = E_{y0} \sin(\omega t + \delta) \mathbf{u}_y \quad (2.71)$$

$$= E_{y0} (\sin \omega t \cos \delta + \cos \omega t \sin \delta) \mathbf{u}_y. \quad (2.72)$$

From Equation 2.70, we have

$$\sin(\omega t) = \frac{E_x}{E_{x0}} \quad (2.73)$$

and a bit of basic trigonometry yields

$$\cos(\omega t) = \sqrt{1 - \sin^2(\omega t)} = \sqrt{1 - \frac{E_x^2}{E_{x0}^2}}. \quad (2.74)$$

We can plug this last result into Equation 2.72 to yield

$$\frac{E_x^2}{E_{x0}^2} - \frac{2E_x E_y \cos \delta}{E_{x0} E_{y0}} + \frac{E_y^2}{E_{y0}^2} = \sin^2 \delta \quad (2.75)$$

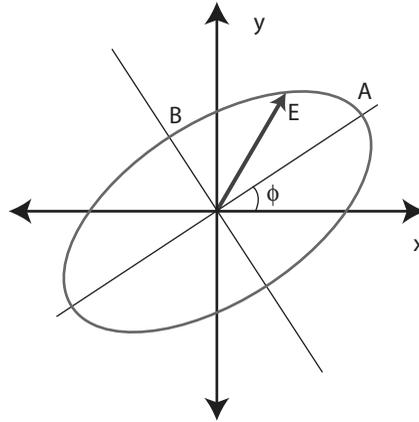


Figure 2.12 The polarization ellipse with two polarization directions x and y . The tip of the electric field vector traces the contour of the ellipse, which has major axis A and minor axis B , and is tilted at an angle of ϕ with respect to the x axis.

which we can write in a slightly more convenient form as

$$\left(\frac{1}{E_{x0}^2 \sin^2 \delta}\right) E_x^2 - \left(\frac{2 \cos \delta}{E_{x0} E_{y0} \sin^2 \delta}\right) E_x E_y + \left(\frac{1}{E_{y0}^2 \sin^2 \delta}\right) E_y^2 = 1. \quad (2.76)$$

This last equation describes an ellipse with the two axes corresponding to the two orthogonal polarization directions, E_x and E_y . As seen in Figure 2.12, the ellipse is defined by the trace of the total electric field vector, \mathbf{E}_T , where A and B represent the major and minor axes, respectively. The tilt of the ellipse with respect to the E_x axis is given by an angle ϕ , where

$$\phi = \frac{1}{2} \arctan \left(\frac{2E_{x0} E_{y0} \cos \delta}{E_{x0}^2 - E_{y0}^2} \right). \quad (2.77)$$

From this general representation, we can derive a few useful special cases. In the case where $E_{x0} = 0$ or $E_{y0} = 0$, we have linear polarization in the y or x directions, respectively. If $E_{x0} = E_{y0}$ but $\delta = 0$, the wave is also linearly polarized but at an angle $\phi = 45^\circ$ with the E_x axis. Finally, for $E_{x0} = E_{y0}$ and $\delta = \pm 90^\circ$, circular polarization results. If $\delta = +90^\circ$, we have *left-handed* circular polarization and when $\delta = -90^\circ$, *right-handed* circular polarization. We recall the definition of left- and right-handedness above.

There are more complete (and complex) means for representing polarization; premier among these is the Poincaré¹⁴ sphere (Iizuka, 2002a, Chapter 7). Useful for describing the change in polarization due to propagation in a material or through a polarizer or

¹⁴ Jules Henri Poincaré (1854–1912), born in Nancy, France, as member of a prominent French family which produced prime ministers and presidents, made a wide variety of fundamental contributions in mathematics and geometry.

polarizing medium are Jones¹⁵ matrices (Jones, 1941)(Yariv, 1991, Chapter 1.5). For optical problems, in micro- or macro-optics, for which polarization plays an important role, these mathematical approaches are of significant value.

Example 2.5: Given an electric field with x and y components

$$\mathbf{E} = E_{x0} \sin(\omega t - kz + \delta_x) \mathbf{u}_x + E_{y0} \sin(\omega t - kz + \delta_y) \mathbf{u}_y,$$

what is the polarization state if $E_{x0} = E_{y0}$, $\delta_x = 0$ and $\delta_y = \pi/2$?

Recalling that $\sin(\phi + \pi/2) = \cos(\phi)$ and defining $E_0 = E_{x0} = E_{y0}$, the field is then

$$\mathbf{E} = E_0 [\sin(\omega t - kz) \mathbf{u}_x + \cos(\omega t - kz) \mathbf{u}_y].$$

Taking the magnitude of \mathbf{E} , we find

$$|\mathbf{E}| = E_0 \sqrt{\sin^2(\omega t - kz) + \cos^2(\omega t - kz)} = E_0$$

which is a constant. Examining the equation describing \mathbf{E} above, we see that the y component (the cosine) *leads* the x component (the sine), so that, using a right-handed coordinate system (positive x to the right, positive y upwards, positive z out of the page toward you), the expression for \mathbf{E} describes clockwise motion. Since the magnitude of \mathbf{E} is constant, the electric field vector thus traces a clockwise circle for a wave moving toward us (along the positive z axis), which, using the definitions above, corresponds to *left circular* polarization.

2.5 Wavefronts

In Section 2.2.2 above, we considered the special case of what we termed a “one-dimensional wave” meaning one for which the axis of propagation was chosen to be coincident with, for example, the z axis. In general, as shown in Figure 2.13, the wave can propagate in any arbitrary direction, such that the general three-dimensional form of the wave equation and its solution, Equation 2.31, must be employed. Thus for the vector \mathbf{r} , whose length in the Cartesian coordinate system is given by

$$|\mathbf{r}| = \sqrt{|\mathbf{x}|^2 + |\mathbf{y}|^2 + |\mathbf{z}|^2}, \quad (2.78)$$

we have the general form

$$\mathbf{E}(\mathbf{r}, t) = E_0 e^{-j(\mathbf{k}_x \cdot \mathbf{x} + \mathbf{k}_y \cdot \mathbf{y} + \mathbf{k}_z \cdot \mathbf{z})} e^{j\omega t} = E_0 e^{-j\mathbf{k} \cdot \mathbf{r}} e^{j\omega t}. \quad (2.79)$$

¹⁵ R. Clark Jones (1916–2004) was an American researcher in optics who spent most of his career at the Polaroid Corporation in Cambridge, Massachusetts.

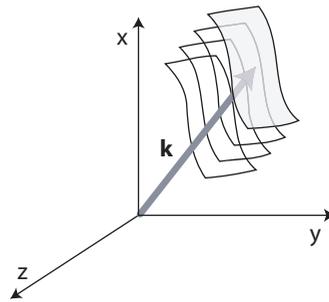


Figure 2.13 A wave vector \mathbf{k} pointing in an arbitrary direction in the Cartesian coordinate system. Normal to the wave vector are the surfaces of constant phase (i.e., constant \mathbf{k}), which form the wavefront of the wave.

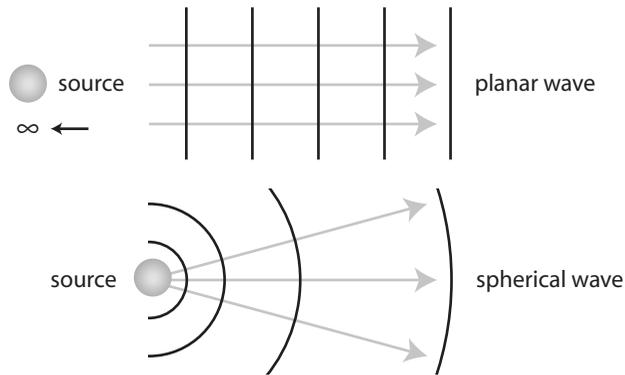


Figure 2.14 Two-dimensional cross sections of two limiting cases for the wavefront; the wavefronts are shown in black and the wave vectors are in gray. A planar wavefront *top* is generated by a source located at infinity and the spherical wavefront *bottom* is emitted by a point source. For distances far from the source, the spherical wavefront becomes approximately planar.

For a propagating wave, regions of constant \mathbf{k} form surfaces normal to the wave vector; this surface is called the wavefront. The form of the wavefront is an essential parameter for characterizing the propagation of a wave, particularly as it moves through media and optical components.

2.5.1 Spherical and planar wavefronts

As Figure 2.13 suggests, the wavefront of a propagating wave is given by the surface of constant \mathbf{k} . Two limiting cases for the wavefront shape, shown in the schematic two-dimensional representations of Figure 2.14, are useful in a wide variety of optical situations: 1) the planar wavefront, which results when the source is infinitely far away and all the wave vectors are parallel; and 2) the spherical wavefront, obtained for emission from and close to a point source, for which the wave vectors propagate radially from the center.

A planar wave is given mathematically by

$$E(z) = E_0 e^{-jkz} \quad (2.80)$$

for the case where propagation is along the z axis. The planar wavefront which results often represents a desirable propagation condition. For this case, since all the wave vectors are parallel, we can imagine a bundle of these propagating close to the propagation axis and can thus speak of a light “beam”, a spatially localized electromagnetic field which propagates in a well-defined direction. A light beam with planar wavefronts is referred to as collimated, and collimation is often an early step in preparing a field for propagation through an optical system. We try to use collimated beams wherever possible, even though perfect collimation (i.e., a perfectly planar wavefront) is not practically achievable; we will see this more clearly in Section 2.6 below.

The spherical wavefront, also shown in Figure 2.14, is generated around a point source which emits uniformly in all directions and is given by

$$E(r) = \frac{E_0}{r} e^{-jkr}. \quad (2.81)$$

The surfaces of constant phase form concentric spheres with the source at its center. The spherical wavefront often represents a good approximation for emission from a small source, such as an LED, for which this representation is again useful for a simple calculation of the propagating field.

The plane and spherical waves are two extreme cases. In a plane wave, the electromagnetic energy is distributed evenly over the wavefront and remains constant with propagation distance. The spherical wave, on the other hand, has diverging wave vectors emerging from an ideal point and the field strength decreases with $1/r$ as the wavefront propagates. As is indicated in the figure, a spherical wavefront slowly evolves into a planar one as the field moves further and further from the source; for the (mathematically) limiting case of infinite distance from the source, the wavefront becomes ideally planar; luckily this limit can often be reached to a good approximation in the confines of the typical (non-infinite) university optics laboratory.

2.5.2 Wavefront characterization

When a wavefront has been transmitted through an optical system, it is typically distorted from its ideal form. We can consider two examples, at extreme size scales: in the eye, the non-ideal shape of the lens can result in aberrations such as astigmatism or defocus. Alternatively, after millions of undistorted years, the light emitted from a star (as close to a source at infinity as we can get) has its planar wavefront deformed in the last nanoseconds of its life when passing through the non-uniform atmosphere of the earth. In both cases, the wavefront is neither ideally spherical nor planar and a more comprehensive means to describe it is required.

If the wavefront is continuous and reasonably smooth, it can often be represented mathematically by a polynomial expansion over its surface. Numerous approaches for accomplishing this have been proposed and applied (Malacara and DeVore, 1992;

Wyant and Creath, 1992), and they have in common that the shape of the wavefront is approximated by summing an appropriately chosen set of two-dimensional polynomial expressions. Experimentally, the mathematical representation of a measured wavefront is usually determined by performing a least-squares fit of the wavefront data with such a polynomial sum.

Zernike polynomials

One of the most established and probably the most widely employed wavefront representation technique is that of Zernike¹⁶ polynomials (Zernike, 1934; Born and Wolf, 1980; Thibos et al., 2002). These are polynomial functions, in general complex, which are orthogonal over the unit circle. Since many optical systems are circularly symmetric, such functions are easily applicable and as a result polar coordinates are employed. In the discussion which follows, we will use (r, θ) , where θ is defined as increasing counterclockwise from the x axis.

The Zernike polynomials with indices (m, n) are defined as

$$Z_n^m(r, \theta) = N_n^m R_n^m(r) e^{jm\theta} \quad (2.82)$$

thus having a radial (r -dependent) component R_n^m and an azimuthal (θ -dependent) component, $e^{jm\theta}$; the term N_n^m is a normalization factor. The indices n and m define the maximum order of the polynomial and the azimuthal frequency, respectively; for a given n , we have $m = -n, -n + 2, -n + 4, \dots, n$ such that $|n - m|$ is always even. Each Zernike polynomial is thus uniquely defined by its indices, which is reflected in the notation Z_n^m .

The harmonic nature of the azimuthal components is more clearly seen if we express Z_n^m in a purely real form, namely

$$Z_n^m(r, \theta) = \begin{cases} N_n^m R_n^m(r) \cos(m\theta) & : m \geq 0 \\ -N_n^m R_n^m(r) \sin(m\theta) & : m < 0. \end{cases} \quad (2.83)$$

We then define the radial function $R_n^m(r)$ as

$$R_n^m(r) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} r^{n-2s} \quad (2.84)$$

and the normalization factor as

$$N_n^m = \sqrt{\frac{2(n+1)}{1 + \delta_{m0}}} \quad (2.85)$$

¹⁶ Frits Zernike (1888–1966), prodigy of an illustrious Dutch family, was professor of mathematical physics in Groningen, turning to optics relatively late in his career. He developed the concept of phase contrast microscopy, which even the optics firm Zeiss ignored until years afterwards, and was awarded the Nobel prize for physics in 1953.

Table 2.4. An indexing scheme for designating Zernike polynomials using a single index, k , for allowed values of (n, m) . Not all workers in the field use the same numbering.

		m										
		-5	-4	-3	-2	-1	0	1	2	3	4	5
n	0						0					
	1					1	2					
	2				3	4	5					
	3			6	7	8	9					
	4		10	11	12	13	14					

where δ_{m0} is the Kronecker¹⁷ delta function, for which $\delta_{m0} = 1$ if $m = 0$ and $\delta_{m0} = 0$ otherwise. Equations 2.83–2.85 thus define a set of r and θ -dependent polynomials over the unit circle which characterize a certain deformation of the wavefront. Through a weighted sum of these polynomials, arbitrary wavefront deformations may often be approximated.

In Figure 2.15, examples of the calculated surfaces defined by eight Zernike polynomials are shown. We see that these lower order functions define several basic wavefront aberrations, such as tilt or defocus; we will study these aberrations in detail in Chapter 7. Although by weighting and adding a number of these, and possibly higher order, Zernike polynomials, many types of arbitrary wavefront distortions can be mathematically described, we note that not all wavefront shapes are necessarily representable by a suitable sum of Zernikes. For rapidly varying or discontinuous wavefronts, or those generated in particular situations (such as, for example, by transmission through certain types of inhomogeneous media), a sum of Zernike polynomials may not provide a suitable approximation (Wyant and Creath, 1992).

Since the designation of Zernike polynomials using the indices n, m is occasionally tedious, the functions are frequently designated using a single index. In Table 2.4 we show one such numbering scheme. The parameter k , found in the table for a given (n, m) defines a “Zernike number”, so that we could, for example, write $Z_7(r, \theta)$ instead of $Z_3^{-1}(r, \theta)$. Whereas such simplified notation is widely used, we caution the reader that not everyone uses the same numbering standards; the safest notation is still $Z_n^m(r, \theta)$, writing the indices (n, m) explicitly.

Example 2.6: We calculate and plot the Zernike coefficient Z_2^2 . Using Equations 2.83–2.85 and either a sharp pencil or our favorite mathematics software, we plug in $m = n = 2$ and can easily determine

$$R_n^m(r) = r^2$$

¹⁷ Leopold Kronecker (1823–1891), a German logician who believed that mathematics should only employ integers.

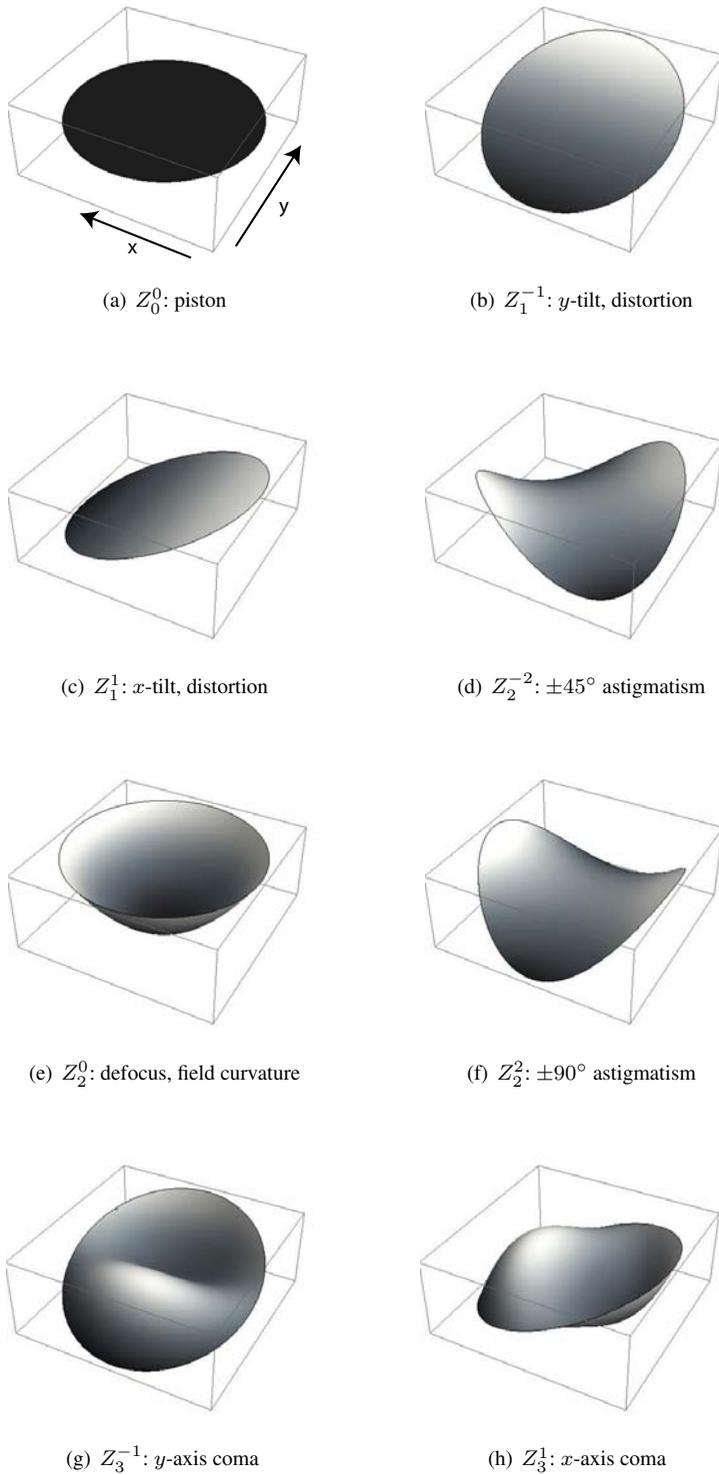


Figure 2.15 Plots of the Zernike functions, Z_n^m , for various coefficients, illustrating the basic wavefront aberrations. The x -axis is at the front; the y -axis extends front-to-back.

$$N_n^m = \sqrt{6}$$

which then yields, using 2.83,

$$Z_2^2 = \sqrt{6}r^2 \cos \theta.$$

If we plot this expression, given in radial coordinates, we obtain the characteristic shown in Figure 2.15(f). This shape of wavefront is known as astigmatism, in which the focus is different along two orthogonal axes. We will discuss this aberration in detail in Chapter 7.

2.6 Gaussian beams

The spherical and planar wavefronts discussed in Section 2.5.1 are ideal, limiting cases and may be used as useful approximations in a variety of situations. Alternatively, the very general wavefront representation just presented in Section 2.5.2, is useful for describing arbitrarily distorted wavefronts. For an undistorted wave propagating close to an optical axis, known as the paraxial condition, the lateral field distribution is best described by yet another representation, the Gaussian beam. The Gaussian intensity distribution applies for the case where the wave vectors are close to and roughly parallel to the z axis, and we will see that the planar and spherical wavefronts are special cases which result from this wavefront description.

2.6.1 Helmholtz equation

The Gaussian beam defines the spatial variation of a paraxial beam and is derived from a solution of the Helmholtz equation¹⁸. The Helmholtz equation applies to a variety of physical problems, and defines the spatial variation of functions which are solutions to time and spatially dependent partial differential equations, such as the wave equation.

Beginning with the three dimensional wave equation of Equation 2.27, we plug in a solution of the form given by Equation 2.79, for which we can write the spatial dependence as $E(r)$, thus

$$E(r, t) = E(r)e^{j\omega t}, \quad (2.86)$$

to yield

$$\nabla^2 E(r)e^{j\omega t} = \frac{1}{v^2}(-\omega^2)E(r)e^{j\omega t}. \quad (2.87)$$

¹⁸ Hermann von Helmholtz (1821–1894), a German with a broad spectrum of interests ranging from physiological optics and acoustics to mathematics, medicine, music and electrodynamics.

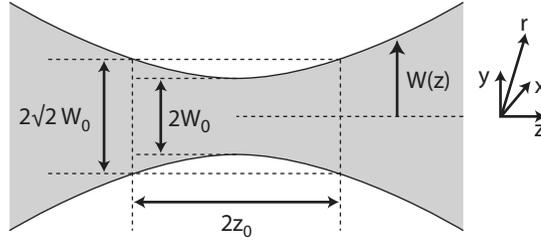


Figure 2.16 Schematic sketch of the cross-section of a Gaussian beam.

Dividing out the $e^{j\omega t}$ terms, we obtain the Helmholtz equation, namely

$$\nabla^2 E(r) + \frac{\omega^2}{v^2} E(r) = 0 \quad (2.88)$$

or alternatively

$$\nabla^2 E(r) + k^2 E(r) = 0. \quad (2.89)$$

As we can see, the Helmholtz equation is an eigenvalue equation, a partial differential equation in r , and thus defines the spatial distribution of $E(r)$. Solutions include planar, spherical, paraboloidal and Gaussian waves.

2.6.2 Gaussian beam

A special case for the solution to the Helmholtz equation may be derived for the case of paraxial waves propagating in the z direction of the form

$$E(r) = E_0 \psi(r) e^{-jkz}, \quad (2.90)$$

where $\psi(r)$ is an envelope function which varies slowly with z . In this case we have

$$\frac{\partial^2 \psi}{\partial z^2} \approx 0 \quad (2.91)$$

and plugging $E(r)$ into the Helmholtz equation using this last condition yields

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} - j2k \frac{\partial \psi}{\partial z} = 0. \quad (2.92)$$

The first two differential terms, $(\partial^2 \psi / \partial x^2 + \partial^2 \psi / \partial y^2)$, are known as the transverse Laplacian operator.

There are numerous solutions to the paraxial Helmholtz equation, Equation 2.92. In addition to the spherical wave, Equation 2.81, of interest is also the paraboloidal wave, given by

$$\psi(r) = \psi_0 \frac{1}{z} \exp \left[-jk \left(\frac{x^2 + y^2}{2z} \right) \right] \quad (2.93)$$

where ψ_0 is a constant and the paraboloidal wavefront is one for which the phase on a contour defined by $(x^2 + y^2)/z$ is a constant, that contour defining a paraboloid of

Table 2.5. Parameters defining a Gaussian beam

parameter	symbol	definition
Rayleigh range	z_0	
minimum beam width	$2W_0$	
depth of focus	$2z_0$	kW_0^2
beam width	$2W(z)$	$2W_0\sqrt{1 + \left(\frac{z}{z_0}\right)^2}$
radius of curvature	$R(z)$	$z\left[1 + \left(\frac{z_0}{z}\right)^2\right]$
Gouy phase shift	$\phi(z)$	$\arctan\left(\frac{2z}{kW_0^2}\right)$

revolution. The paraboloidal wavefront is an approximation to the spherical wavefront for the paraxial case, far from the source. As we move along the z axis, the wavefront emitted from a point source is always spherical but as the radius increases, it becomes approximately paraboloidal and ultimately, when we move far enough away, approximately planar.

The most useful solution to Equation 2.92 is that which defines the Gaussian beam, which takes the form

$$\psi(r) = \psi_0 \left[\frac{W_0}{W(z)} \right] \exp \left[-\frac{r^2}{W^2(z)} \right] \exp \left[jk \left(z + \frac{r^2}{2R(z)} \right) \right] e^{-j\phi(z)} \quad (2.94)$$

where we summarize the definitions of the following parameters:

- propagation direction z ;
- radial parameter $r = \sqrt{x^2 + y^2}$;
- beam half-width $W(z)$ [m], with its minimum value W_0 [m];
- beam radius of curvature $R(z)$ [m]; and
- the Gouy¹⁹ phase shift $\phi(z)$ [rad].

In Equation 2.94, the first term in brackets describes the transverse (in x and y) contraction, the second term the transverse amplitude and the third term the form of the wavefront; the final exponential defines a correction factor or phase shift.

Beam parameters

The sketch in Figure 2.16 shows a schematic cross section of a Gaussian beam as defined by Equation 2.94 and the definition of the beam width $2W(z)$. We see that the Gaussian beam has a minimum width, $2W_0$, at the beam waist and that it expands along the z -axis as we move away from the waist. The range along the z -axis in which the beam width expands to a value $2\sqrt{2}W_0$ is given by $2z_0$, which is defined as the depth of focus (sometimes “depth of field”, and often abbreviated DOF), and is given by

$$2z_0 = kW_0^2. \quad (2.95)$$

¹⁹ Louis Georges Gouy (1854–1926), a French physicist who worked extensively on diffraction, the velocity of light, electromagnetics and Brownian motion.

The parameter z_0 [m] is referred to as the Rayleigh range²⁰ and defines the region over which the beam has a minimum width, and thus represents the “focus” of a beam.

The variation in the width $W(z)$ as a function of position along the z -axis is given as

$$W(z)^2 = W_0^2 \left[1 + \left(\frac{z}{z_0} \right)^2 \right] \quad (2.96)$$

where $z = 0$ is defined at the beam waist where $W(z) = W_0$.

Finally, we can define the radius of curvature of the beam, which describes the shape of the local wavefront, as

$$R(z) = z \left[1 + \left(\frac{z_0}{z} \right)^2 \right] \quad (2.97)$$

and the correction factor, or Gouy phase shift, as

$$\phi(z) = \arctan \left(\frac{2z}{kW_0^2} \right) = \arctan \left(\frac{z}{z_0} \right). \quad (2.98)$$

The Gouy phase shift represents a shift in the phase of a Gaussian beam when compared to that of a plane wave of the same frequency. The total phase shift of a Gaussian beam from far field to far field, through the focus, is exactly $\Delta\phi = \pi$.

Using the parameters introduced above, and summarized in Table 2.5, the shape of the Gaussian beam as it propagates from a source is defined at any point in space.

2.6.3 Gaussian characteristics

From the previous discussion, we see that the wavefront for the propagation of an optical field along an axis (that which we informally refer to as an optical “beam”) as described by the Gaussian beam can vary from a plane wave to a spherical wave and back again. We have furthermore seen that perfect collimation (an ideally planar wavefront) is not possible in a finite system and that perfect focus (a mathematical point, implying $W_0 = 0$ is likewise not achievable.

We see thus that any approximation we may wish to employ for the describing a Gaussian beam depends on where we look at it. As is indicated in Figure 2.17, the wavefront of the beam is exactly planar exactly at the beam waist; for $z = 0$, Equation 2.97 shows that $R(0) \rightarrow \infty$. As we propagate from the waist, thus in the range $z \approx z_0$, the radius decreases and the wavefront becomes spherical. Finally, as we saw in Figure 2.14, the spherical wavefront is well approximated by a planar wavefront again as we move infinitely far away from the waist (which we can take to be the “source” of the wave), so for $z \rightarrow \infty$, $R \rightarrow \infty$ again. We summarize the limiting values for beam

²⁰ Lord Rayleigh (1842–1919), whose civil name was John William Strutt, represented one of the few members of the higher English nobility who pursued a successful career in science. His research covered almost all aspects of physics, with a particular emphasis on optics and electromagnetism. Maxwell’s successor at Cambridge, he left the university to perform experiments in the peace of his estate and was awarded the Nobel Prize for physics in 1904.

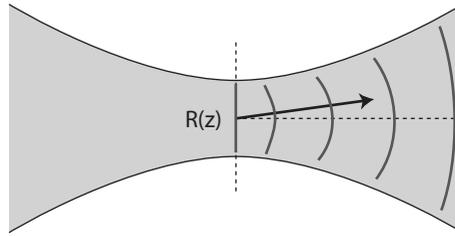


Figure 2.17 The wavefront of a Gaussian beam varies with propagation; it is planar at the beam waist, becomes spherical as it moves away from the waist and is approximately planar again as the wave approaches infinity.

Table 2.6. Limiting values for W , R and ϕ

	$z = 0$	$z = z_0$	$z \gg z_0$
$W(z)$	W_0	$\sqrt{2}W_0$	$\sim z$
$R(z)$	∞	$2z_0$	$z + z_0^2/z$
$\phi(z)$	0	$\pi/4$	$\pi/2$

width, radius and Gouy phase shift for various values of longitudinal position Z in Table 2.6.

Intensity

As the Gaussian beam propagates, through the beam waist off to infinity, the intensity distribution widens. Given an electric field with a radial distribution $\psi(r)$ as in Equation 2.94, we see that the Gaussian shape widens as the wave moves away from the beam waist, shown schematically in Figure 2.18. Since the Poynting vector, a measure for the energy density, is proportional to the square of the field,

$$S \propto \psi^2, \quad (2.99)$$

we may write the intensity distribution as a function of longitudinal position z and radius from the optical axis r as

$$S = S_0 \left[\frac{W_0}{W(z)} \right]^2 \exp(2r^2/W^2(z)). \quad (2.100)$$

It is thus clear that, whereas the total energy is of course conserved, the energy density of the beam decreases rapidly as it moves away from the beam waist minimum.

DOF and beam waist

We see from the relationships given in Table 2.5 that beam waist width, $2W_0$ and depth of focus, $2z_0$, are interrelated. We cannot independently set a certain desired beam waist width and an arbitrary corresponding depth of focus; a small value for W_0 gives a small value for z_0 and vice versa. The third important parameter which plays a role in this relationship is the wavelength. Since the wavevector k is present in Equation 2.95, we

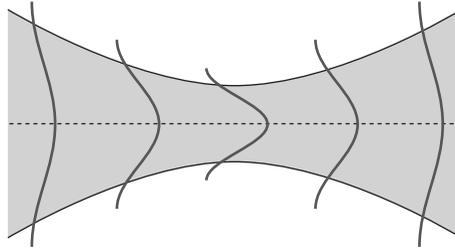


Figure 2.18 Schematic representation of the mode cross-section of a Gaussian beam. The energy density in the beam decreases as it widens.

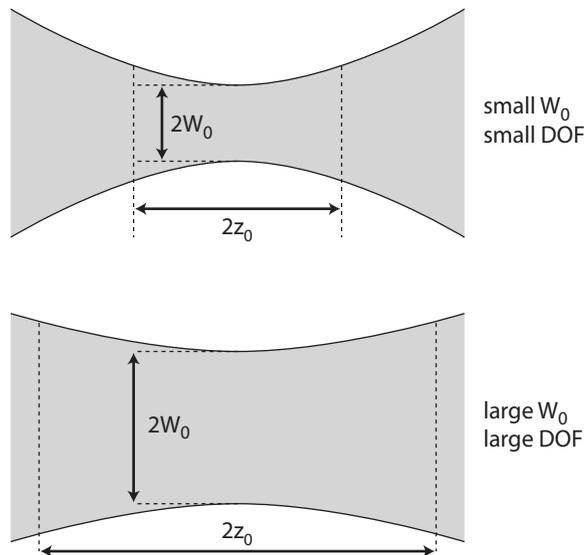


Figure 2.19 The interrelationship between beam waist width and depth of focus. A focussed beam *top* has a small beam waist width but concomitantly a small DOF; an approximately collimated beam *bottom* has a large DOF, but also a large beam width.

can see the role of wavelength explicitly from

$$z_0 = \frac{1}{2}kW_0^2 = \frac{\pi n}{\lambda}W_0^2 \quad (2.101)$$

where n is the refractive index of the transmission medium. Thus knowing any two of the parameters beam waist width, DOF or wavelength, the third is determined.

As a result, and as we see in Figure 2.19, a focussed Gaussian beam, with a small beam waist width, also necessarily has a small depth of focus; thus the beam width increases rapidly as we move longitudinally from the focal point. Likewise, if a collimated beam is desired, we require a large DOF (ideally infinite); this condition can only be achieved with a correspondingly large beam width. The upshot of these considerations is that a well-collimated Gaussian beam, from a laser for example, is not perfectly collimated (the beam width varies, albeit with a large DOF, slowly) and better

collimation is achievable with a large diameter beam. The optical “beam” which we like to draw as a long pencil-thin line is thus an artistic abstraction, not achievable in reality.

Example 2.7: We have a helium-cadmium (HeCd) laser emitting at $\lambda = 442$ nm with an aperture of $2W_0 = 1$ mm and would like to determine the spot size and beam radius of curvature of the laser emission, assuming it is a perfect Gaussian beam, on a piece of paper 10 cm from the aperture, on a wall 5 m from the laser and on the surface of the moon, 3.5×10^8 m away. Let’s also estimate the emission angle.

We can directly apply the relationships given in Table 2.5. Assuming transmission through air or vacuum with $n = 1$, we employ $k = 2\pi/\lambda$ to determine a DOF of $2z_0 = 3.55$ m. Employing the expression for $W(z)$, we find that on the paper, the spot diameter is hardly larger than the aperture, $2W_{10\text{cm}} = 1.002$ mm; the radius of curvature, $R_{10\text{cm}} = 31.7$ m, is reasonably large.

On the wall (proper laser safety precautions having been taken), the spot is still quite small, with a diameter of $2W_{5\text{m}} = 2.986$ mm, but the radius of curvature is reduced, $R_{5\text{m}} = 5.6$ m. Thus the wavefront is less planar and more spherical 5 m from the laser than close to the aperture.

On the moon (ignoring clouds, atmospheric turbulence and an occasional roasted pigeon), the beam width and thus spot diameter is $2W_{\text{moon}} = 197$ km and the radius of curvature roughly equal to the laser moon distance, $R_{\text{moon}} = 3.5 \times 10^8$ m. We can thus consider the wavefront to be planar, but the “spot” is the size of a smaller European country.

Finally, considering the emission angle, if we are far from the laser ($z \gg z_0$), we can approximate

$$\theta = \arctan \left[\frac{2W(z)}{z} \right]$$

from which we find $\theta \approx 0.064^\circ$.

Problems

1. Maxwell’s equations are not completely symmetrical. Explain why.
2. Show the equivalence of the differential and integral forms of Maxwell’s equations. You may find it of value to recall Stoke’s and Gauss’ theorems from vector calculus.
3. Given a vector field

$$\mathbf{A} = A_0 \frac{x\mathbf{u}_x + y\mathbf{u}_y + z\mathbf{u}_z}{\sqrt{x^2 + y^2 + z^2}}$$

sketch a vector plot and calculate $\nabla \cdot \mathbf{A}$ and $\nabla \times \mathbf{A}$.

4. Given a vector field

$$\mathbf{B} = B_0 \frac{(x - y)\mathbf{u}_x + (x + y)\mathbf{u}_y}{\sqrt{x^2 + y^2}}$$

- sketch a vector plot and calculate $\nabla \cdot \mathbf{B}$ and $\nabla \times \mathbf{B}$.
5. Using Faraday's law, show that a TEM wave with $E_x(z)$ has a magnetic field component of only $H_y(z)$.
 6. Take the curl of the electric field vector defined in Example 2.1 and verify that indeed it only has a z component.
 7. Sketch a vector plot of the electric field vector given in Example 2.1 and confirm that it is a rotational field. How does the vector length change with radius?
 8. Derive the wave equation using the magnetic field rather than the electric field.
 9. Show that $E = E_0 \cos[\omega t - kz + \delta]$ is a solution to the wave equation.
 10. Is $E = E_1 \cos[\omega t - kz + \delta] + E_2 \sin[\omega t - kz + \phi]$ also a solution to the wave equation?
 11. For an electromagnetic wave of the form $E_x(z, t) = E_{x0} \exp[\pm kz \pm \omega t]$ and a t or z axis with positive values pointing to the right, show that a $+$ sign for the ωt or kz argument implies movement to the left and a $-$ sign, movement to the right.
 12. Given an electromagnetic wave with $E_0 = 10 \text{ V/m}$ moving in a medium with $\epsilon = 3$ and $\mu = 1$, calculate the impedance of the medium, S and $\langle S \rangle$, the peak magnetic field and the velocity of the wave.
 13. A radio transmitter on local mountain peak 15 km from your laboratory emits a 50 kW signal isotropically. What is the peak electric field you measure in your lab on a sunny day (when the air has $n = 1.00$)? Assuming a dense winter fog gives the air an effective refractive index of $n = 1.05$, how much longer does the signal take for the trip on a foggy day?
 14. A femtosecond laser system emits 1 mJ pulses with a pulse length of 120 fs at a repeat rate of 1 kHz. Calculate the power per pulse and the emitted power averaged over 1 s. How many photons are in each pulse and how many wavelengths long (in space) is each pulse?
 15. Express the emission from a telecommunications laser diode with $\lambda = 1.55 \mu\text{m}$ in frequency, angular frequency, energy and inverse centimeters.
 16. Repeat the calculation of Example 2.5 but using the expressions which define the polarization ellipse, Equations 2.76 and 2.77.
 17. Given an electric field with x and y components

$$\mathbf{E} = E_{x0} \sin(\omega t - kz + \delta_x) \mathbf{u}_x + E_{y0} \sin(\omega t - kz + \delta_y) \mathbf{u}_y,$$

determine the polarization state for the following cases:

- a. $E_{x0} = -E_{y0}$, $\delta_x = -\pi/2$ and $\delta_y = \pi/2$
 - b. $E_{x0} = E_{y0}$, $\delta_x = 0$ and $\delta_y = -\pi/2$
 - c. $E_{x0} = 2E_{y0}$, $\delta_x = 0$ and $\delta_y = -\pi/2$
 - d. $E_{x0} = E_{y0}$, $\delta_x = 0$ and $\delta_y = \pi/4$
18. Given an electric field of the form

$$\mathbf{E} = 0 \cdot \mathbf{u}_x + \pi \cos \left[2.5\pi \cdot 10^{14} \left(t - \frac{x}{c} \right) + \frac{\pi}{2} \right] \mathbf{u}_y + 0 \cdot \mathbf{u}_z$$

what are the frequency, wavelength, direction of motion, amplitude, initial phase angle, and polarization of the wave?

19. Show that the expression for a planar wave, $E(z) = E_0 e^{-jkz}$ represents a solution to the wave equation.
20. Show that the expression for a spherical wave, $E(r) = \frac{E_0}{r} e^{-jk r}$ represents a solution to the wave equation.
21. You are given a point source emitting at 633 nm. How far do you have to go from the source before the deviation of the wavefront from a plane over a circular area with a diameter of 1 cm is less than $\lambda/10$?
22. Calculate and plot the Zernike coefficients Z_2^0 and Z_3^1 . Then plot the sum $Z_2^0 + Z_3^1$. Does this wavefront look easy to correct?
23. You have an Ar ion laser emitting at $\lambda = 514.4$ nm and an aperture of 1 mm from which a Gaussian beam is emitted; the beam waist minimum is assumed to be at the aperture. What is the DOF and how wide is the beam z_0 from the aperture? What can you change to obtain a DOF of 50 m and what is the beam width then? Is it possible to have a DOF of 50 m and a maximum beam width (z_0 from the aperture) of 3 mm?

3 Optical materials

*Le cuisinier sortira une cuisine différente lorsqu'il aura une connaissance de sa matière première.*¹

Alain Passard, 2007

Most optics experiments do not take place in vacuum and the effect of the medium on the propagation of an electromagnetic wave is of fundamental importance in designing an optical component. The behavior of optical components is strongly defined by the materials from which they are made and, whereas the shape of the interface or interfaces is also of primary importance, material characteristics play an equally important role in defining what happens to a light wave impinging on an optical structure.

We begin this section by discussing the means by which we characterize materials optically. A key concept here is the refractive index, the wavelength dependent, complex, frequently direction-dependent and occasionally non-linear parameter which defines most of the important propagation conditions for an optical field. We will discuss its physical origins and implications for wave propagation and also address absorption and scattering. In the second half of the chapter, we will provide an overview of the most relevant materials used in micro-optics; for each of these, we will learn about their most important characteristics and also consider their relative advantages and disadvantages when compared with alternative material systems.

Material choices

As we will see in detail below, most micro-optical components are fabricated using one or more of five basic material systems: glass, semiconductors, dielectrics, plastics and specialized crystals. Glass is perhaps the simplest and most established; semiconductors include silicon and germanium as well as the compound semiconductors in the III-V or II-VI material systems; dielectrics are exemplified by silicon dioxide, silicon nitride, or titanium dioxide; plastics include the polycarbonate of which CDs are made; and in the family of specialized crystals, lithium niobate is one example of a material with very attractive properties.

So which material do we prefer to use? The cheapest which does the job. This trite answer comprises two essential aspects of materials: performance and price. Perfor-

¹ "A cook will produce a different cuisine when he knows his raw materials." Alain Passard (b. 1956), culinary magician, chef of l'Arpège in Paris; quoted in (Astier, 2007)

mance includes a wide variety of potential factors and the design engineer has to ask him or herself a long list of questions: does the material merely have to be transparent at the operating wavelength? Does it have to absorb certain wavelengths and thereby generate electrons? Does the refractive index need to be specified with high accuracy? Does the absolute value of refractive index have to be excessively high? Or low? Does dispersion have to be low? Should it be high? Does the material have to emit light? Should it be birefringent? Are optical non-linear characteristics, such as electro-optic effects, essential?

In general, the more questions which are answered with “yes”, the more expensive the material. If mere transparency is required, cheap glass or plastic may do the job. If the refractive index, however, needs to be exactly specified, or have an uncommonly large or small value, expensive speciality glasses or polymers may be required. Optical absorption in a certain wavelength range and subsequent current generation, such as required for photodetectors, necessitates semiconductors or specialized polymers. Light emission, needed for the fabrication of LEDs or laser diodes, typically lies within the domain of compound semiconductors or even more advanced polymers. Finally, efficient electro-optic or other non-linear effects often require the use of specialized crystals.

Thus materials range widely in their capabilities and this range is typically reflected in their cost. Simple glasses and plastics may be bought in bulk for a pittance, noting however that specialized glasses may quickly become very expensive. Silicon as a semiconductor is relatively economical, but really only when compared to compound semiconductors such as GaAs; these materials also typically require a cleanroom for fabrication, a facility that can easily have a price tag in the tens of Mega-Euros. At the luxury end of the scale, materials such as LiNbO_3 , for example, have prices per substrate comparable to dinner for two at Ducasse in Paris; for that investment, however, the researcher obtains a material that can do virtually anything.

The choice of material is then a function of required functionality and acceptable cost. For the wide variety of micro-optical components we will consider in this text, a spectrum of suitable materials from which they may be manufactured can be defined. The ultimate application will then generally define which of these makes the most sense.

3.1 Refractive index

The difference between vacuum and a material is mathematically most easily represented by the refractive index. We were already introduced to this dimensionless parameter n in Section 2.3.2, where we saw that wave velocity in a material was reduced from c to c/n . We consider refractive index in greater detail here and will realize that it represents the most important means for optically characterizing a material.

3.1.1 Propagation in a material

As we saw in Section 2.3.2, the velocity of a wave in a material is given by

$$v = \sqrt{\frac{1}{\mu_m \epsilon_m}} = \frac{c}{\sqrt{\mu \epsilon}} = \frac{c}{n} \quad (3.1)$$

where the refractive index n was assumed to be a dimensionless number. We also recall the definitions

$$\epsilon_m = \epsilon \epsilon_0 \quad (3.2)$$

and

$$\mu_m = \mu \mu_0. \quad (3.3)$$

As well shall see presently, n is in general complex, direction-dependent and frequency dependent.

The relationship above was derived from definition of the electric flux density in Equation 2.3, namely

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}. \quad (3.4)$$

The change of permittivity from ϵ_0 to ϵ_m is due to the polarizability P of a medium, which we can write as

$$\mathbf{P} = \epsilon_m \mathbf{E} - \epsilon_0 \mathbf{E} = \epsilon \epsilon_0 \mathbf{E} - \epsilon_0 \mathbf{E} = \epsilon_0 (\epsilon - 1) \mathbf{E} \quad (3.5)$$

in which ϵ is the dielectric constant and $\epsilon = 1$ in vacuum. Polarizability, which we will discuss in greater detail below, represents the extent to which the atoms and electrons in a material may be polarized by a propagating electric field; polarizability should not be confused with polarization. Defining the susceptibility χ [] as

$$\chi = \epsilon - 1, \quad (3.6)$$

we see that $\chi = 0$ in vacuum; the polarizability is then given by

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E} \quad (3.7)$$

and is also zero in vacuum. We will re-visit susceptibility again below.

From these considerations we define the simplest form of the refractive index as

$$n = \sqrt{\epsilon} = \sqrt{\chi + 1} \quad (3.8)$$

which has a single, real value. We see from the above discussion that the parameters n , ϵ , ϵ_m , χ and \mathbf{P} all basically define the same thing: the effect of a medium on an electromagnetic wave. Which parameter to use is a matter of taste and training: engineers tend to like n , physicists often prefer ϵ .

Complex propagation constant

Based on our derivation of the wave equation in Section 2.1.4, we can consider the complete form of the analysis which we saw in Equation 2.23, namely

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu_m \left[\sigma \frac{d\mathbf{E}}{dt} + \epsilon_m \frac{d^2 \mathbf{E}}{dt^2} \right]. \quad (3.9)$$

In our discussion of the last chapter, we assumed that the conductivity $\sigma = 0$. We will now perform a more general analysis, assuming that there are no free charges (i.e., $\rho = 0$ so that $\nabla \cdot \mathbf{E} = 0$) but taking $\sigma \neq 0$. This case corresponds to a realistic material; metals of course have high conductivity, but even dielectrics such as glass have low, but non-zero, conductivity. We will see at the end of this analysis that non-zero σ leads to absorption of an electromagnetic wave. Explicitly retaining σ , we then have

$$\nabla^2 \mathbf{E} = \mu_m \left[\sigma \frac{d\mathbf{E}}{dt} + \epsilon_m \frac{d^2 \mathbf{E}}{dt^2} \right] = \mu_m \left[\sigma \frac{d\mathbf{E}}{dt} + \epsilon_0 \frac{d^2 \mathbf{E}}{dt^2} + \frac{d^2 \mathbf{P}}{dt^2} \right] \quad (3.10)$$

This more general form of the wave equation also has harmonic solutions, as we saw in the last chapter. If we take a field of the form (ignoring the vector aspect for the time being)

$$E = E_0 \exp j[\omega t - kz] \quad (3.11)$$

we find that this solution yields a propagation constant of the form

$$k^2 = \mu_m \epsilon_m \omega^2 + j\sigma \mu_m \omega. \quad (3.12)$$

For the case that $\sigma = 0$, this last expression simplifies to

$$k^2 = \frac{\omega^2}{v^2} = \frac{\omega^2}{c^2} n^2 = \frac{\omega^2}{c^2} \epsilon = \frac{\omega^2}{c^2} (1 + \chi) \quad (3.13)$$

which is purely real. This expression for refractive index is the one we considered in the previous chapter.

In the general case for which $\sigma \neq 0$, k becomes complex and thus takes the form

$$k = k_R + jk_I \quad (3.14)$$

for real part k_R and imaginary part k_I . Taking the complex root of k^2 as defined in Equation 3.12, we obtain

$$k = \left[(\mu_m \sigma \omega)^2 + (\epsilon_m \mu_m \omega^2)^2 \right]^{\frac{1}{4}} \left\{ \cos \left[\frac{1}{2} \text{Arg}(\epsilon_m \mu_m \omega^2 + j\mu_m \sigma \omega) \right] + j \sin \left[\frac{1}{2} \text{Arg}(\epsilon_m \mu_m \omega^2 + j\mu_m \sigma \omega) \right] \right\} \quad (3.15)$$

where the function $\text{Arg}(a + jb)$ yields the angle defined in the complex plane by the complex vector $(a + jb)$.

Let us look at a few special cases. We can easily verify, for example, that $\sigma = 0$ in the expression above yields a purely real value for k , namely $k = \omega/c$, as we had previously.

If $\sigma > 0$ but is small (meaning that $k_I^2 \ll k_R^2$), as for a low-conductivity material, we then have

$$k^2 = (k_R + k_I)^2 = k_R^2 - k_I^2 + 2jk_Rk_I \approx k_R^2 + 2jk_Rk_I \quad (3.16)$$

which we can set equal to the expression for k^2 in Equation 3.12,

$$k_R^2 + 2jk_Rk_I = \mu_m \epsilon_m \omega^2 + j\sigma \mu_m \omega. \quad (3.17)$$

Equating the real and imaginary parts, we then obtain

$$k_R = \sqrt{\epsilon_m \mu_m \omega} \quad (3.18)$$

$$k_I = \frac{\sigma}{2} \sqrt{\frac{\mu_m}{\epsilon_m}}. \quad (3.19)$$

Finally, for the case where $\sigma \gg 0$, or more specifically $\sigma \gg \epsilon_m \omega$, which is the case for a high-conductivity material, we may derive

$$k_R = \frac{\sqrt{2}}{2} \sqrt{\mu_m \sigma \omega} \quad (3.20)$$

$$k_I = \frac{\sqrt{2}}{2} \sqrt{\mu_m \sigma \omega} \quad (3.21)$$

such that the real and imaginary parts are equal. We thus have three special cases which allow us to apply simplified forms of the real and imaginary propagation constant to the determination of the propagation behavior.

Complex refractive index

From the propagation constant, we may then easily determine the refractive index, which also becomes a complex quantity of the form

$$n = n_R + jn_I. \quad (3.22)$$

We may easily relate this to k by

$$n = \frac{c}{\omega} (k_R + k_I). \quad (3.23)$$

Let us again look at the useful special cases. For $\sigma = 0$, we have, as previously,

$$n = \sqrt{\epsilon \mu} \quad (3.24)$$

For $\sigma > 0$ but small, the refractive index becomes

$$n = \sqrt{\epsilon \mu} + j \frac{\sigma}{2\omega \epsilon_0 \sqrt{\epsilon \mu}} \quad (3.25)$$

and for $\sigma > 0$ and large,

$$n = \sqrt{\epsilon \mu} + j \frac{c}{\sqrt{2}} \sqrt{\frac{\mu_m \sigma}{\omega}}. \quad (3.26)$$

We see thus that a complex propagation constant, which results when we consider the non-zero conductivity of a material, leads to a complex refractive index.

Kramers-Kronig relation

The complex refractive index, of course, has a real and an imaginary part. For passive frequency-dependent linear functions which have that form, namely

$$\zeta(\omega) = \zeta_R(\omega) + j\zeta_I(\omega), \quad (3.27)$$

the real and imaginary parts are related to each other and one may be found from the other using the Kramers-Kronig relations² (Kittel, 2005, Chapter 15). For our frequency-dependent complex refractive index, $n(\omega) = n_R(\omega) + jn_I(\omega)$, Kramers-Kronig then allows us to determine n_I from n_R by

$$n_I(\omega) = -\frac{2\omega}{\pi} \text{P} \int_0^\infty \frac{n_R(s)}{s^2 - \omega^2} ds \quad (3.28)$$

and n_R from n_I by

$$n_R(\omega) = \frac{2}{\pi} \text{P} \int_0^\infty \frac{sn_I(s)}{s^2 - \omega^2} ds \quad (3.29)$$

where s [rad/s] is an integration variable. In both of the previous expressions, P denotes the principle part of the integral, namely

$$\text{P} \int_0^\infty = \lim_{\delta \rightarrow 0} \left(\int_0^{\omega-\delta} + \int_{\omega+\delta}^\infty \right). \quad (3.30)$$

by which the singularity at $s = \omega$ is avoided.

We see thus that knowledge of either the real or imaginary spectrum of the refractive index allows calculation of the other. Since the integrals are over all ω , we need the complete spectrum, for $0 \leq \omega \leq \infty$; but in practice, only a portion of the spectrum in a certain wavelength range is known. However, making reasonable assumptions about the behavior of $n(\omega)$ outside the region in which it can be measured does allow, for example, real refractive index (typically not easy to measure) to be calculated from measurement of imaginary refractive index (absorption, usually easier to measure) (Madelung, 1978).

Implications for propagation

We recall that the solutions to the wave equation are propagating harmonic functions, given by the mathematical form of Equation 3.11. If we have a purely real propagation constant, the electric field then varies as

$$E = E_0 e^{j\omega t} e^{jk_R z} \quad (3.31)$$

² Hendrik Anthony (Hans) Kramers (1894–1952) was a Dutch physicist, a student of Niels Bohr, who spent most of his career as a professor in Utrecht, Leiden and Delft; Ralph Kronig (1904–1995), a German, also spent most of his career in Delft, and is remembered for his work on X-ray spectroscopy and electron spin.

as we had previously. This solution is a purely harmonic function, varying as the sum of sinusoids. If k is complex, this relationship becomes

$$E = E_0 e^{j\omega t} e^{jk_R z} e^{-k_I z} \quad (3.32)$$

which has an additional exponentially decaying term which varies with the imaginary part of the propagation constant.

The general propagation equation for an electromagnetic wave thus has three components: 1) a sinusoidally varying time variation, oscillating with ωt ; 2) a sinusoidally varying spatial variation oscillating with $k_R z$; and 3) an exponentially decaying spatial variation decreasing with $k_I z$. This last component corresponds to absorption in the material, leading to a decrease in the magnitude of the field.

Example 3.1: We have two semiconductors with differing refractive indices and conductivities: the first with $n_1 = 3.45$ and $\sigma_1 = 5 \times 10^2$ 1/ Ωm and the second with $n_2 = 3.63$ and $\sigma_2 = 8 \times 10^3$ 1/ Ωm . Let us plot the spatial variation of the electric field for incidence of a HeNe laser beam ($\lambda = 633$ nm) onto the surface of these materials.

We may directly employ the general equation for the propagation of a field in a material, Equation 3.32. Since the case here is that $\sigma > 0$ but small, we then employ the definitions we derived above for the propagation constants, namely

$$k_R = \frac{n\omega}{c}$$

$$k_I = \frac{\sigma}{2nc\epsilon_0}$$

where we remember that

$$\omega = 2\pi \frac{c}{\lambda} = 2.98 \times 10^{15} \text{ rad/s.}$$

Entering these values into Equation 3.32, taking the real part and plotting the field variation for the first 5 μm under the surface, we obtain the plots shown in Figure 3.1. We see the harmonic spatial variation of the field, defined by the value of k_R (assuming phase $\phi = 0$ at the surface, $z = 0$) as well as the exponential decay of the envelope, defined by k_I . It is clear that a larger value for conductivity implies a more rapid decay with propagation distance, and thus larger absorption, than that found in materials with lower conductivity.

3.1.2 Susceptibility

It is a legitimate question to ask what the origin of the refractive index is: why is $n \neq 1$ and is thus light propagation slower in a material than in vacuum? We can usefully use the concept of susceptibility to understand the physical principles on which the refractive index is based.

We saw in Section 3.1.1 above that the parameters n , ϵ , ϵ_m , χ and \mathbf{P} are all inter-related and that the polarizability of a material is related to the electric field by the

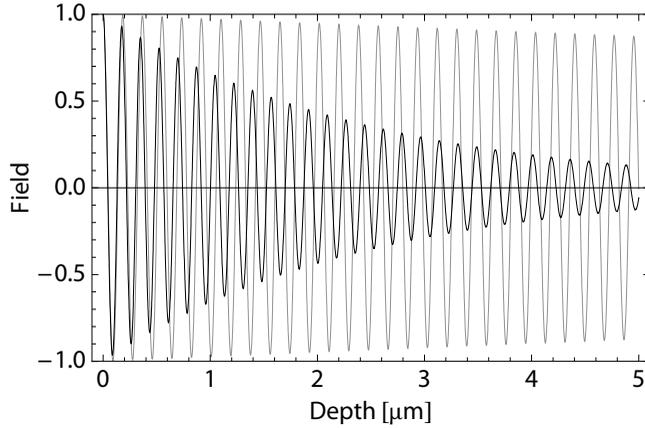


Figure 3.1 Plot of the decay of the electric field incident onto two materials with differing conductivities: the relatively slowly decaying ($\sigma_1 = 5 \times 10^2 \text{ 1}/\Omega\text{m}$ and $n_1 = 3.45$, gray) and the more rapidly decaying ($\sigma_2 = 8 \times 10^3 \text{ 1}/\Omega\text{m}$ and $n_2 = 3.63$, black).

susceptibility χ as $\mathbf{P} = \epsilon_0 \chi \mathbf{E}$. Thus the polarizability is key to understanding refractive index. Heuristically, an electric field propagating through a material interacts with the components of that material, the atoms and electrons. At certain frequencies, the incident field excites the atoms, electrons or combinations thereof, causing them to resonate as harmonic oscillators. These harmonic oscillators also generate electric fields which in turn interact with the incident field, such that propagation of a wave may be imagined as a long trail of interacting fields which emit and absorb energy.

Since atomic or electronic resonances occur at certain energy levels, or between well-defined energy states, the strength of the field interaction is highly energy- or wavelength-dependent. In addition, the response is both real and imaginary: the real response corresponds to a change in the wave velocity, the imaginary response to absorption of energy and thus a change in wave magnitude.

Atomic susceptibility

A simple mechanical model can be used to derive the response of the atoms in a material to an incident electric field (Zappe, 1995, Chapter 3.4). Let us model a dielectric material as a three-dimensional arrangement of masses (the atoms) connected by springs (the forces holding the atoms together), as shown in Figure 3.2. For masses m [kg] connected by springs with spring constant K [N/m], excitation of this system with an electric field E will yield a movement of the masses (which have a charge and thus respond to an electric field) by a displacement d [m]. The restoring force F [N] due to the springs is given by the simple spring equation,

$$F = -Kd. \quad (3.33)$$

If we now include a damping of the system γ [s^{-1}] due to frictional forces, the atomic movement due to the exiting field may be characterized by the general force balance

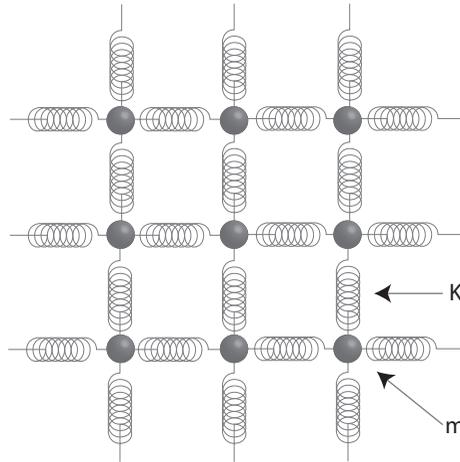


Figure 3.2 Two-dimensional representation of an atomic lattice modeled as masses m connected by springs with spring constant K .

equation

$$\frac{d^2d}{dt^2} + \gamma \frac{dd}{dt} + \frac{K}{m}d = \frac{q}{m}E \quad (3.34)$$

for fundamental electronic charge q . Recalling that oscillatory systems of this nature react to a harmonic excitation

$$E(t) = E_0 e^{-j\omega t} \quad (3.35)$$

with a harmonic response

$$d(t) = d_0 e^{-j\omega t}, \quad (3.36)$$

we can plug these last two expressions into the force balance equation which then simplifies to

$$-\omega^2 d_0 + j\omega\gamma d_0 + \frac{Kd_0}{m} = \frac{q}{m}E_0. \quad (3.37)$$

It can easily be shown that Equation 3.37 has as solutions for the displacement of the form

$$d_0 = \frac{qE_0}{m} \left(\frac{K}{m} - \omega^2 + j\omega\gamma \right)^{-1}. \quad (3.38)$$

We return now to the concept of polarizability: on the atomic level, P is directly related to the sum of the dipole moments of the system, where the latter is defined as the charge of a pair of bodies times their spacing, qd_0 . For N atoms in a system, assuming for this simple model that all have the same induced movement d_0 due to the applied field, we thus have

$$P = qNd_0. \quad (3.39)$$

Using the expression for the induced movement d_0 due to E given above, we may thus simply express the polarizability as

$$P = \frac{\frac{q^2 N}{m} E_0}{(\omega_0^2 - \omega^2 - j\omega\gamma)} \quad (3.40)$$

where we have employed the definition of the resonance frequency,

$$\omega_0 = \sqrt{\frac{K}{m}}. \quad (3.41)$$

Since polarizability is related to electric field as $\mathbf{P} = \epsilon_0 \chi \mathbf{E}$, we have susceptibility given as (ignoring the vector properties of \mathbf{E} and \mathbf{P})

$$\chi = \frac{P}{\epsilon_0 E_0} \quad (3.42)$$

and thus

$$\chi = \frac{q^2 N}{\epsilon_0 m (\omega_0^2 - \omega^2 - j\omega\gamma)}. \quad (3.43)$$

For a material or system in which there are multiple contributions to the polarizability (as would be the case for most materials, in which there are a variety of atomic resonances, all at distinct frequencies), the general susceptibility may then be written as

$$\chi = \sum_n \frac{q^2 N_n}{\epsilon_0 m_n (\omega_{0n}^2 - \omega^2 - j\omega\gamma_n)} \quad (3.44)$$

for n distinct resonance types.

We see that Equation 3.44 is complex, which we would expect based on the discussion of refractive index and propagation constant above. Indeed, in the frequency range near resonance, by assuming $\omega \approx \omega_0$ and $\omega \gg \gamma$, we can approximate this expression for susceptibility as (Siegman, 1986, Chapter 2.4)

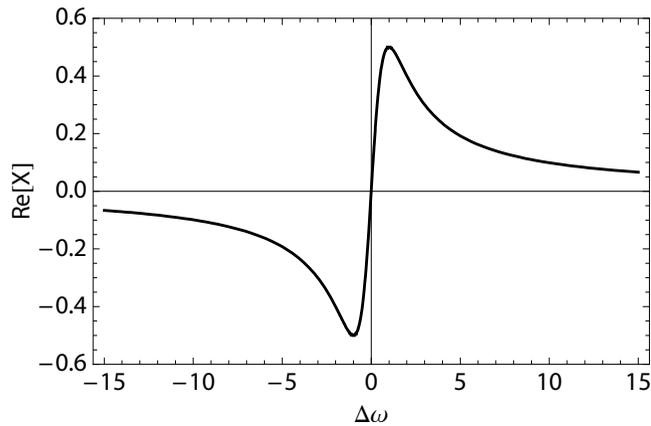
$$\chi \approx \chi_0 \left(\frac{\Delta\omega}{1 + \Delta\omega^2} + j \frac{1}{1 + \Delta\omega^2} \right) \quad (3.45)$$

where we have defined

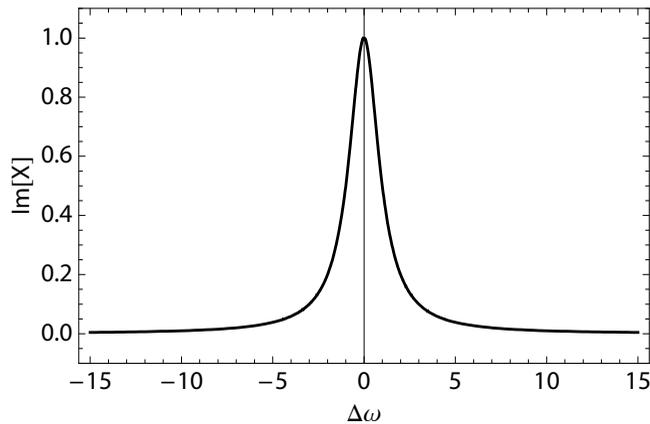
$$\Delta\omega = \frac{\omega - \omega_0}{\gamma}. \quad (3.46)$$

Plotting the variation of the real and imaginary parts of Equation 3.45 as a function of $\Delta\omega$, such that $\Delta\omega = 0$ corresponds to the atomic resonance frequency, we obtain the variations shown in Figure 3.3. Both terms only have significant variations around resonance; the imaginary part peaks at $\omega = \omega_0$ implying a maximum of absorption at that frequency. We will see below that this variation agrees qualitatively with what is measured in experiment.

Thus complex susceptibility has the same implications as complex n or k : the real part defines the propagation velocity and the imaginary part, the absorption. We see from this simple model that the resonances which result from interaction of an electromagnetic wave with the atoms of a material give rise to the complex refractive index.



(a) Real part of the susceptibility.



(b) Imaginary part of the susceptibility.

Figure 3.3 The variation of the real and imaginary parts of the susceptibility (Equation 3.45) as a function of frequency spacing from resonance, $\Delta\omega = (\omega - \omega_0)/\gamma$.

Electronic susceptibility

In systems with a high concentration of free electrons, such as metals, a similar model may be used to derive the susceptibility due to electronic resonances. Metals are conventionally characterized using the dielectric constant,

$$\epsilon = (n_R + jn_I)^2 \quad (3.47)$$

and the Drude model³ (Ziman, 1972, Chapter 8.6), which assumes non-interacting free electrons, is usefully employed. The electronic susceptibility derived from the Drude

³ Paul Drude (1863–1906), a German physicist who worked extensively on the optics of materials, is remembered today for his eponymous model relating the optical, electrical and thermal properties of media.

representation may be written as

$$\chi = -\frac{n_e q^2}{\epsilon_0 m_e \omega^2} \quad (3.48)$$

for density of electrons n_e [m^{-3}] and electron mass m_e [kg]. By defining the plasma frequency ω_p [rad/s] as (Seeger, 1999, Chapter 11.7)

$$\omega_p = \sqrt{\frac{n_e q^2}{\epsilon_0 m_e}}, \quad (3.49)$$

the dielectric constant due to electronic resonances in metals may thus be expressed as

$$\epsilon_{Drude} = 1 - \frac{\omega_p^2}{\omega^2}. \quad (3.50)$$

Plasma frequency is to first order a function of electron density and, in metals, typical values for ω_p are the terahertz range.

For low frequencies, with respect to the plasma frequency, we see from the previous expression that $\epsilon_{Drude} < 0$, so that the imaginary part of the refractive index dominates the real part and $n \rightarrow n_I$ such that $\sigma \rightarrow \infty$. In this case the electric field is *reflected*, as we expect from our experience with metals. For high frequencies, $\omega > \omega_p$, the electric field is transmitted but, depending on the magnitude of n_I , part of it is absorbed.

For metals in general, the reflectivity r [] (defined as the ratio of reflected to incident electric field and which we will consider extensively in Chapter 4 below) is given by

$$r = \frac{(n_R - 1)^2 + n_I^2}{(n_R + 1)^2 + n_I^2} \quad (3.51)$$

where we recall that both n_R and n_I are strongly frequency-dependent. For the realistic case that $\sigma < \infty$, ϵ is not purely imaginary and thus $r < 1$. We thus see that metal mirrors are not perfect reflectors and we also that the value of reflectance varies with frequency.

Total polarizability

If we examine the polarizability of a material over wide range of electric field excitation frequencies, we will find frequency regimes in which either the imaginary or real (or both) parts vary strongly (a resonance) and others in which little or no variation is seen. Correspondingly, the susceptibility (or dielectric constant or refractive index) varies strongly with frequency in regions where an atomic or electronic resonance is found, and is relatively constant elsewhere.

The typical form of susceptibility variation for a dielectric such as glass over a broad portion of the electromagnetic spectrum is seen in Figure 3.4. Atomic and ionic resonances dominate below about 10^{14} Hz, thus for infrared and longer wavelengths, whereas electronic resonances are relevant for frequencies above 10^{16} Hz, thus ultraviolet or shorter wavelengths. On the coarse scale of this figure, real susceptibility is relatively constant and imaginary susceptibility approximately zero for this material at visible wavelengths; we will see below, however that a more precise look at these val-

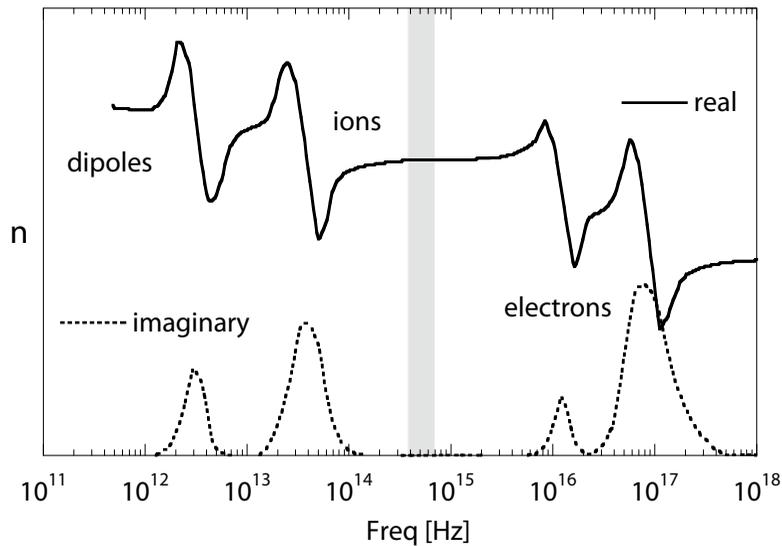


Figure 3.4 Variation of the real and imaginary parts of susceptibility (or dielectric constant or refractive index) as a function of excitation frequency for a glass-like dielectric. The contributions of the various components of a material for different frequency ranges is shown. The gray bar roughly indicates the visible wavelength region.

ues does show variation at visible wavelengths and, depending on the application, this variation may be significant.

Complex refractive index, again

By regarding the atomic and electronic components of the material as a whole and taking screening into account, such that the local fields generated by the applied field are considered, the susceptibility as calculated above may be related to the refractive index (or dielectric constant) by the Clausius-Mossotti equation⁴ (Kittel, 2005, Chapter 16). This equation, given by

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{n^2 - 1}{n^2 + 2} = \frac{4\pi}{3}\chi \quad (3.52)$$

relates the macroscopically employed parameters n or ϵ to the microscopically-derived susceptibility; this relation is valid for materials in which electronic screening occurs. By solving the Clausius-Mosotti equation for n or ϵ as a function of χ , we can determine the value of refractive index at any excitation wavelength, assuming that susceptibility is known.

⁴ Rudolf Clausius (1822-1888), a German, was one of the founders of the field of thermodynamics, proposing that which is known today as the Second Law of Thermodynamics and naming the concept of entropy; the Italian Ottaviano-Fabrizio Mossotti (1791-1863), fled his native country due to his revolutionary political views and became professor of physics in Argentina.

3.2 Dispersion

We saw in Section 3.1.1 above that the propagation of an electromagnetic field, such as a light wave, through a material leads to an interaction which is strongly frequency (or wavelength) dependent; as outlined in Section 3.1.2, this dependence is due to various atomic and electronic resonances in the material. The consequence of these effects is that refractive index n , the simplified parameter with which an optical material is generally characterized, varies with wavelength, a phenomenon we consider in detail in the following two sections.

As we also saw repeatedly above, n is complex and thus has real and imaginary parts, $n = n_R + n_I$. The optical transmission characteristics of a material are a function of the relative values of the real and imaginary components. In general, the response of the material to an electromagnetic field with increasing frequency (decreasing wavelength) qualitatively varies as follows:

- n_R large, n_I small — transmission
- n_R large, n_I maximum — absorption
- n_I small, n_I large — reflection
- n_R small, n_I small — transmission

Both n_R and n_I vary with wavelength independently but related through the Kramers-Kronig relation outlined in Section 3.1.1 above. We will consider the real and imaginary parts separately: the variation of n_R with wavelength is known as dispersion, affects the propagation velocity, and is the subject of this section; n_I defines attenuation in the material, affects the wave magnitude, and is the subject of Section 3.3.

3.2.1 Dispersion at visible wavelengths

Examination of the variation of real and imaginary susceptibility in Figure 3.4 seemed to show that neither vary very strongly in the visible wavelength range, in which much of optics takes place. If we look at a material in more detail at these frequencies, we see that there is indeed a variation of n_R with wavelength and, as we will see in Section 3.3 below, n_I is small but non-zero, even in “transparent” materials such as glass.

Figure 3.5 shows the variation of real refractive index of SiO_2 , the primary constituent of most glasses, with wavelength in the visible and near-infrared wavelength regimes. We see that in the visible wavelength range, roughly 400 to 700 nm, the refractive index varies from about 1.50 to 1.46 and decreases with increasing wavelength; in the near infrared, the decrease with wavelength is smaller, typical behavior for glasses (Hecht, 2002, Chapter 3.5). For applications in which the exact value of refractive index is relevant, dispersion characteristics such as this one are essential for an accurate optical design.

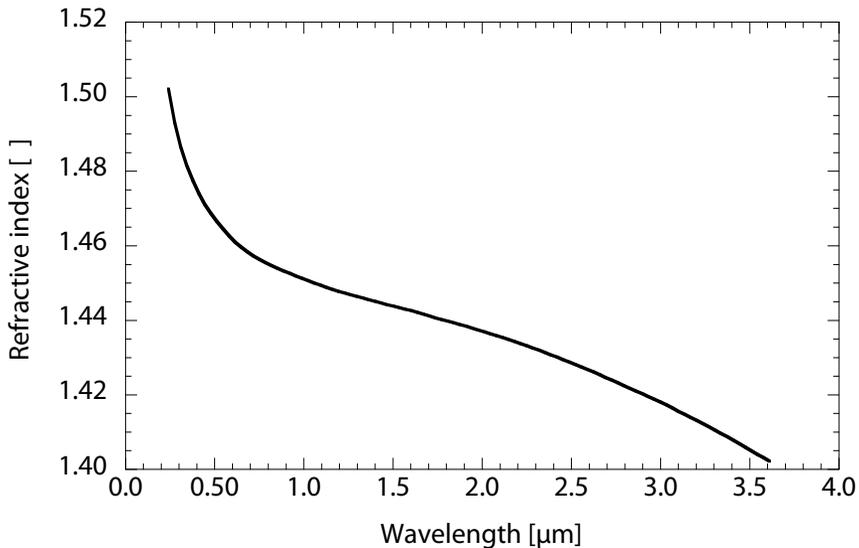


Figure 3.5 The variation of real refractive index n_R as a function of wavelength for SiO_2 at visible and near-infrared wavelengths.

Sellmeier formula

The dispersion characteristic is often of primary importance when employing a material in an optical system, and designers have tables and charts with this data available for many important materials. Whereas an analytic expression for $n(\lambda)$ is not derivable from first principles, a curve fit to the data using the Sellmeier dispersion formula⁵ (Sellmeier, 1872) provides a very useful approximation (Born and Wolf, 1980, Chapter 2.3). The Sellmeier formula is a polynomial fit to the dispersion characteristic and thus approximates $n(\lambda)$ as

$$n^2(\lambda) = 1 + \sum_{i=1}^3 \frac{B_i \lambda^2}{\lambda^2 - \lambda_i^2} \quad (3.53)$$

where B_i and λ_i are fitting parameters, defined for a wide range of popular materials. The fit to the dispersion characteristic is often better than 10^{-5} so that this expression is very useful for optical design. The Sellmeier parameters have been derived for a variety of optically relevant materials, especially SiO_2 (Malitson, 1965), and the numerous III-V semiconductor materials, including AlGaAs (Adachi, 1985; Blakemore, 1982), InP (Pettit and Turner, 1965) and InGaAsP (Fiedler and Schlachetzki, 1987). Values are also frequently given in glass catalogues, or may be determined using curve fitting software from measurement data.

⁵ W. Sellmeier of Berlin published his useful extension of the Cauchy formula for wavelength-dependent refractive index in 1872.

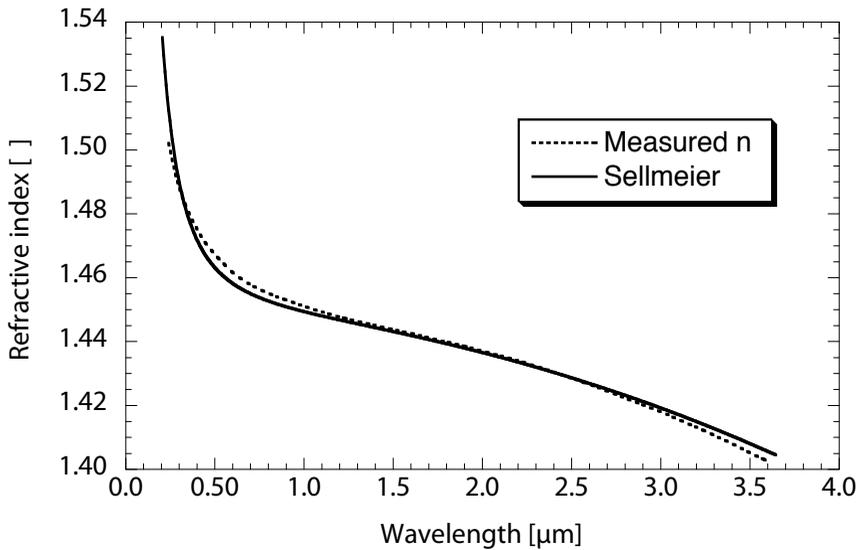


Figure 3.6 The dispersion of SiO_2 as calculated using the Sellmeier formula *solid line* compared with the measured characteristic shown in Figure 3.5 *dotted line*.

Example 3.2: The Sellmeier dispersion formula can be simplified somewhat mathematically to take the form (Simmons and Potter, 2000, Chapter 3.4)

$$n^2(\lambda) = 1 + A + \frac{B}{\lambda^2} - C\lambda^2.$$

For fused silica (SiO_2), these coefficients have been determined to be

$$A = 1.0994$$

$$B = 10974$$

$$C = 9.5988 \cdot 10^{-9}$$

for λ in nanometers.

In Figure 3.6, the dispersion characteristic as predicted by this simple approximation is plotted and compared with the measured $n(\lambda)$ shown in Figure 3.5. We see that the agreement is acceptable over much of the wavelength range. Applying Equation 3.53 in full would lead to even better agreement.

Table 3.1. Standard wavelengths for definition of the Abbe number ν

color	source	wavelength	index
red	Fraunhofer C line of H	656.2 nm	n_C
yellow	Fraunhofer d line of He	587.6 nm	n_d
green	Fraunhofer e line of He	546.1 nm	n_e
blue	Fraunhofer F line of H	486.1 nm	n_F

3.2.2 Abbe numbers

The dispersion characteristics of a material in the visible wavelength range are often summarized in a single numerical value, the Abbe number⁶. The Abbe number of a material, ν , is defined as the ratio of several well-defined refractive indices. These refractive indices, n_d , n_e , n_F and n_C , correspond to values measured at standard wavelengths; for the visible wavelength range, these are emission wavelengths of H and He and known as Fraunhofer lines, given in Table 3.1.

Using the yellow Fraunhofer d-line as a reference, we have the Abbe number ν_d defined as

$$\nu_d = \frac{n_d - 1}{n_F - n_C} \quad (3.54)$$

or perhaps more conveniently

$$\nu_d = \frac{n_{yellow} - 1}{n_{blue} - n_{red}}. \quad (3.55)$$

Alternatively, the Abbe number ν_e uses the green Fraunhofer e-line as a reference, and is thus given by

$$\nu_e = \frac{n_e - 1}{n_F - n_C}. \quad (3.56)$$

The Abbe number of a material is a measure for the level of dispersion, limited in this case to visible wavelengths. A small value for ν implies large dispersion (large variation of n with λ) and a large ν , small dispersion. Many glasses are specified with respect to their dispersion by means of the Abbe number; indeed, in many cases, n and ν are the most important initial catalogue values consulted by the optical engineer. The spectrum of glasses available from glass manufacturers often reflects this: the classic glass diagram published by the German manufacturer Schott is shown in Figure 3.11 (Section 3.4, below), and plots the company's speciality glass products on axes corresponding to n_d and ν_d . One can find, for example, the popular BK7 material, the sub-

⁶ Ernst Abbe (1840 – 1905), the great German optician, whose advances in lens and microscope design led him to the position of research director at Carl Zeiss, while still maintaining a professorship in Jena. Socially as well as technically advanced, he introduced, among other things, sick pay, holidays and pensions for his workforce.

strate of preference for many micro-optical components, on the chart at the intersection of $n_d = 1.512$ and $\nu_d = 64$.

Two designations which are frequently used, particularly for those designing achromatic lenses, also micro-optical ones, are those of *crown* and *flint* glasses. Crown glasses are those which have ($n_d > 1.6$ and $\nu_d > 50$) or ($n_d < 1.6$ and $\nu_d > 55$); all others are flint glasses.

3.2.3 Temperature dependence

Refractive index not only varies strongly with wavelength, it is also a function of temperature, due primarily to two effects: 1) thermal expansion of the material, which leads to a decrease in n with increasing temperature T ; and 2) increase in polarizability with temperature, which results in an increase in n . As a general rule, we can say that materials with a low thermal coefficient of expansion have n which increases with increasing T whereas those with a high thermal coefficient of expansion have n which decreases with increasing T .

The change in n with temperature can be expressed as a function of the variation of polarizability and thermal coefficient of expansion with T , namely (Fanderlik, 1983)

$$\frac{dn}{dT} = \frac{(n^2 - 1)(n^2 + 2)}{6n} \left[\frac{1}{P} \frac{dP}{dT} - \frac{1}{V} \frac{dV}{dT} \right] \quad (3.57)$$

where V is the molar volume of the material and P the polarizability. In this expression we can see clearly the relative weights of the two effects, which affect refractive index with opposite polarity.

3.2.4 Birefringent materials

A further parameter with which refractive index may vary in some materials is propagation direction. In our definition of the electric flux density in Equation 3.4, we tacitly assumed that the permittivity ϵ was not a function of the direction in which the wave vector points. In general, this is not the case: ϵ is direction-dependent, such that the electric flux density D as a function of electric field E , ignoring P for the time being, must be written in tensorial form, namely

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}, \quad (3.58)$$

in which we have taken the Cartesian coordinate system as a basis. For most materials with direction-dependent index of refraction, we can take the coordinate axes to correspond to the primary optical axes of the material, and thus write the previous definition in a much more practical, diagonal form, namely

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_{xx} & 0 & 0 \\ 0 & \epsilon_{yy} & 0 \\ 0 & 0 & \epsilon_{zz} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}. \quad (3.59)$$

We see, nevertheless, that, insofar as the diagonal elements of the matrix are not equal, different polarization directions for the electric field lead to different values of ϵ and thus n ; this effect is known as birefringence.

Depending on the relative values of ϵ_{xx} , ϵ_{yy} , and ϵ_{zz} , different types of materials with differing propagation conditions result. For the case that $\epsilon_{xx} = \epsilon_{yy} = \epsilon_{zz}$, the material is *isotropic*, the case we assumed implicitly until now. Isotropic materials tend to be cubic or centrosymmetric crystals, such as Si or NaCl, or amorphous materials such as SiO₂ or glass.

If $\epsilon_{xx} = \epsilon_{yy} \neq \epsilon_{zz}$, a *uniaxial* material results. In uniaxial materials, generally trigonal, tetragonal and hexagonal crystals, one crystalline axis has a different value of permittivity than the two others. The axis with the unique value for permittivity, ϵ_{zz} in this case, is known as the *extraordinary* axis, whereas the remaining two are termed the *ordinary* axes. When employing birefringent materials, knowing the orientation of the axes is thus essential. Typical uniaxial materials include quartz (crystalline SiO₂), calcite (CaCO₃), LiNbO₃ and KDP (KH₂PO₄), the latter a popular non-linear crystal used extensively in optical modulators.

For the case in which all axes have different permittivity values, $\epsilon_{xx} \neq \epsilon_{yy} \neq \epsilon_{zz}$, the material is known as *biaxial*, and is predominantly found amongst orthorhombic, monoclinic and triclinic crystals. Popular examples include mica (a crystalline silicate, typically structured as sheets) and lead oxide.

Birefringent materials have a broad range of applications in optics, most frequently as polarizers, polarization rotators and in certain types of beam splitters. As a popular demonstration of birefringence, the splitting of an image into two laterally shifted images using a calcite crystal serves as a useful means to illustrate this effect (Hecht, 2002, Figure 8.18 ff).

Example 3.3: Lithium niobate (LiNbO₃) is a popular, if very expensive, uniaxial birefringent material. A substrate is usually specified as “z-cut”, implying that the z -axis (sometimes labeled the c -axis) is the one with the extraordinary refractive index, n_e ; the two orthogonal axes have the ordinary refractive index, n_o . For LiNbO₃, the index values are typically $n_o = 2.2967$ and $n_e = 2.2082$.

As seen in Figure 3.7, differing orientations of the electric field will lead to differing values of refractive index seen for the wave. For a propagating field moving in the y -direction, TE (transverse electric) polarization implies the electric field in the x -direction, whereas TM (transverse magnetic) polarization leads to the electric field pointing in the z -direction (TE and TM will be defined more extensively in Chapter 4). Thus TE and TM polarized waves will see refractive indices which differ by about 4%.

Polarization retarders

One important family of applications for birefringent materials is polarization retarders, typically in the form of quarter-wave or half-wave plates. These take the form of a plate

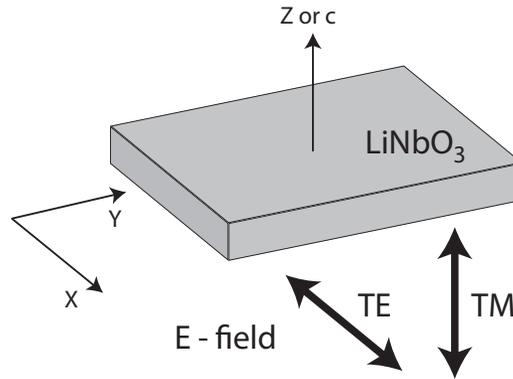


Figure 3.7 A LiNbO_3 crystal slab, indicating the z -axis with refractive index n_e and the x, y axes with refractive index n_o . The orientation of the electric field for two possible polarizations, TE and TM, is also shown.

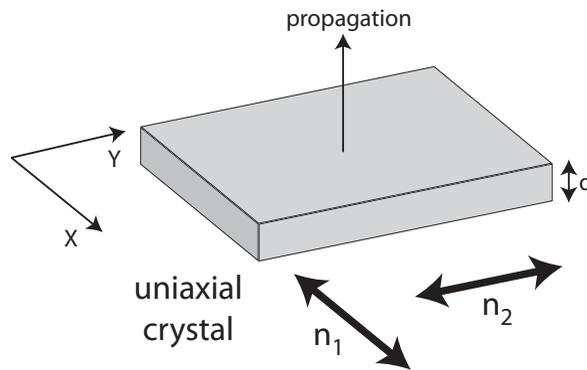


Figure 3.8 A polarization retarder made from a uniaxial crystal of thickness d with two differing refractive indices, n_1 and n_2 in the x - and y -directions, respectively. Propagation direction is normal to the xy -plane

of uniaxial material, as shown in Figure 3.8; the slab has a thickness d and has two differing refractive indices, n_1 and n_2 in the x - and y -directions, respectively.

The electric field components of a wave propagating vertically in the z -direction then have different propagation conditions, implying that the optical thickness (physical thickness d times refractive index n) is different for E_x and E_y . If we assume $n_2 < n_1$, then we obtain a phase retardation Γ [rad] of

$$\Gamma = \frac{2\pi(n_2 - n_1)}{\lambda}d \quad (3.60)$$

for wavelength λ . The retardation Γ defines the induced phase shift between the E_x and E_y components of the field, and we have two important special cases:

- $\Gamma = \frac{\pi}{2}$: linear polarization is converted to circular polarization and vice versa; this configuration is known as a *quarter-wave plate*.

- $\Gamma = \pi$: linear polarization is rotated by 90° ; this configuration is known as a *half-wave plate*.

We see that the value of Γ depends on the wavelength, the refractive index difference between the two axes and the thickness of the plate. Thus for a given birefringent material and fixed wavelength, changing the thickness of the plate changes the optical function. The function of phase retarding plates strongly depends on the operating wavelength and angle of incidence onto the plate.

Example 3.4: Given a birefringent plate of mica, with thickness d , it is known that the following refractive indices apply for the x and y axes:

$$n_1 = 1.594$$

$$n_2 = 1.599$$

What thicknesses are required for a quarter- and a half-wave plate for use with a HeNe laser, $\lambda = 633 \text{ nm}$?

We apply Equation 3.60 and determine that $\Gamma/d = 16\pi \text{ rad/mm}$. For a quarter-wave plate, $\Gamma = \pi/2$ is required, which results when $d = 31.25 \text{ }\mu\text{m}$; for a half-wave plate, $\Gamma = \pi$, obtained with $d = 62.5 \text{ }\mu\text{m}$.

3.2.5 Non-linear materials

Another tacit assumption remains in the definition of the electric flux density of Equation 3.4: it was assumed that the permittivity ϵ was not a function of the electric field strength, so $\epsilon \neq f(E)$. In non-linear materials, this is no longer the case: we have $\epsilon = f(E)$ or, perhaps more usefully for the following analysis, the susceptibility $\chi = f(E)$. As a result, the propagation conditions in non-linear materials change with the intensity of the electric field.

In non-linear materials, the simple definition for polarizability we had in Equation 3.7 is replaced by a series expansion of the form

$$P_j = \epsilon_0 \sum_{k=1}^3 \chi_{jk} E_k + \epsilon_0 \sum_{k=1}^3 \sum_{l=1}^3 \chi_{jkl} E_k E_l + \dots \quad (3.61)$$

where (j, k, l) correspond to the (x, y, z) components of the field and the χ_{jkl} are different susceptibility terms. By restricting our attention to linearly polarized light, with polarization in either the x , y or z direction, the previous expression can be simplified into the form

$$P = \epsilon_0 \chi_1 E + \epsilon_0 \chi_2 E^2 + \epsilon_0 \chi_3 E^3 + \dots \quad (3.62)$$

in which the χ_i are higher-order susceptibility terms, non-zero in non-linear materials. The susceptibility is thus field-dependent; writing the first three terms, we obtain

$$\chi = \frac{P}{E} = \epsilon_0\chi_1 + \epsilon_0\chi_2 E + \epsilon_0\chi_3 E^2. \quad (3.63)$$

The variation of susceptibility with applied field leads to the so-called electro-optic effects (Yariv, 1991, Chapter 9). In Equation 3.63, the first term corresponds to linear optics, the case we have considered implicitly up to now. The second term, $\epsilon_0\chi_2 E$, defines the linear electro-optic effect, a variation of susceptibility (or permittivity, or refractive index) linearly with the applied field; this variation is often referred to as the Pockels effect.⁷ The third term, $\epsilon_0\chi_3 E^2$, corresponds to the quadratic electro-optic effect, also known as the Kerr effect,⁸ which shows a quadratic variation of susceptibility with the applied field.

The magnitude of the electro-optic effect is a strong function of the material employed (Zappe, 1995, Chapter 11.3). The Pockels effect is significant only in non-centrosymmetric materials, including KDP, LiNbO₃, GaAs and InP; it is negligible in silicon. The effect is often a critical function of crystalline orientation (Eimerl, 1987), but is relatively wavelength-independent over a wide range. In contrast, the Kerr effect is present in silicon and in glass, with a strong wavelength dependence; it is weak in III-V semiconductors.

The Pockels and Kerr effects are widely used in optical systems, predominantly as the basis for phase and intensity modulators as we will see in Chapter 10, but also for second harmonic generation and parametric amplification. For the latter case, a non-linear crystal may be used to amplify an optical beam; using a strong pump beam and an “idler” beam at a frequency near that of the weak signal beam, the non-linear interactions allow amplification of the signal beam without energy transfer to the crystal (Cerullo and Silvestri, 2003).

Example 3.5: Second harmonic generation (SHG) uses non-linear materials to generate an optical field with frequency 2ω from an input field with frequency ω (Yariv, 1991, Chapter 8); frequency tripling is also possible. SHG relies on the linear electro-optic effect and thus requires a material for which χ_2 is sufficiently large; KDP and LiNbO₃ are both good candidates.

Given an excitation field

$$E = E_0 \cos(\omega t),$$

⁷ Friedrich Pockels (1865 – 1913) was professor of theoretical physics in Heidelberg. Of frail health, he was cared for by his older sister Agnes Pockels, who, self-educated since neither her parents nor contemporary society at large found higher education to be an activity suitable for women, made fundamental and important contributions to the field of surface science, and is arguably the better known of the pair.

⁸ John Kerr (1824 – 1907), an ordained Scottish minister with a strong interest in physics, was a friend and assistant of Lord Kelvin. He is best known for his discovery of the linear electro-optic effect which bears his name and his efforts to introduce the metric system into the UK.

the second-order term in Equation 3.63 then becomes

$$P \propto \epsilon_0 \chi_2 E_0^2 \cos^2 \omega t = \frac{1}{2} \epsilon_0 \chi_2 E_0^2 (1 + \cos 2\omega t).$$

From the previous expression, it is clear that the non-linear behavior of the material leads to the generation of a field at twice the excitation frequency. For effective frequency doubling, a material with a sufficiently large value of χ_2 is thus required.

SHG, or frequency doubling, is a useful mechanism for a variety of applications. The effect was first demonstrated in 1961, when a ruby laser was frequency-doubled ($\lambda = 694 \text{ nm} \rightarrow 347 \text{ nm}$) using quartz as the non-linear medium; the very low efficiency ($\approx 10^{-8}$) implied that the intensity required could only be reached using a laser, newly available at the time.

Today, a variety of materials with considerably higher SHG efficiencies are available, including KDP or LiNbO_3 ; for crystals several centimeters long, efficiencies of 0.3 are available. Lithium triborate (LBO) is very popular for frequency doubling or tripling the emission of Nd:YAG lasers (1 064 nm to 532 nm or 355 nm) and barium borate (BBO) may be employed to reach wavelengths in the UV, below 300 nm. The currently popular green laser pointer, with which professors like to attract their students' attention to the front of the lecture hall, is, contrary to popular belief, not a diode laser but a frequency-doubled Nd:YAG laser, using a $\text{Ba}_2\text{NaNb}_5\text{O}_{15}$ crystal. Instructors are still waiting for a portable CO_2 laser with which cell phones that ring in class may be vaporized.

3.3 Attenuation

We saw extensively in Section 3.2 that the real part of the refractive index, n_R , can have, in general, a strong wavelength variation and may also be a function of propagation direction, temperature and field intensity. The imaginary part of the refractive index may also vary with these parameters but the effect on the optical field is different: n_I defines the attenuation of the wave.

We may usefully divide the sources of wave attenuation into two categories: *absorption* implies the loss of photons and *scattering* a change in their wave vector, where both phenomena may lead to a reduction in field intensity.

3.3.1 Absorption

Optical absorption defines the loss of photons, typically by conversion to electrons or to vibrations in the atomic lattice of the material. As we saw in Equation 3.32, a non-zero imaginary part of the wave vector (and thus refractive index) causes an exponential decay of the field magnitude with propagation distance. The imaginary part of the refractive index is commonly written as the absorption coefficient, $\alpha \text{ [m}^{-1}\text{]}$, where $\alpha = n_I/2$.

Since optical measurements are usually of power, or power density times the known area of the photodetector, it is practical to use the optical intensity, $I(z)$ [W/m^2], as a measure for field magnitude; fields themselves (E or H) are relatively difficult to measure. Since intensity is a measure for power, and thus $I(z) \propto E^2 \propto H^2$, the decay of intensity with propagation distance z is given by

$$I(z) = I_0 e^{-\alpha z} \quad (3.64)$$

for initial intensity I_0 . This expression is known as Beer's law⁹ (occasionally referred to as the Beer-Lambert¹⁰ or Bouguer-Beer¹¹ law), and is characteristic for the loss of intensity of any optical wave propagating through an absorbing medium.

Example 3.6: Let us derive Beer's law by *a priori* consideration of the absorption coefficient, α . Consider for simplicity propagation in only the z direction of an optical field with z -dependent intensity $I(z)$ in a material with uniform absorption coefficient α . In the increment dz , the intensity changes by

$$dI(z) = -\alpha I(z) dz$$

such that

$$\begin{aligned} \frac{dI}{dz} &= -\alpha I \\ \frac{1}{I} dI &= -\alpha dz. \end{aligned}$$

By integrating this last expression from the point of incidence of the field onto the material, $z = 0$ where $I(0) = I_0$, to a point z , we obtain

$$\int_{I_0}^{I(z)} \frac{1}{I} dI = - \int_{z'=0}^z \alpha dz'$$

which evaluates to

$$\ln \left(\frac{I(z)}{I_0} \right) = -\alpha z$$

or

$$I(z) = I_0 e^{-\alpha z}$$

which is again Beer's law. Thus by considering the absorption coefficient phenomenologically, we obtain the same predicted behavior as by consideration of the propagation as a solution to the wave equation.

⁹ August Beer (1825 – 1863), was a professor of mathematics in Bonn, Germany, and is remembered today primarily for his optics textbook in which his eponymous law is presented.

¹⁰ Johann Heinrich Lambert (1728 – 1777), a German mathematician, born in the romantic little town of Mulhouse, in Alsace, France, was very active in astronomy and geometry.

¹¹ Pierre Bouguer (1698 – 1758), a French naval architect and cartographer, first described what is known today as Beer's law in 1729.

Decibels

The exponential decay of the optical intensity with propagation distance can result, in strongly absorbing media, in a change of intensity over several orders of magnitude. As a result, the optical loss due to absorption is usually expressed in decibels, logarithmic units derived from the ratio of intensity at a given point to an initial intensity. The power decibel (dB) is defined as¹²

$$\text{dB} = 10 \log \frac{I(z)}{I_0} \quad (3.65)$$

such a loss of -10 dB implies a reduction in intensity by a factor 10, -20 dB by a factor 100, and so forth.

The absorption loss in a material may then be given either by the absorption coefficient α , or in decibels per unit length, typically dB/cm. The two are simply related as

$$\alpha_{\text{dB}} = \frac{10}{\ln 10} \alpha \quad (3.66)$$

where the traditional units for α [cm^{-1}] and for decibels [dB/cm] are employed.

Absorption mechanisms

The absorption of photons is primarily due to transfer of their energy to electrons, atoms or molecules. We saw in the susceptibility model of Section 3.1.2 that, at certain resonant wavelengths, energy is transferred from a photon to a component of the material; for atoms or molecules, the resonance can also correspond to higher-order vibrational or rotational states.

An illustrative example is that of water molecules, or, more exactly, OH^- radicals, in glass. Water has a fundamental absorption peak at a wavelength of $\lambda = 2.73 \mu\text{m}$, such that significant optical absorption is expected at that wavelength. This fundamental absorption frequency also has numerous higher harmonics, such that we also expect absorption peaks at 1.13, 1.24 and 1.39 μm . Indeed, the measured absorption plot for glass shown in Figure 3.9, shows peaks at exactly those wavelengths, the most prominent at $\lambda = 1.39 \mu\text{m}$.

The total absorption characteristic shown in Figure 3.9 has a minimum around a wavelength of 1.55 μm and increases for shorter and longer wavelengths. The increase to shorter wavelengths is due to Rayleigh scattering, which we discuss further below, and the increase at longer wavelengths derives from the infrared band-edge of SiO_2 as well as phonon absorption. At the absorption minimum, an excess loss (summing over all contributions, absorption as well as scattering) of about 65 dB/km/ppm OH^- is found, such that water concentrations in the ppb range are required to obtain the minimum losses shown in the figure. The issue is not just academic; glass optical fiber is used for long-range optical data communications, and the lasers employed emit at 1.55 μm

¹² The power decibel should not be confused with the decibel defined for field values (E or H , or, in electrical engineering, voltage or current) which has a factor 20 rather than 10 as a multiplier.

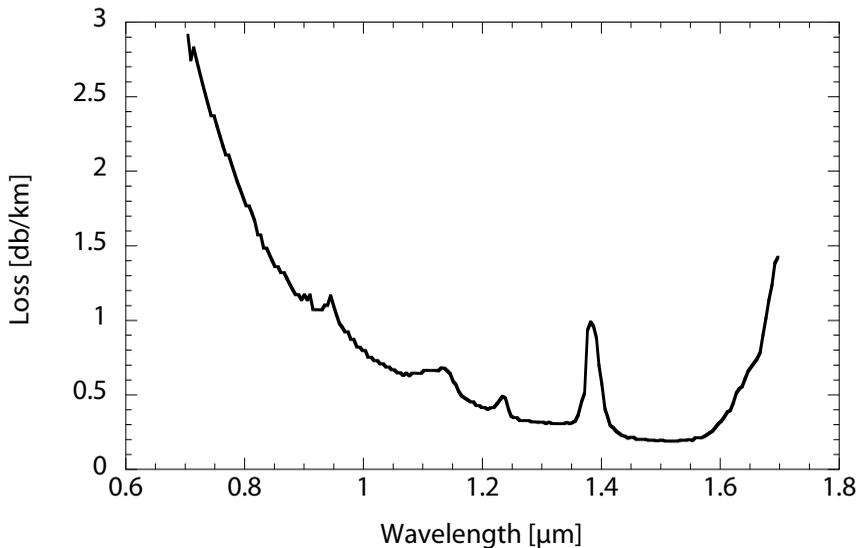


Figure 3.9 The optical absorption in glass as a function of wavelength in the visible and near-infrared ranges.

to take advantage of the loss minimum. Every extra dB/km reduces the maximum range a signal may be transmitted.

All components of a material, even those present in small quantities, may potentially give rise to excess absorption. In glasses, the dopants employed to adjust the (real) refractive index may also increase the (imaginary) refractive index. Silicon-oxygen bonds, for example, have a fundamental absorption resonance at $\lambda = 9 \mu\text{m}$ and the Ge-O bonds due to germanium doping induce a resonance at $\lambda = 11 \mu\text{m}$.

Semiconductor absorption

Semiconductors, which we will discuss in greater depth in Section 3.5 below, have a bandgap which strongly affects the optical absorption behavior. Unlike atomic resonances, which, as in SiO_2 , are found at mid- to far-infrared wavelengths, many semiconductors have a bandgap energy, E_g [eV], at visible and near-infrared wavelengths. Recalling from Equation 2.65 the relationship between energy and wavelength, bandgap energies of roughly $1 \text{ eV} \leq E_g \leq 3 \text{ eV}$ are typical for a wide variety of these materials.

For photons with energies greater than the bandgap energy, absorption in semiconductors is very strong; this is true for direct bandgap semiconductors as well as indirect bandgap semiconductors such as silicon. For photon energies below the bandgap, absorption can still be significant, as may be seen in Figure 3.10. This exponential, but not step-like, increase in absorption with decreasing wavelength is due to band tails or Urbach tails (Urbach, 1953); these are a temperature-dependent extension of available states from the conduction and valence bands into the bandgap due to imperfections in the crystal, dopants and impurities. For that reason, the use of a semiconductor for an application in which it needs to be transparent requires that the operating wavelength be

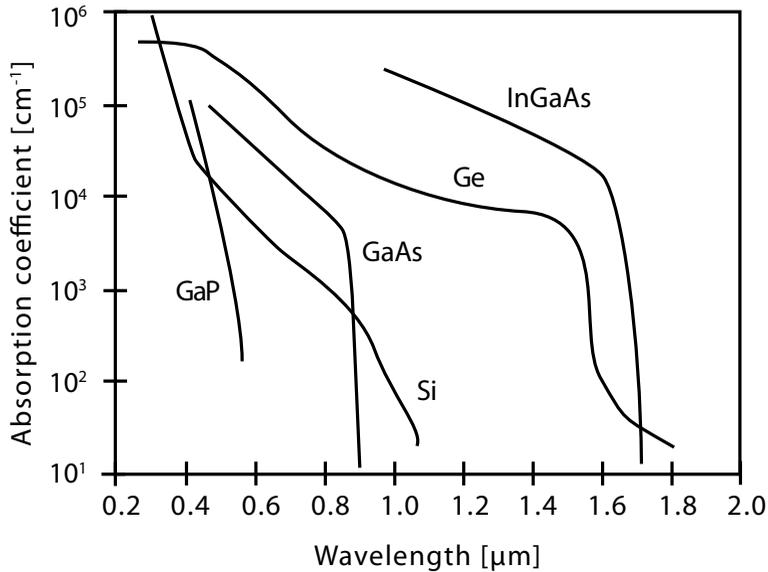


Figure 3.10 Wavelength-dependent absorption of various semiconductor materials, including III-V compound semiconductors (GaAs, GaP and InGaAs) and the elemental semiconductors silicon and germanium.

sufficiently long to avoid residual absorption, which occurs even for photons with $h\nu$ below the bandgap energy.

In addition to electronic inter-band transitions (valence to conduction band) leading to photon absorption, intra-band transitions may also cause considerable absorption loss in semiconductors. Electrons already in the conduction band can absorb photons of relatively low energy and thereby be raised to higher energies in the same band. The electrons then thermalize back to the bottom of the band, losing their energy to phonons. Such intra-band or free-electron absorption has a broad spectrum and can have a significant magnitude.

3.3.2 Scattering

Whereas absorption typically results in the complete loss of a photon, scattering changes the direction or magnitude (or both) of photons. The result is ultimately the same: the intensity of the optical field decreases, such that scattering processes may contribute to the overall value of α measured for a material sample. Elastic scattering events are those which only change the direction of \mathbf{k} , whereas inelastic scattering changes both the magnitude and direction of \mathbf{k} .

Scattering is a complex optical phenomenon and considerable effort has been extended in understanding and characterizing it (Stover, 1995). Several types of scattering are of primary importance in optical materials characterization, and we consider these briefly here.

Rayleigh scattering

Rayleigh scattering results from local fluctuations in real refractive index or from particles suspended in an otherwise transparent medium. The size of these fluctuations or particles is much smaller than the wavelength, so scattering at visible wavelengths is due to inhomogeneities on nanometer size scales. Rayleigh scattering varies in strength with wavelength as λ^{-4} , so it becomes more significant for shorter wavelengths.

For optical applications, Rayleigh scattering in glass is responsible for the increase in optical loss for wavelengths shorter than $\lambda = 1.55 \mu\text{m}$; we see this as the increase in α moving toward the left of the characteristic in Figure 3.9. Outside of the laboratory, Rayleigh scattering is the reason that the sky and some people's eyes are blue and that sunlight is polarized parallel to the ground.

Mie scattering

For particles or index fluctuations on size scales on the order of and larger than the wavelength, Mie scattering¹³ results. It also varies inversely with wavelength, but as λ^{-1} , implying again an increase in scattering rate with decreasing wavelength. Colored glasses, sometimes used in optical setups for filtering ranges of wavelengths, are often made using colloidal suspensions of metals in the glass, and rely in Mie scattering for the induced color.

Other scattering mechanisms

For particle or inhomogeneity sizes much larger than the wavelength, scattering may also occur but typically does not vary with wavelength, i.e., all wavelengths are scattered equally. The white hair some of us develop due to excessive administrative burdens is white for that reason: visible wavelengths are scattered uniformly on large-scale structures in the hair.

Brillouin¹⁴ and Raman¹⁵ scattering are both processes which rely on the emission of absorption of phonons; the former uses acoustic phonons, the latter employs optical phonons. Both of these are inelastic processes and result in a frequency shift of the scattered optical field. Whereas Rayleigh scattering, for example, results in no frequency shift, Brillouin scattering can give rise to $\Delta f \approx \pm 10^{12}$ Hz and Raman scattering $\Delta f \geq \pm 10^{12}$ Hz. Due to their sensitivity to composition of the medium, Brillouin and Raman scattering are more important as tools in materials analysis than they are annoying sources of optical loss in a sample.

¹³ Gustav Mie (1869–1957) was professor of physics at the University of Freiburg, Germany (the author's current employer), and studied scattering from colloidal nanoparticles a half-century before the word nano came into use.

¹⁴ Léon Brillouin (1889–1969), a French physicist who like his father, Marcel Brillouin, made important contributions to quantum mechanics. He spent the latter part of his career at different universities and at the IBM research laboratory in the USA.

¹⁵ Chandrasekhara Venkata Raman (1888–1970) was an illustrious Indian physicist, professor at the University of Calcutta, who received the Nobel Prize in 1930 for his discovery of the Raman effect.

3.4 Glass

Glass is the oldest and most established optical material for classical optics (Fanderlik, 1983; Bach and Neuroth, 1995) but is also of considerable relevance for micro-optics. Consisting essentially of silicon dioxide with a controlled concentration of impurities, glass is typically linear, highly stable, homogeneous, isotropic, inert and reasonably robust. Manufacturable in large quantities, glass can be very inexpensive, whereas specialty glasses with customized characteristics and tight tolerances can strain the research budget. Glass optics ranges in size from microlenses with micrometer dimensions (Merz et al., 2003; Krause et al., 2006) to multiple-meter-diameter telescope mirrors (Doehring et al., 2005), the latter fabricated with nanometer surface precision.

3.4.1 Composition

The primary constituent of most silicate glasses is SiO_2 , with dopants and impurities added in well-defined quantities to give a desired refractive index, dispersion or other optical characteristic. Glass components may include GeO_3 or B_2O_3 as well as the oxides Na_2O , K_2O or CaO . Silicate-based glasses usually have visible-wavelength refractive indices in the range 1.4 to 1.5; for higher index values, the titanates replace silicon with titanium and can have refractive index values up to about 2.4. Oxygen-based glasses are more or less transparent in the wavelength range $200 \text{ nm} \leq \lambda \leq 2500 \text{ nm}$; for transparency into the mid-IR, halide glasses replace the oxygen with a halide (F or Cl) and these may transmit for wavelengths longer than $10 \mu\text{m}$. For transparency at shorter wavelengths, the UV with $\lambda \leq 200 \text{ nm}$, materials such as calcium fluoride (CaF_2), magnesium fluoride (MgF_2) or lithium fluoride (LiF) have to be used.

For long-wavelength applications, a glass family also of interest in micro-optics is that of the chalcogenides. These are glasses in which the oxygen is usually replaced by elements from column VI of the periodic table, the chalcogens such as S, Se or Te; the silicon is typically replaced by metals such as As or Ge. Chalcogenides generally absorb strongly and are thus opaque at visible wavelengths; they typically transmit well at IR wavelengths up to $12 \mu\text{m}$ and have refractive indices larger than 2. Typical examples of chalcogenide glass include As_2S_3 , with $n = 2.35$, the exact value a function of composition, and $\text{As}_{11}\text{Se}_{55}\text{Ge}_{35}$, whose index of $n = 2.6$ may be modified by irradiation. Chalcogenides are generally deposited as thin films by evaporation and are useful in applications due to their strong acousto-optic effect.

3.4.2 Forms

Glasses are typically amorphous; quartz, in contrast, is crystalline SiO_2 . Glass is available in variety of forms, most popularly as solid substrates in a variety of shapes, but also as raw material for the generation of thin or thick-film layers, usually deposited by sputtering.

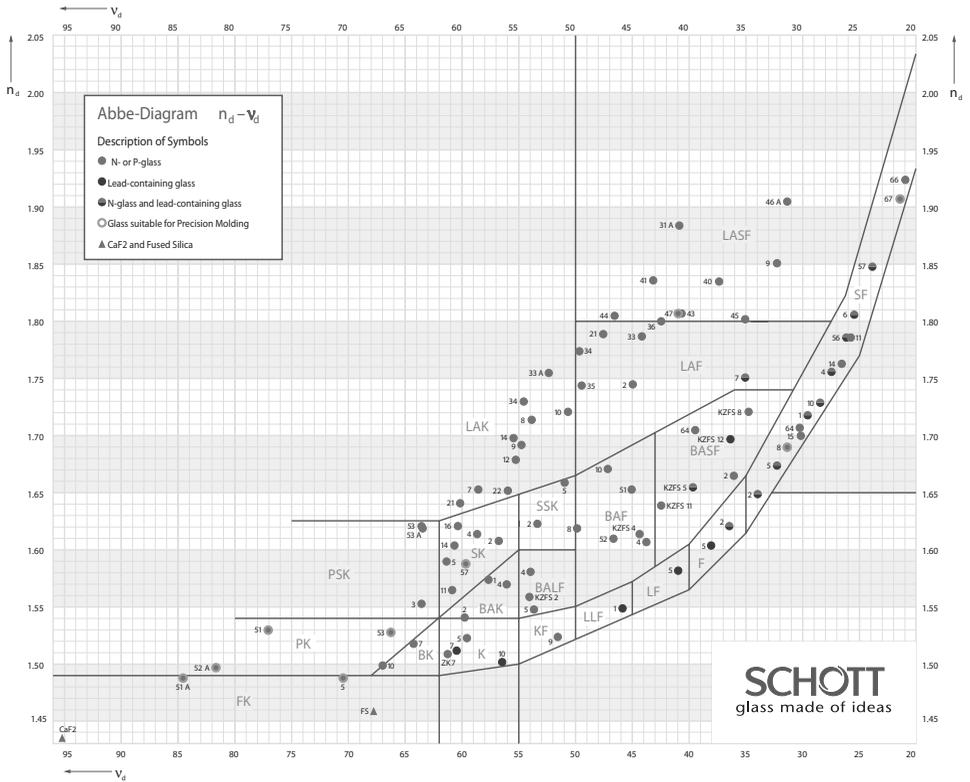


Figure 3.11 The Abbe diagram published by the German glass manufacturer Schott. Each dot corresponds to a glass product, plotted on axes corresponding to n_d vertical and ν_d horizontal; n_d and ν_d are defined in Section 3.2.2. The dots and triangles correspond to various glass products, grouped into material families. Diagram ©SCHOTT AG; reprinted courtesy of SCHOTT AG.

Substrates

For the micro-optics researcher, glass is most conveniently obtained as rectangular or circular plates, as lens blanks or as prisms. Easily available substrate sizes range from several millimeters to 1 m edge lengths for rectangular slabs, the same range for circular radii, both with thicknesses from 2 to 40 mm. Depending on the application, the most important considerations are typically transparency, refractive index and dispersion. An indication of the variety of glasses available may be seen in the Abbe diagram published by the German manufacturer Schott, reproduced in Figure 3.11. Plotted on axes of refractive index n_d and Abbe number ν_d , both using the the yellow Fraunhofer d-line (587.6 nm) as a reference (recall the definitions in Section 3.2.2), each point represents a standard glass product, grouped into areas with comparable composition.

As an example, BK7 is a popular standard Schott glass, relatively inexpensive and thus often used as a substrate for many micro-optical components or assemblies; this glass may be found on the chart at the intersection of $n_d = 1.51680$ and $\nu_d = 64.17$.

The label “BK7”, incidentally, is the manufacturer’s designation for borosilicate crown glass #7, BK being the abbreviation for the German “Bor-Kron”. There are a number of other high-quality optical glass manufacturers world-wide, including Corning in the US and Nippon Sheet Glass in Japan, and each uses its own catalogue designation for a particular glass. Six-digit numbers in which the first three derive from the refractive index and the last three from the Abbe number are popular.

From the Abbe diagram of Figure 3.11, we see that there is a large spectrum of glasses available. For a higher index glass, for example, we could use Schott SF6, a lead-silicate glass with $n_d = 1.80518$ and $\nu_d = 25.36$. For extreme values of index or dispersion, diamond, with $n_d = 2.4173$ and $\nu_d = 55.3$, or CaF_2 with $n_d = 1.43388$ and $\nu_d = 95.36$, may be employed.

Other types of glass substrates popular amongst micro-opticians include standard soda-lime glass ($n = 1.512$), the material from which standard photolithography masks are made; Pyrex ($n = 1.472$), a borosilicate glass known to cooks¹⁶ and chemists as the material of oven-proof glassware, but also available in optical-quality substrates and very useful in micro-optical systems assembled by anodic bonding; and low-alkali aluminosilicate glass, such as Corning 1737, which is expressly designed to withstand multiple thermal cycles, such as those found in semiconductor and microsystems processing.

Sputtered layers

Whereas standardized glass substrates are often the preferred starting material for much of glass-based micro-optics, sputtered glass layers are of considerable utility in thin- and thick-film based multi-layer structures. The advantage of using sputter deposition processes is that different types of substrates (semiconductors, metals or ceramics) may be used as a support for the glass layers, and that specialized glass products with customized characteristics may be deposited.

Sputtered glass layers, most simply SiO_2 or sapphire (Al_2O_3), have long been used as deposited waveguide layers (Pitt, 1973). Advanced glass materials, such as TeO_2 , with strong electro-optic, acousto-optical and non-linear optical effects, may also be deposited as thin films (More et al., 2003) providing the basis for advanced optical and integrated optical components. We discuss dielectric films, many of which are glass-like, further in Section 3.6.1 below.

3.4.3 Characterization

When employing glass substrates for critical micro-optical applications, the characteristics of the material typically need to be known to considerable precision. The values for refractive index may usually be determined to $\pm 5 \cdot 10^{-6}$ and Abbe number to $\pm 2 \cdot 10^{-6}$; the tolerances (deviation from the desired, or specified, value) are $\pm 0.0005\%$

¹⁶ Pyrex is a Corning trademark for its glass with low thermal coefficient of expansion. Pyrex cookware is now made from soda-lime glass in the US but continues to be a borosilicate glass for products sold in the EU.

and $\pm 0.8\%$, respectively. Within a single glass sample, striae define regions where the refractive index deviates from the norm; typical homogeneity values are on the order of $\Delta n = \pm 5 \cdot 10^{-6}$ whereas $\Delta n = \pm 5 \cdot 10^{-7}$ may be specified for certain speciality glasses. The absolute index values are a critical function of the raw materials and the production processes employed, where the cooling rate of the melt is of primary importance.

Mechanically, the residual stress in a glass substrate, which may lead to undesired birefringence, is a critical function of glass blank cooling rate and the nature of the high-temperature processes the substrate has undergone. The residual birefringence which results is then characterized by an optical path length difference (in nm/cm) for different transmission directions in the crystal. Values below 10 nm/cm are standard and precision annealed glasses can have residual birefringence less than 4 nm/cm. In addition, the climatic resistance (to moisture, acids or alkalis) is specified related to ISO norms, and the Knoop hardness and viscosity may also be of interest.

Transparency

Transmittance, implying a measure for transparency and thus specifying absorption, is generally specified for a standard thickness, typically 25 mm, at a range of blue wavelengths. The wavelengths at which transmission decreases to 80% and 5% of maximum are then specified, defining the so-called UV-dip. In general, higher index glasses have the UV-dip at longer wavelengths, implying poorer transmission at higher frequencies.

Scattering in glass substrates is due to natural inhomogeneities in the material, giving rise to Rayleigh scattering as described above, but may also derive from bubbles and inclusions in the glass. The density of bubbles is generally specified using a total bubble cross-section (for bubbles larger than typically $30 \mu\text{m}$) in a given volume. Typically less than 0.03 mm^2 of total bubble area, with fewer than 10 bubbles in total, are permitted per 100 mm^3 volume for high-quality optical glass.

Thermal behavior

The thermal behavior, finally, must be well characterized for precision optical systems and is especially relevant for the use of glass in microsystems employing a variety of materials. The coefficient of linear thermal expansion, the specific heat capacity and the thermal conductivity are all factors which play a role in hybrid systems subject to varying temperature values. The refractive index and dispersion (Hoffmann et al., 1990) of glass are also a function of temperature, as we saw in Section 3.2.3 above, and these effects also need to be taken into account when designing robust glass-based optical microsystems.

Glass with extremely low coefficient of linear thermal expansion has been developed, for use in high-precision optical systems, notably telescopes (Petzoldt and Pannhorst, 1991; Bach, 1995). Zerodur is a glass ceramic developed by Schott Glass; it consists of precisely-defined crystallites in the glass matrix, which have a thermal expansion coefficient opposite to that of amorphous glass. Thermal expansion coefficients less than $2 \times 10^{-8} \text{ }^\circ\text{C}^{-1}$ in the temperature range $0 - 50^\circ\text{C}$ have been achieved. A related product, ULE (Ultra Low Expansion) glass, manufactured by Corning, is actually titanium

silicate (SiO_2 with 7.5% TiO_2) and has similarly low values of thermal expansion coefficient.

3.5 Semiconductors

Micro-optics is one facet of microsystems engineering, a widely diverse field in which one unifying aspect is the use of materials and process technologies ultimately based on semiconductor microelectronics (Gianchandani et al., 2007). As a result, not only do many of the microsystems fabrication techniques play an important role in micro-optics, but many of the materials do so as well. Thus semiconductors, which have relatively little significance in macroscopic classic optics, are considerably more important for micro-optics.

The semiconductor field is vast, as is the literature (for example, (Smith, 1978; Madelung, 1978; Ridley, 1988; Brennan, 1999)). We provide only the merest hint of an overview here to give the reader a cursory introduction to some of the relevant materials and properties.

3.5.1 Characteristics

Semiconductors are a class of generally crystalline materials whose defining characteristic is the band structure. Metals have a very high conductivity ($\sigma \sim 10^6$ $1/\Omega\text{cm}$) due to the high density of free electrons available for conduction; insulators do not conduct well ($\sigma \sim 10^{-14}$ $1/\Omega\text{cm}$) since the valence band electrons must overcome a large bandgap (10 eV or more) to reach a conduction band. Semiconductors, *nomen est omen*, have conductivity between these two extremes, generally in the range $10^{-4} \leq \sigma \leq 10$ $1/\Omega\text{cm}$, and the nature of the electronic conduction is a function of the bandgap. Their optical properties, usually of greater relevance for optical applications, are likewise defined by the details of the band structure (Zappe, 1995, Chapter 2 and references therein).

Band structure

The energy bands in a semiconductor relate the energy of electrons, W [eV], to their wave vector, \mathbf{k} . Unlike in free space or in most unstructured materials, where the relationship is given simply by

$$E = \frac{\hbar^2 |\mathbf{k}|^2}{2m_e} \quad (3.67)$$

for electron mass m_e [kg], the periodic nature of crystals gives rise to a more complex relationship, whose calculation can be quite involved (Cohen and Chelikovsky, 1988) and which is typically plotted as a band structure plot of E vs. \mathbf{k} .

As a somewhat generic example, a simplified schematic band structure of a direct bandgap semiconductor, such as GaAs, is shown in Figure 3.12. The energy bands define the allowed values of E and \mathbf{k} such that absorption and emission of energy by

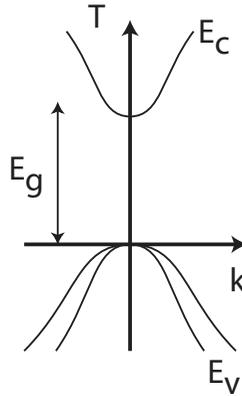


Figure 3.12 Simplified schematic band structure of a direct-bandgap semiconductor, such as GaAs. Plotted is energy E as a function of wavevector k ; the valence band, with energy E_v , has all electron states filled at low temperature and the conduction band, E_c , has all electron states empty. E_g , the energy gap, is the energy between maximum E_v and minimum E_c .

electrons must take place within the bands. At temperature $T = 0$, the valence band energy, E_v , defines the maximum electron energy and the valence band is completely occupied by electrons. The conduction band, defined by E_c , is completely empty (of electrons) at zero temperature. For electronic conduction, electrons need to be thermally or optically raised into the conduction band; at non-zero temperatures, the electron population of the conduction band then defines the electrical properties of the material.

The separation between valence and conduction bands defines the bandgap energy E_g [eV], whose value varies strongly between different kinds of materials but typically lies in the range $\sim 0.5 - 1.5$ eV for most commonly used semiconductors; for Si, for example, $E_g = 1.12$ eV and for GaAs, $E_g = 1.42$ eV, both at room temperature. Electron transitions between the valence and conduction bands define most of the optical behavior of the material. A photon with energy $h\nu \geq E_g$ will raise an electron from E_v to E_c and thus be absorbed. Alternatively, an electron falling from E_c to E_v may emit a photon, giving rise to luminescence. Finally, transitions within the conduction or valence bands can give rise to free-carrier absorption as we saw in Section 3.3.1.

Direct and indirect bandgap

An important consideration for the definition of the optical properties of semiconductors is not only the magnitude but also the position of the bandgap. In the bandgap sketch of Figure 3.12, the valence band maximum and the conduction band minimum are both at the same value of \mathbf{k} , $|\mathbf{k}| = 0$. This situation corresponds to that of a direct bandgap semiconductor, for which both absorption, but particularly emission, of photons with $E = E_g$ is very efficient. GaAs is a direct bandgap semiconductor and is thus useful for the fabrication of LEDs and laser diodes.

In some materials, notably silicon, the bandgap is indirect: the valence band maximum and the conduction band minimum are not at the same value of \mathbf{k} , so that $\Delta|\mathbf{k}| > 0$ is required for an electronic transition. As a result, electrons typically lose their energy

when falling from E_c to E_v through emission of numerous phonons, which have considerable momentum but low energy, but not photons. For that reason, silicon is a very poor emitter of light and can only be employed as a laser material using, for example, Raman effects (Rong et al., 2005, 2008), albeit with limited efficiency; we will address this issue again in Chapter 10.

3.5.2 Silicon

The most established semiconductor with arguably the most advanced manufacturing technology on the planet, silicon dominates as the substrate material for microelectronics. Its use in optics has been hampered by the indirect bandgap, just discussed above, which renders light emission inefficient; electro-optic effects are likewise weak. Silicon has, nonetheless, proven to be of value as an efficient photodetector, as a base material for refractive or diffractive micro-optics, for passive integrated optics and, possibly most significantly, as a substrate on which micro-optics, micromechanics, microfluidics and microelectronics may be combined yielding highly-integrated, multi-functional optical microsystems. We will see some examples in Chapter 11.

Silicon is generally obtained as polished wafers, whereby 200 or 300 mm (4- or 6-inch) diameters are commonly used in university research cleanrooms; wafer thickness is between 300 and 500 μm . Industry standard diameter for the microelectronics industry is 600 mm (12 inches) but the process infrastructure for these Frisbees is generally not economically feasible for low-volume optics or microsystems applications. The wafers are generally oriented in the (100) direction and are p- or n-doped; the orientation and doping are reflected in the relative positions of the wafer flats which are also used as alignment features during lithography. The surface quality on the polished side is typically excellent and double-side polished wafers are available for structuring both sides of the substrate. Silicon ingot and wafer fabrication is easily as well-established as that for glass (Pearce, 1983).

3.5.3 Compound semiconductors

Silicon is an elemental semiconductor, found in column IV of the periodic table, of which a portion is shown in Figure 3.13. It is thus compositionally a simple material and this is one of the reasons that silicon may be fabricated at such large volumes with such exceptional purity and reproducibility. A second important class of semiconductors results from combining elements from columns III and V of the periodic table, yielding so-called III-V compound semiconductors. These semiconducting alloys, while more difficult to grow as crystals and process than silicon, often have very useful optical properties and have thus become an essential material for a broad range of optical and optoelectronic applications.

Binary, ternary and quaternary materials

By combining gallium from column III and arsenic from column V in a crystal melt, GaAs results, providing an example of a binary compound semiconductor, consisting of

IIB	IIIA	IVA	VA	VIA
	B	C	N	
	Al	Si	P	
Zn	Ga	Ge	As	Se
Cd	In	Sn	Sb	Te
Hg				

Figure 3.13 A subsection of the periodic table showing the elements of primary importance for most optically-relevant semiconductors. The elemental semiconductors (Si, Ge) are found in column IV; the III-V compound semiconductors (GaAs, InP, AlGaAs, ...) are combined from elements in columns III and V; and the II-VI semiconductors (ZnSe, HgCdTe, ...) from those in columns II and VI.

two elements. GaAs is one of the grandparents of the large extended compound semiconductor family and we discuss some of its properties below. Alternatively, combining indium and phosphorous yields InP, an important substrate material, particularly for laser diodes in the near infrared. Many of these binary compound semiconductors have been the subject of considerable study for decades (Blakemore, 1982).

The properties of binary materials may be extended by incorporating two different column III or V elements into the alloy, thereby yielding a ternary compound semiconductor. A well established example is AlGaAs, (Adachi, 1985) in which aluminum and gallium from column III combine with arsenic from column V. This ternary compound is typically denoted as $\text{Al}_x\text{Ga}_{1-x}\text{As}$, where x is the mole fraction of Al. For $0 \leq x \leq 1$, the aluminum mole fraction may be varied continuously, giving different compositions of AlGaAs; as limiting cases, $x = 0$ gives GaAs and AlAs results for $x = 1$. Further examples of ternary compounds currently of interest include InGaN (used for blue laser diodes) or InGaAs (a useful photodetector material for the infrared).

As is suggested in Figure 3.13, compound materials may also be generated from elements in columns II and VI of the periodic table, leading to II-VI semiconductors. Prominent examples of useful II-VI materials are HgCdTe (the material used in many long-wavelength “night vision” goggles) or the binary ZnSe (employed for blue laser diodes, but now eclipsed by GaN-based III-Vs).

Alloy composition may be made even more complex by combining two elements each from columns III and V; a quaternary compound semiconductor results. A widely used example is $\text{In}_{1-x}\text{Ga}_x\text{As}_{1-y}\text{P}_y$, a quaternary material in which the In/Ga and As/P ratios may be varied independently. InGaAsP is used for near-infrared laser diodes employed in optical telecommunications and is just one example of a useful quaternary system (Pearsall, 1982); others include $(\text{Al}_x\text{Ga}_{1-10x})_y\text{In}_{1-y}\text{P}$, used for the laser diodes found in the popular red laser pointers.

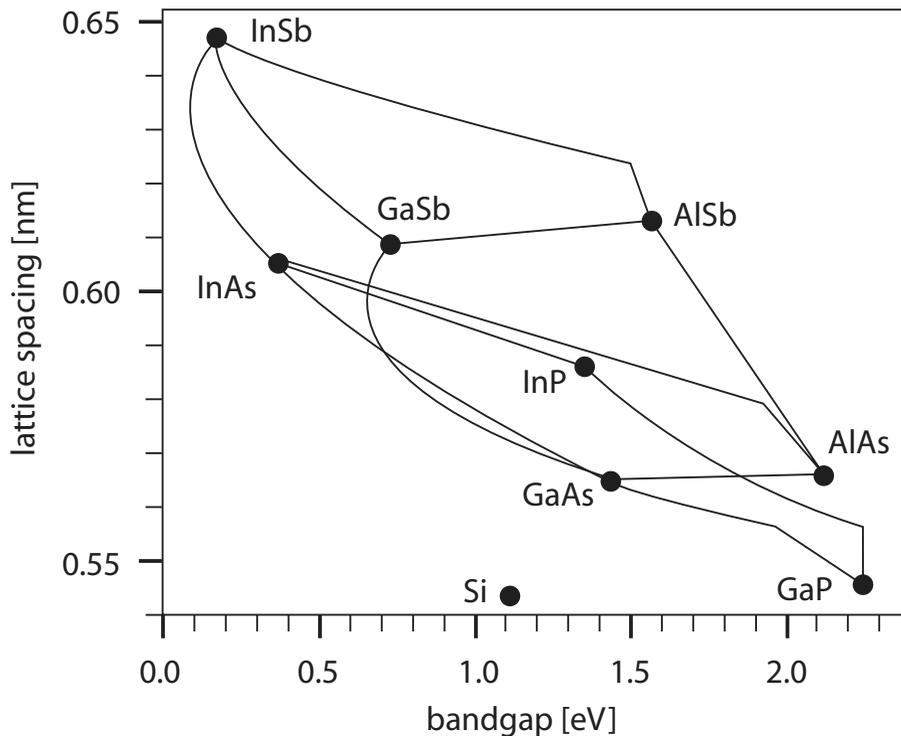


Figure 3.14 Variation of the lattice constant and energy gap, along with the associated transition wavelength, with composition for a selection of III-V compound semiconductors. The dots represent binary materials and the solid lines between them the ternary compound; thus AlGaAs is found on the line between GaAs and AlAs, lower right.

Bandgap and lattice constant

The primary reason that so much effort has been extended on the development of these complex binary, ternary and quaternary compound semiconductors is that a change in the material composition leads to change in their electrical and optical properties. Key among these are the lattice constant of the crystal unit cell, a [nm], the bandgap energy, E_g and the refractive index n . As can be seen in Figure 3.14, varying the composition of a ternary material typically results in a change of both lattice constant and energy gap. For example, InGaSb, found at the upper left of the figure, is on the line between InSb and GaSb, and we note a significant change in a and E_g as the composition changes.

Lattice constant is of relevance for crystal growth. Most III-V and II-VI semiconductors may be deposited by epitaxial crystal growth techniques, such as MBE, MOCVD or LPE (Zappe, 2004, Chapter 3.3), allowing the formation of precisely defined epitaxial (crystalline) layers with sub-nanometer accuracy on an appropriate substrate. Key to the success of this process is lattice matching: the lattice constant of the deposited material must match that of the substrate. Therefore the composition of a material can be chosen so that the atomic spacing matches that of the substrate, allowing the epitaxial depo-

Table 3.2. Typical material parameters for several commonly used semiconductors, at room temperature and near infrared wavelengths. λ_g is the bandgap wavelength, $\Delta L/\Delta L\Delta T$ the thermal coefficient of expansion and σ_{th} the thermal conductivity. Data is from (Zappe, 1995, 1999).

symbol	Si	Ge	GaAs	InP	units
E_g	1.12	0.66	1.42	1.35	eV
λ_g	1.11	1.88	0.87	0.919	μm
n	3.44	4.00	3.65	3.45	
a	0.5431	0.5657	0.5654	0.5869	nm
$\Delta L/\Delta L\Delta T$	2.6	5.8	5.7	4.6	$\cdot 10^{-6} \text{ K}^{-1}$
σ_{th}	1.45	0.6	0.46	0.68	W/cm K

sition of heterostructures, layered single-crystal structures of varying composition and bandgap.

Along with lattice constant, the bandgap energy also varies with composition. If we use AlGaAs, found on the line between GaAs and AlAs at the lower right of Figure 3.14, as an example, we see that E_g varies from 1.42 eV to 2.1 eV. A detailed look at the lattice parameter reveals that a changes by less than 0.15% between GaAs and AlAs, implying that all compositions of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ may be grown lattice-matched to GaAs; this flexibility is one of the reasons the material system has enjoyed considerable popularity. Other material systems are less obliging: $\text{In}_{1-x}\text{Ga}_x\text{As}_{1-y}\text{P}_y$, for example, used for mid-IR laser diodes, is only lattice-matched to InP substrates for $0 \leq x \leq 0.47$ and $(\text{Al}_x\text{Ga}_{10x})_y\text{In}_{1-y}\text{P}$, used for red diode lasers, is lattice matched to GaAs only for $y = 0.5$.

Calculation of the band structure in compound semiconductors is generally complex (Bastard, 1988) and a close look at the energy bands is required for a detailed materials design. In $\text{Al}_x\text{Ga}_{1-x}\text{As}$, for example, a look at the band structure shows that this material only has a direct bandgap for $x \leq 0.45$; thus the transitions for $E_g > 1.9$ eV are indirect and thus inefficient for light emission.

Characteristics

The variations in lattice constant, bandgap energy and refractive index available within the families of compound semiconductors make these useful contenders for the fabrication of a wide variety of optical devices. Numerous numerical approximations for the variation of n and E_g with composition have been derived (Zappe, 1995, Chapter 4.2), as have the variations in these parameters with temperature and wavelength. Table 3.2 summarizes a few characteristics of several important semiconductors.

The bandgap energy of the material is of primary interest in the design and fabrication of laser diodes at various wavelengths. The bandgap may be expressed in energy units (traditionally [eV]) or more practically as the emission/absorption wavelength λ_g , which is simply related to the energy using the relationship of Equation 2.65, as $E [\text{eV}] = 1.23985/\lambda_g [\text{nm}]$. In Table 3.3, we list a few compound semiconductor mate-

Table 3.3. Approximate wavelength ranges for laser diodes and detectors fabricated from several popular compound semiconductors.

material	λ_g range [nm]	application
CdZnSe	300–500	blue laser diodes
InGaN	350–700	high-density optical data storage
InGaAs	600–2000	infrared photodetectors
AlGaInP	600–700	laser diode pointers, data storage
GaAs	780–850	first generation optical data storage
InGaAsP	1300–1600	long-range optical telecommunications

rial families and the wavelength ranges for which these may be usefully employed as laser diodes or photodetectors. We refer the reader to the literature for a more comprehensive treatment (Zappe, 2004).

3.6 Other materials

A range of other materials for micro-optical and integrated optical applications has been developed (Eldada, 2004). Many of these provide advantages in functionality, ease of fabrication or compatibility with other materials in more complex optical microsystems (Suleski and Te Kolste, 2005). We briefly discuss dielectric films, polymers, sol-gels and the menagerie of specialized crystals below.

3.6.1 Dielectric films

Both the fields of microelectronics and microsystems employ a host of thin-film materials which may be deposited on substrates using a variety of means. Many of these are dielectrics, implying the materials are generally amorphous and non-conducting. Although many of these were developed for their electrical properties, their optical behavior is in many cases also of interest. Thin surface films may be used, for example, as multi-layer periodic structures, such as those used in optical interference filters, or as waveguides in integrated optical circuits.

The silicon-based dielectrics, SiO_2 and Si_3N_4 , are well established in process technology. SiO_2 is the primary constituent of glass, but may also be grown thermally on silicon, or either evaporated or sputtered on other arbitrary substrates. Also a well-standardized material, Si_3N_4 may be deposited by similar means; the oxide/nitride hybrid, SiO_xN_y is also popular. For most optical applications, the refractive index is of primary importance and its exact value can be a critical function of film deposition conditions. An excessive amount of hydrogen incorporated into a Si_3N_4 film, for example, can cause its index to be reduced from its stoichiometric value; alternatively, the ratio of oxygen to nitrogen in SiO_xN_y can be used to tune the index deliberately, from

Table 3.4. Approximate refractive indices of various dielectrics around 589 nm (Na D line); both indices are given for birefringent materials

dielectric	name	n
MgF ₂	magnesium fluoride	$n_o = 1.38, n_e = 1.39$
LiF	lithium fluoride	1.39
CaF ₂	calcium fluoride	1.43
SiO ₂	silicon dioxide	1.46
Al ₂ O ₃	sapphire	$n_o = 1.77, n_e = 1.76$
Si ₃ N ₄	silicon nitride	2.05
SiO _x N _y	silicon oxynitride	1.5 – 2.0
Ta ₂ O ₅	tantalum pentoxide	2.2
Nb ₂ O ₅	niobium oxide	2.4
TiO ₂	titanium oxide / rutile	$n_o = 2.621, n_e = 2.919$

that of oxide to nitride. The approximate refractive index at visible wavelengths for a selection of dielectrics is given in Table 3.4.

For many applications, particularly in integrated optics or in thin-film multilayer structures, a high refractive index value is desirable. Most standard dielectric films have an index near that of glass (~ 1.5); for higher index values, specialized dielectrics have to be employed. As seen in the table, TiO₂, Nb₂O₅, and Ta₂O₅ have higher index values; these materials may be deposited by sputtering onto substrates which can withstand a degree of exposure to high temperatures.

3.6.2 Polymers

Polymer materials are currently the subject of considerable research and development for many aspects of micro- and nanotechnology in general (Li and Huck, 2002) and are also of considerable and increasing relevance for micro-optics. Representing a large family of very diverse materials, polymers or plastics were originally attractive due to their low price (compared to glass) and ease of manufacturability; using molding or embossing processes, optical structures could be fabricated at high volume and low cost. The field has developed wildly in recent years, and the functionality now available within the very broad spectrum of polymer materials has greatly increased the impetus for their use in micro-optics and expanded the application possibilities. Optical functionality now includes electro-optic effects, light emission, thermo-optic response, non-linear behavior, photodetection and sometimes a combination of these. These capabilities are a long way from cheap plastic lenses for disposable film cameras, both now virtually obsolete.

Common types

Since polymers are chemically synthesized compounds, there are almost endless variations in possible compositions; indeed, many optical polymers are synthesized on

Table 3.5. Popular polymer materials, with refractive index and Abbe number at visible wavelengths, wavelength range for transparency and birefringence^a; data from (Simmons and Potter, 2000, Table 4.1), (Gale, 1998, Table 6.1) and (Yasufuku, 1992).

polymer	acronym	n	ν	λ range [nm]	birefringence
polycarbonate	PC	1.586	35	380–1600	0.106
poly methyl methacrylate	PMMA	1.491	57	400–1100	- 0.0043
polystyrene	PS	1.590	31	350–1600	-0.10
polyvinyl chloride	PVC	1.54		400–2200	0.027

^a Birefringence is defined as the index difference between the molecular chain axis and the perpendicular axis.

demand for particular applications. There are, however, a few standard polymers, well-established in the optics field. These include polycarbonate (PC), probably best known as the substrate material for compact discs and DVDs; poly methyl methacrylate (PMMA), usually better known under its various trade names plexiglass, lucite (in the USA), perspex (in the UK) or acrylic; polystyrene (PS); and polyvinyl chloride (PVC). These polymers are typically transparent in the visible and near-IR (NIR) wavelength ranges (with loss below 0.1 dB/cm) (Lytle et al., 1979), are generally moldable and are of variable robustness.

Characteristics

The refractive index of polymers is typically around 1.5 for visible wavelengths; as with glass, the refractive index decreases with increasing wavelength and Sellmeier coefficients have been determined for some popular materials (Nikolov and Ivanov, 2000). Abbe numbers are generally in the range 30–60. Transparency usually covers the visible wavelength range, but decreases sharply in the violet and UV due to polymerization effects which can take place for wavelengths below about 380 nm. In the infrared, atomic resonances (C-H vibrational modes) become relevant above 1000 nm and increase absorption for longer wavelengths. Table 3.5 gives a few detailed values for these parameters.

Most polymers have a large but negative thermo-optic coefficient ($\Delta n/\Delta T$), opposite to that of glass, such that they are advantageously used in glass/polymer hybrid optical systems for athermalization, compensation for temperature variations in the optical behavior. Intrinsic birefringence is usually very low but stress-induced birefringence can be significant. Specially engineered polymers can have electro-optic coefficients which very large, so that these materials are attractive for use as optical modulators, particularly integrated optical ones.

Mechanical properties

Polymers tend to be less environmentally robust than most glasses and semiconductors: thermal stability, scratch resistance and chemical sensitivity are all worse for these materials. However, their malleability allows them to be structured using techniques not

available to the more rigid counterparts: injection molding, hot embossing and laser patterning are fabrication tools with which polymer-based micro-optics may be generated (Heckele and Schomburg, 2004; Kopitkovas et al., 2006). Machining and polishing, as is done in classical optics, are less suitable for plastics, and dry or wet etch techniques derived from microfabrication processes are in development.

As a result, complex surface features may relatively easily be generated in polymers, structures which are frequently difficult to generate in semiconductors or glass. Molding or embossing techniques, which we will discuss in greater detail in Chapter 11, then allow replication of these structures at large volumes and over large areas, thereby permitting low-cost fabrication of complex micro-optical components.

Most polymer materials are purchased as granulate or in liquid form, and then deposited by spin coating or formed into thicker blocks. Solid substrates, rectangular or circular, are also available, not least of which are the CD blanks fabricated by the tens or hundreds of millions. The micro-optician who has a chemistry group as a research partner often benefits from access to custom-synthesized polymers with specified characteristics, often materials not available on the open market.

3.6.3 Sol-gels

The expression “sol-gel” refers more to a fabrication process than a particular material, but common usage now implies that sol-gels refer to colloidal suspensions which are processed to yield glass-like materials, usually as surface layers (Brinker and Scherer, 1990). There are numerous compositional variations, with distinct nomenclature, but these have in common that one begins with a liquid sol material, typically a silicon source and solvents, such as water and ethanol. Upon aging and evaporation of the solvent, this sol becomes a gel and this gel may be patterned or shaped to form an optical component; alternatively, the sol itself may be deposited into a mold. The gel is subsequently annealed at high temperature to form a glass. By changing the composition of the sol, the mechanical and optical properties, including refractive index and dispersion, may be controlled to a significant degree. Considerable experience in the synthesis of the sol is necessary such that, as with polymers, opticians and chemists collaborate advantageously in this area.

The sols are most often silica- or titanium-based (Yuan et al., 2005); since these inorganic sols may be spun on to a surface, the term “spin-on glass” is thus frequently employed. These materials typically suffer from shrinkage, high porosity, and crack formation during annealing, such that the formation of layers thicker than about 1 μm is typically not possible, a limitation when using these materials for refractive micro-optical components (Krause et al., 2006). The incorporation of organic compounds into the mixture yields hybrid sol-gels, which may be made photo-sensitive and may be more stable for thicker layers; however, optical absorption in the UV reduces the transmission spectrum when compared to inorganic sol-gels. Such modifications can lead to sol-gels that are ceramic- or polymer-like, and these hybrid sol-gels are known variously as ormocers (organically modified ceramics) (Haas and Rose, 2003), ceramers, ormosils (organically modified silicates) or polycerams. Sol-gels with an incorporated

photo-initiator may be hardened using exposure to UV, which can speed up processing significantly.

The potential of sol-gels for micro-optics fabrication, which will be discussed further in Chapter 11 below, has been illustrated in the fabrication of refractive microlenses (Krause et al., 2006; Yuan et al., 2005), gratings (Blanc et al., 1999) and lasers (Lo et al., 1993).

3.6.4 Specialized materials

There exists a host of optical materials with specialized properties for use in applications with particular demands for which standard materials are inadequate. Most of these are fabricated in small volumes and tend to be expensive, such that their use is generally restricted to those areas in which their optical characteristics are essential. We discuss two important examples here.

LiNbO₃

A prime example is LiNbO₃, which we discussed in Section 3.2 above, a material with good birefringent, non-linear, electro-optic and acousto-optic properties (Weis and Gaylord, 1985; Binh, 2006; Wooten et al., 2000). As outlined in Example 3.3, LiNbO₃ is usually obtained as a crystalline substrate labeled as “z-cut”; when properly oriented in an optical system, the TE polarization then sees the ordinary refractive index n_o and TM the extraordinary refractive index n_e . Of primary interest is the large electro-optic coefficient of the material, which makes it attractive for use in optical modulator applications, but LiNbO₃ has also been extensively developed for use as a photodetector or laser and for a variety of integrated optical applications (Eldada, 2004).

Whereas LiNbO₃ is usually obtained as solid substrate, the size and appearance of glass cover slip at up to a million times the price, the material may also be deposited by sputtering on glass or grown epitaxially on LiTaO₃ substrates, albeit with limitations in quality and optical performance. As a substrate, LiNbO₃ is environmentally quite robust, a reason which has contributed to its popularity despite its price.

Magneto-optical materials

Whereas magnetic effects play a subordinate role in optics, magneto-optical materials are useful for the fabrication of optical isolators, components which allow transmission of light in only one direction (Dötsch et al., 2005). Magneto-optics relies on the Faraday effect: in certain materials, the application of a magnetic field results in non-reciprocal polarization rotation, meaning that the rotation direction is the same for forward and backward propagation; using an appropriate arrangement of polarizers, an optical isolator may then be conceived. We will look at these components in more detail in Section 10.6.1.

An important family of magneto-optical materials is that of the garnets, such as YIG, yttrium iron garnet (Y₃Fe₅O₁₂) or TGG, terbium gallium garnet (Tb₃Ga₅O₁₂); the former is used for NIR applications and the latter for visible wavelengths. In the same material family as YAG (yttrium aluminum garnet, used in Nd:YAG lasers), these gar-

nets are crystalline materials and may be grown as single crystals or deposited epitaxially on lattice-matched substrates such as $\text{Gd}_3\text{Ga}_5\text{O}_{12}$ (GGG). Refractive index values are in the range of 1.9 to 2.2. Numerous other types of advanced materials with interesting magneto-optical properties, including photonic crystals (Lyubchanskii et al., 2003), are being developed.

Problems

1. Show that the complex root of k^2 given in Equation 3.15 is correct.
2. Show that the units of k_R and k_I in Equations 3.18 – 3.21 are correct.
3. Show that the units of n_R and n_I in Equations 3.25 and 3.26 are correct.
4. We illuminate a 100 μm thick plate of non-magnetic material with $\epsilon = 2.25$ and conductivity $\sigma = 20 \text{ 1}/\Omega\text{m}$ using a HeNe laser, $\lambda = 633 \text{ nm}$. Determine the real and imaginary propagation constants of the material at that wavelength and calculate the fraction of the optical intensity which is transmitted as well as the phase shift induced by the plate.
5. We illuminate another 100 μm thick plate of a different non-magnetic material with $\epsilon = 2.25$ but this time with conductivity $\sigma = 5 \cdot 10^5 \text{ 1}/\Omega\text{m}$ using the same HeNe laser, $\lambda = 633 \text{ nm}$. Again determine the real and imaginary propagation constants of the material and calculate the fraction of the optical intensity which is transmitted as well as the phase shift induced by the plate. For what electromagnetic field wavelength would a transmission of about 1% be achieved?
6. Show that d_0 as given in Equation 3.38 is a solution to the force balance equation Eq. 3.37.
7. Verify that the form of the resonance approximation of Equation 3.45 may be derived from the general expression for χ for the case that $\omega \approx \omega_0$.
8. Given a semiconductor with a free electron concentration of $n_e = 10^{15} \text{ cm}^{-3}$, calculate the plasma frequency and plot the dielectric constant derived from the Drude model for the range $10 \text{ GHz} \leq f \leq 1 \text{ THz}$. What is notable about the sign of ϵ for low frequencies?
9. Estimate the Abbe number ν_d of the silicon dioxide material whose dispersion characteristic is shown in Figure 3.5. Is the value for ν_e significantly different? Is this likely to be data from crown or flint glass? Is the dispersion high?
10. Determine the thickness required for a piece of sapphire that will rotate the polarization of a linearly-polarized Ar ion laser ($\lambda = 514 \text{ nm}$) by 90° .
11. We would like to fabricate a quarter wave plate using LiNbO_3 to be used with a diode laser emitting at $\lambda = 780 \text{ nm}$. What dimensions are required? If any of these are impractical, what other dimensions would work?
12. Using the expression for non-linear susceptibility, Equation 3.63, show how second harmonic generation and parametric amplification may be accomplished using a non-linear material. *Hint:* combine two fields and look for sums of frequencies.
13. A 10 mm thick piece of material has an optical absorption coefficient of $\alpha = 13.816 \text{ cm}^{-1}$. Calculate the loss in dB/cm and determine what fraction of intensity incident onto it emerges on the other side. Is the value reasonable?

-
14. For an optical communications application, we can tolerate a reduction of a factor 1000 in optical signal intensity after transmission through 65 km of a fiber material. What value of absorption coefficient would such a material have? Express your answer in dB/km, dB/m and cm^{-1} . If glass were a candidate for the material to be used, to what wavelength range would you be limited?
 15. How much further can light be transmitted through glass using a wavelength of $\lambda = 1.55 \mu\text{m}$ rather than $\lambda = 850 \text{ nm}$? Would this explain why long-range optical telecommunications systems use $1.55 \mu\text{m}$ InGaAsP lasers?
 16. Would PMMA be an appropriate material for transmitting light from a GaAs-based laser diode?
 17. Design an optical system with which we could see both the front and back sides of a silicon wafer simultaneously. Define the wavelength and the possible imaging photodetector materials.

4 Optical interfaces

*They had perhaps some of the weird inhuman beauty of minerals, the lustre of coal, the weight and blueness and resistance of iron, the transparency of glass. Elemental creatures, weird and distorted, of the mineral world!*¹

D.H. Lawrence, 1928

Much, if not most, interesting optics happens at interfaces, the transition from one material to another. The phenomena of reflection and refraction are both interface effects, and since these are the basis for the functional behavior of almost all optical components, we must address the question of what happens when an electromagnetic wave impinges on the boundary between two materials.

We begin this chapter with a discussion of refraction, that which occurs when a beam of light strikes an interface and is deflected in its propagation direction. Based on consideration of the electromagnetic fields and their polarization across the material transition, we will then derive relationships between the reflected and transmitted parts of the optical field; both of these considerations, we will see, are to first order a function of the refractive indices of the materials on either side of the interface. After looking at the reflection and transmission of power, we will look at the important difference between external and internal incidence, and for the latter consider total internal reflection and the evanescent field.

4.1 Reflection and refraction

When a light beam is incident from a material with refractive index n_i onto an interface formed with a material of refractive index n_t , as shown in Figure 4.1, part of the light is reflected back into n_i and a portion is transmitted into n_t . Of interest in this situation are the angles of reflection and transmission as well as the fractions of the fields which are reflected and transmitted.

¹ D.H. Lawrence, *Lady Chatterley's Lover*, p. 160

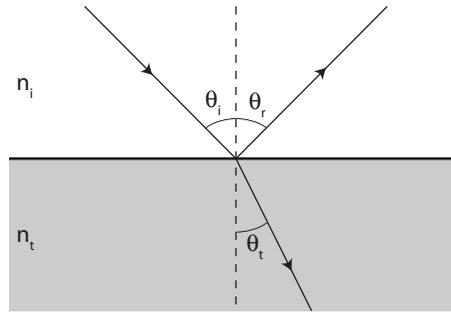


Figure 4.1 Reflection and refraction at an interface between two materials, the light incident from material with refractive index n_i and transmitted into material with refractive index n_t . The angles of incidence, reflection and transmission are measured with respect to the surface normal and given by θ_i , θ_r , and θ_t , respectively. Shown is the case of external incidence, in which the light moves from the material with lower n (*clear*) to that with higher n (*gray*).

4.1.1 Fermat's principle

Helpful for a basic understanding of reflection and refraction is Fermat's principle², which places restrictions on the optical path along which an optical wave moves. Recall, as we saw in Equation 2.49, that the speed of light in a medium v is reduced with respect to its speed in vacuum c by the refractive index of the transmissive medium, n , or

$$v = \frac{c}{n}. \quad (4.1)$$

The optical path length, L_{opt} [m], through a medium is then given by

$$L_{opt} = \int_A^B n(\mathbf{r}) \cdot d\mathbf{S} \quad (4.2)$$

where the integration is along a contour $d\mathbf{S}$ from point A to B along which $n(\mathbf{r})$ in general varies with position.

Fermat's principle states that light moves (and bends) in such a manner that L_{opt} is minimized; this expression is sometimes known as the "principle of least time." It is often explained by considering that all the waves that take the shortest optical path length will be in phase, and thus constructively interfere, whereas any which arrive by more circuitous routes are out of phase and thus tend to cancel each other out.

The implication of Fermat's principle is that light will bend to traverse optically more dense (i.e., those with higher n) regions more quickly. The principle may be employed to understand heuristically how light "bends" through a refractive lens but also explains effects well known in nature even to non-opticians. Mirages in the desert, in which

² Pierre de Fermat (1601–1665), the Frenchman known as *le prince des amateurs*, an amateur poet, linguist and mathematician. He never published his extensive mathematical developments and his contributions to science are known from letters to friends or scribbles in the margins of books. His famous "last theorem", one such a margin scribble, kept leagues of mathematicians occupied until 1993.



Figure 4.2 Due to reflection of sunlight from a hot surface air layer, the cyclist in Kyoto's Imperial Palace, at the center of the photo, looks as if she were riding on water.

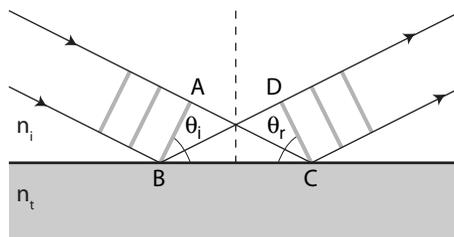


Figure 4.3 Reflection of an optical field from a surface. The wavefronts (gray) are normal to the propagation direction. Incidence is at an angle θ_i and reflection at an angle θ_r .

objects located below the horizon are seen in the distance, is due to light being bent into a region of hot air, with a lower refractive index, present along the earth's surface. The sun we admire (with suitable eye protection) during a particularly romantic sunset has actually set long before, but the orange orb is visible above the horizon since light is bent into the optically denser regions of the atmosphere. And the "wet roadway" effect noticeable on hot days, as shown in Figure 4.2, results from sunlight being channeled through a region of hot, low refractive index air near the earth's surface, thereby essentially reflecting an image of the sky.

4.1.2 Reflection

Light incident onto an interface at an angle θ_i [rad], as shown in Figure 4.1, is partially reflected at an angle θ_r [rad]. Fermat's principle tells us that the light beams, in the homogeneous incident medium with n_i , move in a straight line. If we consider the wavefronts, normal to the propagation direction as shown in Figure 4.3, of the incident and reflected beams, AB and CD, respectively, we see that these impinge and reflect at angles θ_i and θ_r , respectively, onto and from the surface.

For the wavefronts to be maintained, we can see from the geometry that the length AC must equal the length BD. Considering the two right triangles ABC and BCD, we find that

$$\sin \theta_i = \frac{AC}{BC} \quad (4.3)$$

$$\sin \theta_r = \frac{BD}{BC}. \quad (4.4)$$

Since $AC = BD$, we thus have $\sin \theta_i = \sin \theta_r$ or $\theta_i = \theta_r$. Therefore, the angle of reflection is equal to the angle of incidence. This relationship holds for any planar reflecting surface and is independent of the values of n_i and n_t .

4.1.3 Refraction

The other portion of the light incident onto the interface at an angle θ_i shown in Figure 4.1 is transmitted into the material with n_t at an angle θ_t [rad]; as seen in the figure, the propagation direction changes after traversing the interface, the phenomenon known as refraction. Considering the geometry of the wavefronts again, shown in Figure 4.4, we see that the *optical* path lengths AC and BD must be identical for the wavefronts to be maintained. Since these paths are in two different materials, we have

$$n_i AC = n_t BD. \quad (4.5)$$

Considering again the right triangles ABC and BCD, which share the segment BC, we find that

$$\sin \theta_i = \frac{AC}{BC} \quad (4.6)$$

$$\sin \theta_t = \frac{BD}{BC}. \quad (4.7)$$

Combining the previous three equations, we can write

$$n_i BC \sin \theta_i = n_t BC \sin \theta_t \quad (4.8)$$

or finally

$$n_i \sin \theta_i = n_t \sin \theta_t. \quad (4.9)$$

The previous expression is generally known as Snell's law³ and is of considerable relevance in calculating the transmission of light through optical surfaces, as we will see below.

³ Willebrord Snell (or Snel) (1591–1626), professor of mathematics in Leiden, The Netherlands, is remembered today for his discovery of the law of refraction, but was possibly more renowned in his day for his contributions to geodesy, due to which he was known as the “father of triangulation.”

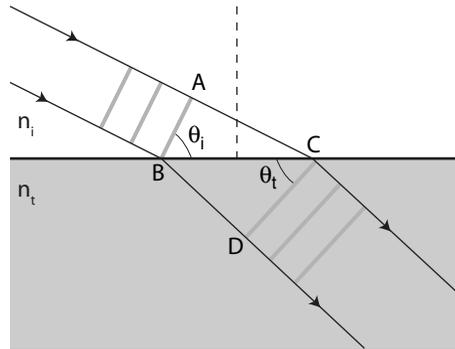


Figure 4.4 Refraction of an optical field at an interface. The wavefronts (gray) are normal to the propagation direction. Incidence is at an angle θ_i and transmission of the refracted beam is at an angle θ_t .

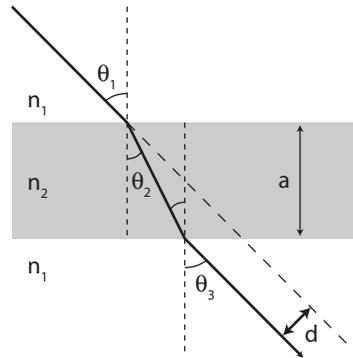


Figure 4.5 Light beam incident from a medium with refractive index n_1 onto a slab of glass with thickness a and index n_2 at an angle θ_1 .

Example 4.1: Consider the plate of glass, with index n_2 and of thickness a , as shown in Figure 4.5. Show that the beam transmitted through the entire slab propagates at the same angle as the incident angle, θ_1 , and calculate the lateral shift of the beam.

The problem is solved in a straightforward manner using Snell and some geometry. If we apply Snell's law to determine θ_2 , we find that

$$\theta_2 = \arcsin\left(\frac{n_1}{n_2} \sin \theta_1\right)$$

and applying Snell again at the bottom interface then yields

$$\theta_3 = \arcsin\left(\frac{n_2}{n_1} \sin\left(\arcsin\left(\frac{n_1}{n_2} \sin \theta_1\right)\right)\right) = \theta_1.$$

From basic geometry, we find that the path length through the glass slab (the hypotenuse of two right triangles) is given by

$$\frac{d}{\sin(\theta_1 - \theta_2)} = \frac{a}{\cos \theta_2}$$

such that the lateral shift of the beam, d , is easily determined to be

$$d = \frac{a \sin(\theta_1 - \theta_2)}{\cos \theta_2}.$$

For the case where the outside medium is air, $n_1 = 1$, we can employ Snell's law to write the lateral shift only as a function of the slab index n_2 and the incident angle θ_1 , as

$$d = a \left(1 - \frac{\cos \theta_1}{\sqrt{n_2^2 - \sin^2 \theta_1}} \right) \sin \theta_1.$$

4.2 Reflected and transmitted fields

The law of reflection and Snell's law tell us the angles at which reflected and transmitted beams propagate beyond an interface. Equally of interest is knowing what fractions of the field are reflected and transmitted; to determine this, we need to employ an electromagnetic model for fields at an interface.

4.2.1 Electromagnetic waves at boundaries

To calculate what happens to an electromagnetic wave at an interface, we apply the boundary conditions well known from basic electrostatics (Kraus and Carver, 1992; Sadiku, 2001). Given a boundary between two materials with refractive indices n_1 and n_2 , the tangential components (thus parallel to the interface) of the electric field \mathbf{E} must be continuous across the boundary,

$$E_{t1} = E_{t2} \quad (4.10)$$

as must the normal components of the electric flux density,

$$D_{n1} = D_{n2}. \quad (4.11)$$

From the definition of electric flux density, the latter equation implies

$$n_1 E_{n1} = n_2 E_{n2}. \quad (4.12)$$

Similar conditions result for the magnetic fields: the tangential components of \mathbf{H} are continuous across the boundary,

$$H_{t1} = H_{t2} \quad (4.13)$$

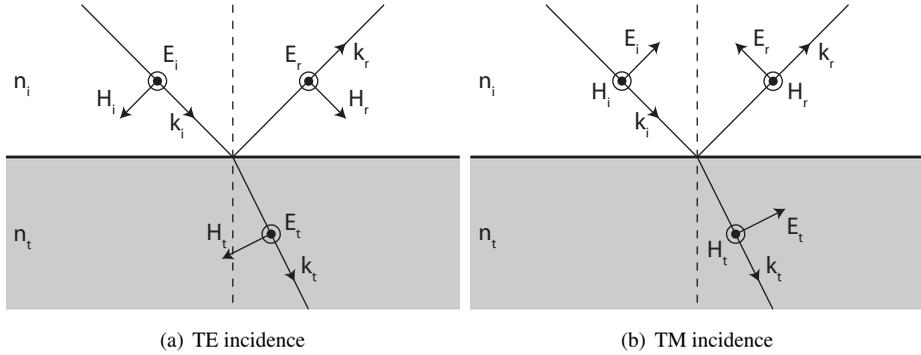


Figure 4.6 Electric and magnetic field orientations for transverse electric (TE) and transverse magnetic (TM) polarizations.

as are the normal components of the magnetic flux density,

$$B_{n1} = B_{n2}, \quad (4.14)$$

implying that

$$n_1 H_{n1} = n_2 H_{n2}. \quad (4.15)$$

Polarization conditions

We can define two useful special cases for the orientation, or polarization, of the electric and magnetic fields, realizing that any arbitrarily polarized wave can be decomposed into a sum of these two cases; these are shown in Figure 4.6. For the transverse electric, or TE case (Figure 4.6(a)), the electric field is always parallel to the interface and the orientation of the magnetic field varies. In contrast, for the transverse magnetic, or TM case (Figure 4.6(b)), the magnetic field remains parallel to the interface.

The two special cases TE and TM may be combined to represent an arbitrary polarization of the incident electromagnetic field; TE and TM fields will appear repeatedly in our optical analyses. As we will see presently, the form of the transmitted and reflected fields are different for these two cases.

4.2.2 The Fresnel equations

To calculate the magnitudes of the transmitted and reflected fields at an interface, we can consider a boundary between two materials, with refractive indices n_i and n_t for the incident and transmitted sides, respectively, as we had in Figure 4.1. Considering the case for TE polarization, as defined in Figure 4.6(a), we realize that the electric field is always tangential to the interface and obtain from the boundary conditions

$$E_i + E_r = E_t \quad (4.16)$$

for the incident, reflected and transmitted components of \mathbf{E} . For the magnetic field, it is the tangential *components* of the field, $\mathbf{H}_{||}$, which are continuous across the boundary, such that

$$H_i|_{||} + H_r|_{||} = H_t|_{||} \quad (4.17)$$

which, considering the directions of H and the component of the magnetic field parallel to the interface, yields

$$H_i \cos \theta_i - H_r \cos \theta_r = H_t \cos \theta_t. \quad (4.18)$$

Recalling that E and H are related by the impedance Z , which we defined in Equation 2.57, such that

$$H = \frac{1}{Z} E = \sqrt{\frac{\epsilon_m}{\mu_m}} E, \quad (4.19)$$

we can re-write Equation 4.18 as

$$\sqrt{\frac{\epsilon_i}{\mu_i}} E_i \cos \theta_i - \sqrt{\frac{\epsilon_r}{\mu_r}} E_r \cos \theta_r = \sqrt{\frac{\epsilon_t}{\mu_t}} E_t \cos \theta_t. \quad (4.20)$$

Realizing that $\theta_i = \theta_r$, $\mu_i = \mu_r = \mu_t$, $\epsilon_i = \epsilon_r$, and recalling that, for non magnetic material, $\sqrt{\epsilon} = n$, we can simplify Equation 4.20 to read

$$n_i E_i \cos \theta_i - n_i E_r \cos \theta_i = n_t E_t \cos \theta_t \quad (4.21)$$

or

$$n_i (E_i - E_r) \cos \theta_i = n_t E_t \cos \theta_t. \quad (4.22)$$

Combining the two boundary conditions, Equations 4.16 and 4.22, to replace E_t , we obtain

$$n_i (E_i - E_r) \cos \theta_i = n_t (E_i + E_r) \cos \theta_t \quad (4.23)$$

$$E_i (n_i \cos \theta_i - n_t \cos \theta_t) = E_r (n_i \cos \theta_i + n_t \cos \theta_t). \quad (4.24)$$

From this last expression, we can take the ratio E_r/E_i , which corresponds to a reflection coefficient for the TE electric field, r_{TE} . Likewise, solving the above expressions for E_i and E_t and taking the ratio E_t/E_i allows definition of a transmission coefficient, t_{TE} . Repeating the whole process for TM polarization yields the coefficients r_{TM} and t_{TM} . Collectively, these relationships are known as the Fresnel equations⁴ and are given for TE polarization by

$$r_{TE} = \frac{E_r}{E_i} = \frac{n_i \cos \theta_i - n_t \cos \theta_t}{n_i \cos \theta_i + n_t \cos \theta_t} \quad (4.25)$$

$$t_{TE} = \frac{E_t}{E_i} = \frac{2n_i \cos \theta_i}{n_i \cos \theta_i + n_t \cos \theta_t} \quad (4.26)$$

⁴ Jean Augustin Fresnel (1788–1827), an engineer at the École des Ponts et Chaussées, lost his post under Napoleon due to his royalist leanings, but then had time to work on his wave theory of light. His segmented “Fresnel” lenses were developed for and are still used in lighthouses.

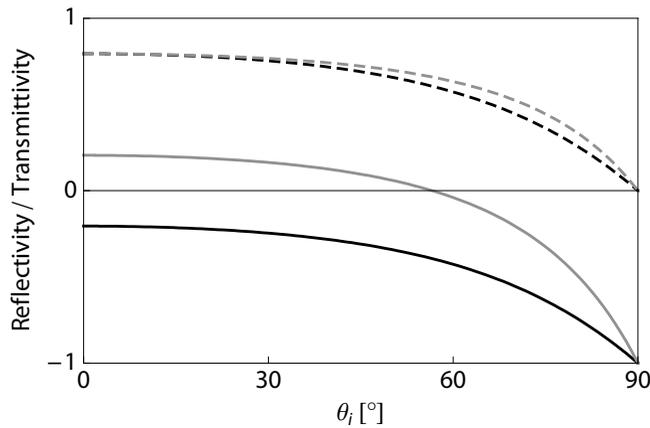


Figure 4.7 Reflectivity *solid* and transmittivity *dashed* for TE *black* and TM *gray* polarizations as a function of incidence angle θ_i for incidence from air, $n = 1$, to BK7 glass, $n = 1.5168$.

and for TM polarization by

$$r_{TM} = \frac{E_r}{E_i} = \frac{n_t \cos \theta_i - n_i \cos \theta_t}{n_t \cos \theta_i + n_i \cos \theta_t} \quad (4.27)$$

$$t_{TM} = \frac{E_t}{E_i} = \frac{2n_i \cos \theta_i}{n_t \cos \theta_i + n_i \cos \theta_t}. \quad (4.28)$$

The Fresnel equations allow us to determine the fractions of the incident field which are reflected from and transmitted across an interface. It is important to emphasize that these relationships are for the fields, not the intensity or power, to which we will return in Section 4.3 below. The reflection coefficient gives the *reflectivity* of an interface, whereas the transmission coefficient defines its *transmittivity*.

4.2.3 Reflection and transmission behavior

The Fresnel equations, Equations 4.25 – 4.28, show that r and t vary as a function of the refractive indices on either side of the interface and the angle of incidence, θ_i ; the angle of transmission, θ_t , can easily be determined from Snell's law, Equation 4.9. Figure 4.7 shows the variation of reflection and transmission coefficients as a function of θ_i for a very frequently employed interface, air ($n = 1$) to BK7 glass ($n = 1.5168$). The example shown is one of external incidence, meaning the light is incident from the material with lower refractive index, $n_i < n_t$; we will discuss the relevance of external and internal incidence in Section 4.4.

We see from the figure that transmittivity t varies from a maximum at $\theta_i = 0$ to zero at $\theta_i = 90^\circ$ with a small difference between TE and TM. For the reflectivity r , TE and TM are significantly different, but both converge to a value -1 at $\theta_i = 90^\circ$. We see thus for large incidence angles, corresponding to grazing incidence as $\theta_i \rightarrow 90^\circ$, all of the light is reflected and none transmitted.

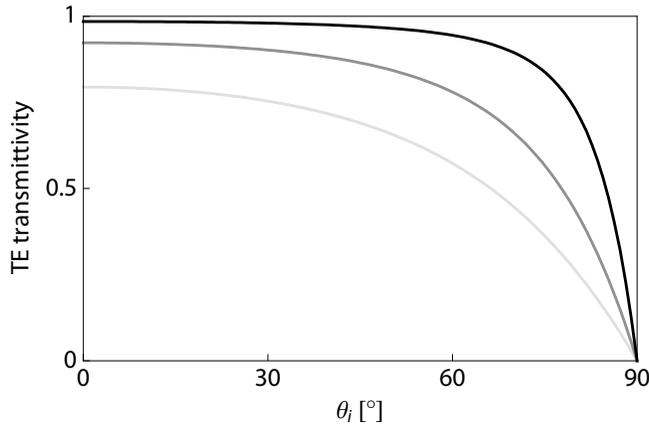


Figure 4.8 Reflectivity for external incidence onto BK7 glass ($n_t = 1.5168$): from air ($n_i = 1$) light gray, water ($n_i = 1.3318$) gray, and Pyrex ($n_i = 1.472$) black.

Example 4.2: How does TE transmittivity into a block of BK7 glass ($n_t = 1.5168$) as a function of θ_i vary if we change the medium from which we couple the light? Let us consider air ($n_i = 1$), water ($n_i = 1.3318$ at 650 nm and 20° C (Lynch and Livingston, 2001, Chapter 4.3)), and Pyrex ($n_i = 1.472$).

We simply apply the Fresnel equation for t_{TE} , Equation 4.26, determine θ_t using Snell's law (Equation 4.9) and plot the transmittivity. The result, shown in Figure 4.8, shows clearly that the refractive indices of both the incident and transmitted media play a role in determining transmission. For a large index step (such as from air), maximum transmittivity is limited and decreases rapidly with increasing incidence angle. Transmittivity from water is higher, and from another glass with a close index value, such as Pyrex, t_{TE} is close to unity for a large θ_i range.

Phase shift

For a range of incidence angle values, the reflectivity characteristics in Figure 4.7 have $r < 0$. Whereas a positive value for r implies a phase shift of zero, a negative value for reflectivity implies a phase shift of π upon reflection. This latter point can be seen simply if we consider the propagation of the incident wave to have the form

$$E_i = E_0 e^{-jkz}. \quad (4.29)$$

If $r < 0$, the reflected wave must be $-E$, and we see that a phase shift of π achieves exactly this, namely

$$E_r = rE_i = |r|E_0 e^{-jkz + j\pi} = |r|E_0 e^{-jkz} e^{j\pi} = -|r|E_0 e^{-jkz}. \quad (4.30)$$

We see from the reflectivity characteristic that reflection of the TE component always leads to a phase shift of π and TM reflection has a transition between phase shift 0

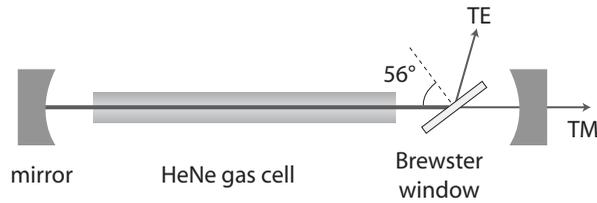


Figure 4.9 Schematic diagram of the inside of a typical HeNe laser; the Brewster window is used to assure lasing in a single linear polarization direction.

and π . As an aside, reflection from a metallic mirror surface, for which these Fresnel relationships do not generally apply, is *always* accompanied by a phase shift of π .

Polarization angle

A further look at Figure 4.7 reveals that TM reflectivity crosses through zero. For a particular incidence angle, termed the polarization or Brewster⁵ angle, the TM polarization is completely transmitted, whereas a portion of the TE polarization is still reflected; this effect can be used to polarize an optical beam, hence the designation polarization angle.

Since we are considering here external incidence ($n_i < n_t$), we designate the polarization angle θ_{pe} to distinguish it from the polarization angle derived for internal incidence, θ_{pi} , which has the same definition but a different value and will be discussed in Section 4.4. By setting $r_{TM} = 0$ in Equation 4.27, and recalling Snell's law, we see that

$$\theta_{pe} = \arcsin \sqrt{\frac{n_t^2}{n_i^2 + n_t^2}}. \quad (4.31)$$

Example 4.3: If we open the laser tube of a HeNe laser (thereby certainly voiding the warranty; don't try this at home) and look inside, we often find a glass plate at a seemingly oddly skewed angle to the optical axis, as shown schematically in Figure 4.9. What is the function of this plate?

We measure the angle of the glass plate and find it is rotated at about 56° with respect to the optical axis. With a bit of trial and error, we find that for $n_t = 1.5$ in air, a value of $\theta_{pe} = 56^\circ$ results from Equation 4.31. At that angle, the TM polarization is transmitted without loss through the plate, whereas the TE polarization has a small reflected component, and thus a small amount of loss. Since the TE polarization thus has a slightly higher loss than the TM in the laser cavity, the laser resonates only in the polarization with lower loss, thus TM. As a result, the emission is linearly polarized. This so-called Brewster window is therefore a simple means to assure linear polarization without wastefully absorbing one polarization with a polarization filter.

⁵ David Brewster (1781-1868), an ordained Scottish minister, extensively studied the reflection and polarization of light. Also concerned with improving the optics of lighthouses, his most endearing and enduring invention is, however, the kaleidoscope.

4.3 Power transmission and reflection

In many cases, it is reflection and transmission of power, or intensity, which are of greater relevance than the electromagnetic fields; optical measurements are usually of intensity (in units of W/cm^2), not of field strength. We thus need to derive relationships between incident power and the fractions transmitted and reflected at an interface.

4.3.1 Power densities

We recall from Section 2.3.3 that the power density S [W/cm^2], also referred to as the optical intensity I [W/m^2], are proportional to the square of the electric field,

$$I = \frac{n}{Z} E^2 = \sqrt{\frac{\epsilon_m}{\mu_m}} E^2; \quad (4.32)$$

the total power, I' [W], is thus a spatial integral of the intensity over the area of illumination, or

$$I' = \int I \cdot dA = I \cdot A, \quad (4.33)$$

the latter equivalence for uniform illumination over the area A [m^2].

If we illuminate a surface with beam of constant cross-sectional area at an angle of incidence $\theta_i \neq 0$, the area illuminated is a function of the incidence angle. Referring to Figure 4.10, in which an optical beam of finite (non-zero) cross-section illuminates an area A on a surface, we see that A is an angle-dependent projection of the incident, reflected and transmitted beams. If we consider A to be constant for all three (a certain fixed spot size on the interface), the cross-sections of the beams must scale with the cosine of the incident angle so that power is conserved. As a result, the incident, reflected and transmitted powers are given by

$$I'_i = I_i A \cos \theta_i = I_i A_i \quad (4.34)$$

$$I'_r = I_r A \cos \theta_r = I_r A_r \quad (4.35)$$

$$I'_t = I_t A \cos \theta_t = I_t A_t \quad (4.36)$$

for power densities $I'_{i,r,t}$ and the corresponding areas $A_{i,r,t}$.

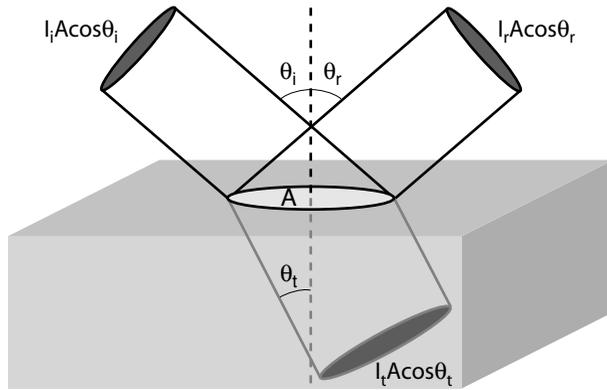


Figure 4.10 Illumination of an area A on an interface by a beam with intensity I_i incident at an angle θ_i ; the incident, reflected and transmitted beams must all project an area A onto the surface.

4.3.2 Reflectance and transmittance

From the definitions above, we can define power reflection and transmission, the reflectance R [] and transmittance T [], respectively. The former is given by

$$R = \frac{I_r A_r}{I_i A_i} = \frac{A \cos \theta_r \sqrt{\frac{\epsilon_r}{\mu_r}} E_r^2}{A \cos \theta_i \sqrt{\frac{\epsilon_i}{\mu_i}} E_i^2} = \frac{E_r^2}{E_i^2} = r^2 \quad (4.37)$$

since $\theta_i = \theta_r$ and the materials are the same for incident and reflected beams. The transmittance, on the other hand, is given by

$$T = \frac{I_t A_t}{I_i A_i} = \frac{A \cos \theta_t \sqrt{\frac{\epsilon_t}{\mu_t}} E_t^2}{A \cos \theta_i \sqrt{\frac{\epsilon_i}{\mu_i}} E_i^2} = \frac{n_t \cos \theta_t E_t^2}{n_i \cos \theta_i E_i^2} = \frac{n_t \cos \theta_t}{n_i \cos \theta_i} t^2. \quad (4.38)$$

Thus, whereas the reflectivity can simply be squared to yield the reflectance, transmittance has an additional angle- and index-dependent term which multiplies the square of transmittivity. The same considerations for TE and TM polarizations as for reflectivity and transmittivity apply to reflectance and transmittance. Since R and T are proportional to the squares of r and t , respectively, reflectance and transmittance are always positive.

R and T behavior

Plots of R and T correspond to the values that would be measured in an experiment. In Figure 4.11, for example, we plot reflectance and transmittance for TE and TM polarizations for external incidence from air onto BK7 glass. The values for R and T are always positive, as we expect. For $\theta_i = 0$, R_{TE} and R_{TM} have the same value, about 0.04. This reflectance of $\approx 4\%$ is typical for glass and is a useful rule of thumb for this material.

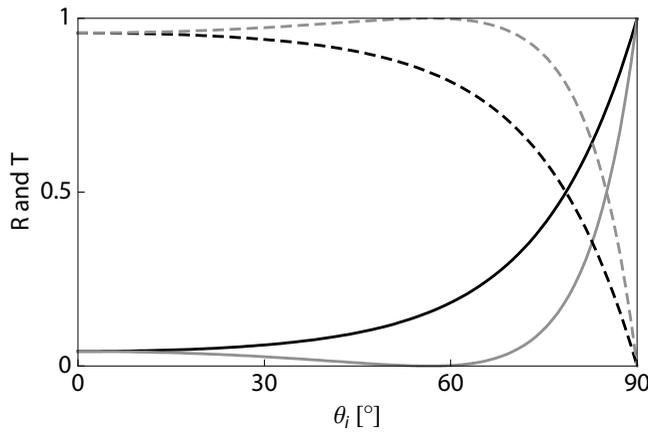


Figure 4.11 TE *black* and TM *gray* reflectance *solid* and transmittance *dashed* for incidence from air ($n_i = 1$) onto BK7 glass ($n_t = 1.5168$); note the polarization angle around 56° for TM reflectance.

From the reflectance plot, we see that reflection from glass in air is relatively small unless we illuminate at large incidence angles; this effect corresponds to our everyday experience, in which glass seems “transparent” unless we position it at grazing incidence, at which point it reflects well. The polarization angle is also seen, the point at which R_{TM} goes to zero (or $T_{TM} \rightarrow 1$), from which we find $\theta_{pe} \approx 56^\circ$.

Since we are considering powers or intensities here, conservation of energy requires that transmittivity T and reflectance R always sum to unity, or

$$R + T = 1 \quad (4.39)$$

so that plotting both is almost superfluous; we note that the sum of r and t , however, is *not* unity.

4.3.3 Normal incidence

The situation where light is incident normal to the interface ($\theta_i = 0$) is a useful special case for reflection and transmission, and occurs in a number of optical applications. We apply the Fresnel equations, Equations 4.25 – 4.28, and the definitions for reflectance and transmittance, Equations 4.37 and 4.38, to the case of normal incidence, where we see that \mathbf{E} and \mathbf{H} are both always parallel to the interface, such that the distinction between TE and TM disappears. We can thus write R and T in a simplified form, namely

$$R_{normal} = \left(\frac{n_i - n_t}{n_i + n_t} \right)^2 \quad (4.40)$$

$$T_{normal} = \frac{4n_i n_t}{(n_i + n_t)^2}. \quad (4.41)$$

The normal incidence case occurs in a number of common situations, such as transmission through a glass plate, for the coupling into a fiber or when obtaining reflection

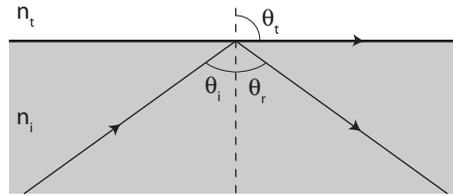


Figure 4.12 Internal incidence from material with $n_i > n_t$ at the critical angle θ_c ; $\theta_t = 90^\circ$, so that no field is transmitted beyond the interface.

from a cleaved semiconductor facet for a laser diode. We also see from the previous expressions that the distinction between internal and external incidence disappears: the reflectance is same whether we come from air onto a glass slab or from inside the glass to an air interface. The 4% rule for glass mentioned above holds only at normal incidence, but in calculating the amount reflected from a slab of glass, we have *twice* 4% reflection, once on the external surface, and once again on the internal surface.

4.4 Internal reflection

The examples outlined in the previous section were cases of external incidence, for which $n_i < n_t$; this situation generally arises when light is coupled into a material. For the case where $n_i > n_t$, we have internal incidence, which occurs when light emerges from a material or is in general coupled from region of higher to lower refractive index.

4.4.1 Critical angle

The Fresnel equations apply for both the internal and external cases, so long as we enter the correct values for n_i and n_t . Examining Snell's law, Equation 4.9, however, for the case that $n_i > n_t$, we see that $\theta_t > \theta_i$. As we thus increase θ_i , we will reach $\theta_t = 90^\circ$ for values $\theta_i < 90^\circ$. We can see that $\theta_t = 90^\circ$ is reached for an incidence angle θ_c given by

$$n_i \sin \theta_c = n_t \sin 90^\circ \quad (4.42)$$

or

$$\theta_c = \arcsin \frac{n_t}{n_i} \quad (4.43)$$

where θ_c is known as the critical angle. For angles of incidence greater than the critical angle, $\theta_i \geq \theta_c$, the transmission angle is 90° , and no light is transmitted across the interface; the effect, known as total internal reflection, is shown in Figure 4.12.

As we see from the above expression, θ_c is a function of the relative values of the refractive index on either side of the boundary. The closer n_i and n_t are to each other, the larger the critical angle. Therefore, in material systems with closely matched refractive indices (such as, for example, interfaces formed by doping in glasses, for which $\Delta n \approx$

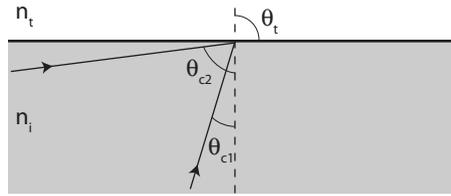


Figure 4.13 The critical angles, drawn to scale, for the internal transitions GaAs to air ($n_i = 3.45$, $n_t = 1$, yielding θ_{c1}) and inside a doped K^+ ion exchange waveguide ($n_i = 1.51$, $n_t = 1.50$, yielding θ_{c2}).

10^{-3} is typical), the critical angle is close to 90° and total internal reflection only occurs for angles approaching grazing incidence.

Example 4.4: Let us compare the critical angle for a two oft-employed material combinations: the transition from GaAs to air, as would be found for the emission of light from a III-V LED substrate; and the transition between the core and cladding of a K^+ ion exchange waveguide.

For the case of GaAs, we have $n_i = 3.45$ and $n_t = 1$, yielding a critical angle of 17° (θ_{c1} , drawn to scale, in Figure 4.13). For the K^+ ion exchange waveguide, the small difference in refractive indices, $n_i = 1.51$ and $n_t = 1.50$, due to the incorporation of K^+ ions into a section of the glass, yields $\theta_c = 83^\circ$, θ_{c2} in the figure.

These two examples demonstrate the relevance of the critical angle. Referring to the rays drawn in Figure 4.13, we see that light emission from GaAs only takes place for a narrow cone of angles with $\theta_i < \theta_{c1} = 17^\circ$. Since LEDs generate light internally with a uniform volume distribution, only a small fraction of this light is successfully emitted from the substrate; most of the light is reflected back into the material and is re-absorbed. Overcoming this inefficiency by, for example, roughening or patterning the substrate surface to allow a greater range of emission angles, is an important aspect of LED design.

The K^+ ion exchange waveguide, on the other hand, relies on total internal reflection, which we discuss in the next section, to maintain the light inside the waveguide (waveguides are the topic of Chapter 9). We see, again from the figure, that total reflection only takes place for $\theta_i \geq \theta_c = 83^\circ$, such that virtually grazing incidence is required to maintain the light in region of higher refractive index. We see thus that a large or a small critical angle is desirable, depending on the application.

4.4.2 Total internal reflection

For internal incidence at angles larger than the critical angle, we then only have a reflected component of the field, the phenomenon known as total internal reflection,

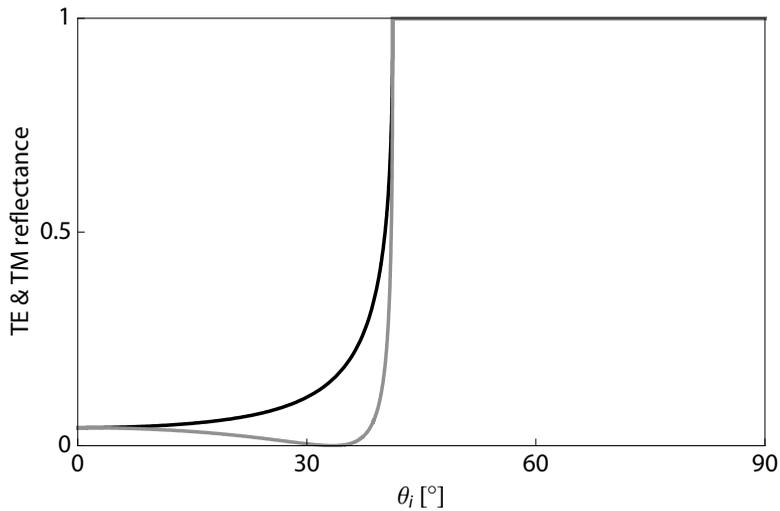


Figure 4.14 Internal reflectance R_{TE} black and R_{TM} gray for incidence from BK7 glass ($n_i = 1.5168$) to air ($n_t = 1$); the internal polarization angle has a value of about 33° .

or TIR. In the case of TIR, $r \rightarrow 1$ and the field is fully reflected back into the material. We emphasize that total internal reflection is only possible when incidence is from the optically denser medium, i.e. $n_i > n_t$.

By calculating reflectance for internal incidence, we see the effect of total internal reflection clearly. In Figure 4.14, R_{TE} and R_{TM} are plotted for incidence from inside BK7 glass onto an air interface. We see a qualitatively similar characteristic to external incidence for incidence angles below the critical angle. But at $\theta_i = \theta_c = 41^\circ$ for this material combination, R goes to unity and remains there for all higher incidence angles: all power is reflected.

The internal polarization, or Brewster, angle is also seen in the plot, since R_{TM} goes to zero at $\theta_i = \theta_{pi} = 33.4^\circ$. The same relationship used for the external incidence case, Equation 4.31, can be used to calculate θ_{pi} , but the numerical value is different; it is no accident that, for a given interface, $\theta_{pe} + \theta_{pi} = 90^\circ$.

Total internal reflection can occasionally be a problem (as with the GaAs LED in the previous example) but is most often a useful phenomenon, and is that on which guided wave optics, namely waveguides and fibers, is based; we will see how these work in Chapter 9.

4.4.3 Phase shift

We saw in Section 4.2.3 that reflection can give rise to a phase shift of the reflected field with respect to the incident field; for negative r_{TE} or r_{TM} , the phase shift is $\Delta\phi = \pi$, for positive values (for r_{TM}), $\Delta\phi = 0$.

For the case of total internal reflection, the phase shift varies in a more subtle manner. We can see from the Fresnel equations for reflectivity, Equations 4.25 and 4.27, that

setting $n_i > n_t$ for internal incidence leads to opposite polarities in the values for r when compared to external incidence: r_{TE} is always > 0 and r_{TM} is < 0 for $\theta_i < \theta_{pe}$ and > 0 for larger values. Therefore the phase shifts for internal incidence are $\Delta\phi = 0$ for r_{TE} and $\Delta\phi = \pi$ for r_{TM} when $\theta_i < \theta_{pe}$.

For internal incidence when $\theta_i > \theta_c$, total internal reflection occurs and the phase shift becomes more complicated. We can most easily see what happens by writing, for example, the expression for the TE Fresnel reflectivity of Equation 4.25 only as a function of n_i , n_t and θ_i ; we obtain

$$r_{TE} = \frac{\cos \theta_i - \sqrt{\left(\frac{n_t}{n_i}\right)^2 - \sin^2 \theta_i}}{\cos \theta_i + \sqrt{\left(\frac{n_t}{n_i}\right)^2 - \sin^2 \theta_i}}. \quad (4.44)$$

For internal incidence ($n_i > n_t$) and $\theta_i > \theta_c$, we see that the argument of the square root becomes negative, such that r_{TE} becomes complex for TIR. Taking the argument of the complex expressions yields relationships for the phase shift of the form (Born and Wolf, 1980, Chapter 1.5)

$$\Delta\phi_{TE} = 2 \arctan \left[\frac{\sqrt{\sin^2 \theta_i - \left(\frac{n_t}{n_i}\right)^2}}{\cos \theta_i} \right] \quad (4.45)$$

$$\Delta\phi_{TM} = 2 \arctan \left[\frac{\sqrt{\sin^2 \theta_i - \left(\frac{n_t}{n_i}\right)^2}}{\left(\frac{n_t}{n_i}\right)^2 \cos \theta_i} \right]. \quad (4.46)$$

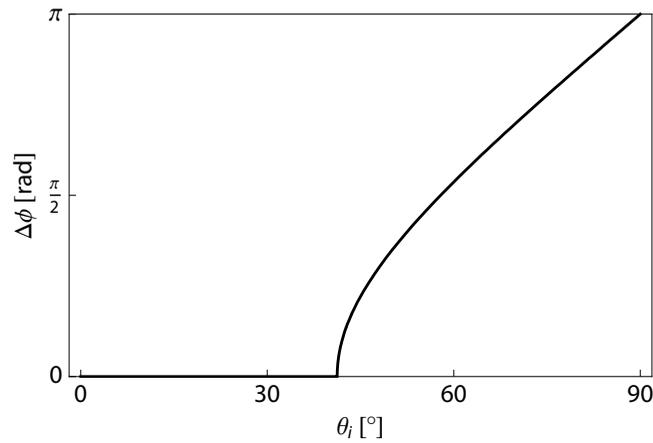
We can summarize the variation of the phase shift upon internal reflection in the plots of Figure 4.15. The distinct values of 0 or π for incidence below the critical angle are clearly seen, as is the jump in $\Delta\phi_{TM}$ at θ_{pe} . For $\theta_i > \theta_c$, there is a continuous variation of phase shift in the range $0 \leq \Delta\phi \leq \pi$.

4.5 Evanescent field

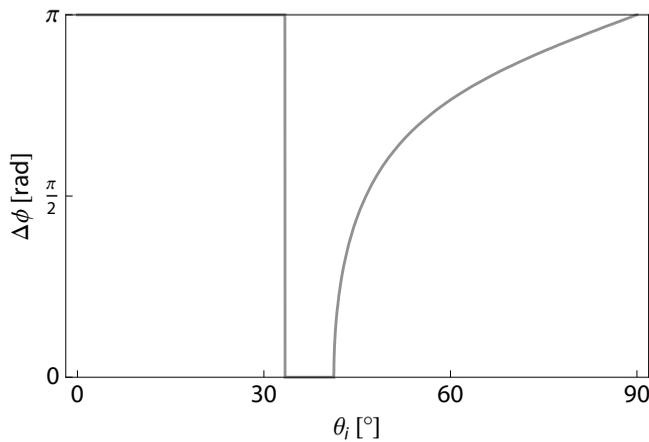
Total internal reflection, as just discussed, results in complete reflection from an interface, and thus zero transmission. Whereas no energy is transmitted across the interface in that case, detailed examination of the electromagnetic fields at the boundary shows that there is an electric field which extends beyond the interface, the evanescent field. This evanescent field decays exponentially beyond the boundary as

$$E = E_0 e^{-\alpha y}, \quad (4.47)$$

for the y -direction defined normal to the interface, as shown in Figure 4.16, and a decay constant α [m^{-1}]. Whereas a propagating field (of the form $E_0 \exp(-jk_y y)$) is found



(a) TE incidence



(b) TM incidence

Figure 4.15 Phase shift, $\Delta\phi$, for internal incidence ($n_i > n_t$) calculated for TE and TM polarizations and incidence from BK7 glass ($n_i = 1.5168$) to air ($n_t = 1$).

on the incident side of the interface (the region n_i in the figure), the exponentially decaying evanescent field is not a propagating solution. As a result, the field extends beyond the boundary but does not propagate in the y direction beyond the interface; there is thus no energy transfer into region n_t .

4.5.1 Evanescent versus propagating waves

The evanescent field decays exponentially with a decay constant α , as given in Equation 4.47. To determine the value of this constant and thus derive a measure for how far the evanescent field extends beyond the boundary, we need to consider solutions to the wave equation or electric fields across the interface.

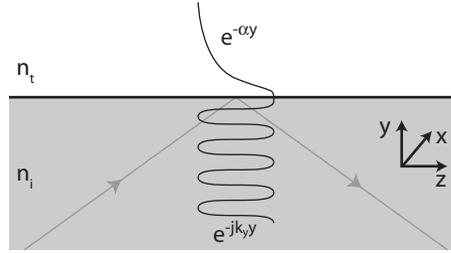


Figure 4.16 Schematic representation of the evanescent field at an interface between two materials, with internal incidence giving rise to TIR at the boundary. A propagating field is found below the interface, but an exponentially decaying evanescent field above it. No energy is transferred into the region n_t .

Helmholtz equation revisited

We recall from Chapter 2 that the wave equation, Equation 2.27, may be written in the Helmholtz, or eigenvalue, form, Equation 2.89. We may adapt this to our present situation where several materials are being considered as

$$\nabla^2 E(r) + k^2 E(r) = \nabla^2 E(r) + (n_x k_0)^2 E(r) = 0 \quad (4.48)$$

where n_x is either n_i or n_t , depending on which side of the interface we solve the equation. Referring again to Figure 4.16, we see that the refractive index only varies in the y direction, that wave propagation may take place in the y and z directions, and that everything is invariant in the x direction. We may then simplify

$$\nabla^2 E(r) \rightarrow \frac{\partial^2 E}{\partial y^2} - k_z^2 E \quad (4.49)$$

where

$$k_z = n_i k_0 \sin \theta_i, \quad (4.50)$$

the z -component of the propagation vector of the electric field incident onto the boundary at angle θ_i . The electric field propagating in the z direction, parallel to the interface, thus has the form

$$E_z = E_0 e^{-jk_z z}. \quad (4.51)$$

The eigenvalue wave equation may then be written as

$$\frac{\partial^2 E}{\partial y^2} - k_z^2 E + (n_x k_0)^2 E = 0 \quad (4.52)$$

$$\frac{\partial^2 E}{\partial y^2} + (n_x^2 k_0^2 E - k_z^2 E) = 0 \quad (4.53)$$

$$\frac{\partial^2 E}{\partial y^2} + (n_x^2 k_0^2 E - n_i^2 k_0^2 \sin^2 \theta_i E) = 0 \quad (4.54)$$

$$\frac{\partial^2 E}{\partial y^2} + k_0^2 (n_x^2 - n_i^2 \sin^2 \theta_i) E = 0. \quad (4.55)$$

The type of solution which applies to Equation 4.55 depends on the sign of $(n_x^2 - n_i^2 \sin^2 \theta_i)$: if this term is > 0 , propagating solutions of the form $E_0 \exp(-jk_y y)$ are obtained; if it is < 0 , decaying solutions, $E_0 \exp(-k_y y)$, result.

Solutions

We can solve Equation 4.55 below and above the boundary. Below the interface, $n_x = n_i$ and since $n_i^2 (1 - \sin^2 \theta_i) > 0$, we have propagating (harmonic) solutions. Above the interface, $n_x = n_t$, and, remembering that $n_i > n_t$ for internal incidence, we can show that $(n_t^2 - n_i^2 \sin^2 \theta_i) < 0$ for $\theta_i > \theta_c$. In that case, Equation 4.55 has an exponentially decaying solution of the form

$$E = E_0 e^{-k_y y} = E_0 e^{-k_0 \sqrt{n_i^2 \sin^2 \theta_i - n_t^2} y} = E_0 e^{-\alpha y}, \quad (4.56)$$

for, we reiterate, $\theta_i > \theta_c$. Thus the solution to the wave equation in the region beyond the interface for the total internal reflection condition is an exponentially decaying one, showing that the field extends into the material beyond the boundary but that there is no propagation, and thus no energy transfer, across the boundary.

4.5.2 Penetration depth

From the evanescent decay constant, α , we can easily define a penetration depth, d [m], given by

$$d = \alpha^{-1} = \left(k_0 \sqrt{n_i^2 \sin^2 \theta_i - n_t^2} \right)^{-1} \quad (4.57)$$

which represents the $1/e$ depth into the medium beyond the interface, thus the depth at which the field has decayed to $1/e$ of its maximum value. A small penetration depth implies a rapid decay of the field away from the boundary. In some applications, such as in evanescent wave sensors or some regimes of near-field microscopy, a larger penetration depth is desirable, since, although there is no energy transferred across the boundary, the evanescent field is affected by refractive index conditions beyond the boundary. We will consider these effects in Chapter 9.

Example 4.5: How do the penetration depths for incidence from a BK7 glass slab ($n_i = 1.5168$) into air ($n_t = 1$), water ($n_t = 1.3318$) and PMMA ($n_t = 1.491$) compare? Let's calculate the values for a light beam with $\lambda = 633$ nm, incident at 75° , 80° , and 85° .

Noticing that the professor was smirking slightly when he posed the problem, we are a bit suspicious and decide to quickly calculate the critical angles for the material combinations proposed. Doing so by applying Equation 4.43, we find that

$$\theta_c|_{air} = 41^\circ \quad \theta_c|_{water} = 61^\circ \quad \theta_c|_{PMMA} = 79^\circ.$$

We thus discover that the BK7/PMMA combination has a critical angle larger than 75° , so that for that incident angle, TIR does not take place and we can therefore not calculate a penetration depth. The professor has stopped smirking.

Continuing for those combinations for which the calculation makes sense, we obtain the following penetration depths d using Equation 4.57:

$$\begin{aligned}\theta_i = 75^\circ: & \quad d|_{air} = 94 \text{ nm} & \quad d|_{water} = 165 \text{ nm} \\ \theta_i = 80^\circ: & \quad d|_{air} = 91 \text{ nm} & \quad d|_{water} = 149 \text{ nm} & \quad d|_{PMMA} = 1111 \text{ nm} \\ \theta_i = 85^\circ: & \quad d|_{air} = 89 \text{ nm} & \quad d|_{water} = 141 \text{ nm} & \quad d|_{PMMA} = 411 \text{ nm}\end{aligned}$$

From these values, we see two important characteristics of the evanescent field and the penetration depth: 1) the penetration depth is larger, for a given incident angle, for material combinations in which the refractive index step ($n_i - n_t$) is smaller, so penetration is deeper into PMMA than air; and 2) the penetration depth is larger for angles of incidence close to the critical angle, and decreases as θ_i increases. For the example of PMMA, we see that d decreased from over $1 \mu\text{m}$ at $\theta_i = 80^\circ$ to around 400 nm by increasing the incidence angle by only 5° , realizing that 80° incidence is just a hair over the critical angle, 79° .

Problems

1. Derive Snell's law using Fermat's principle.
2. Using sketch, explain how a mirage works.
3. Using another sketch, explain the "wet roadway" effect seen in Figure 4.2.
4. With all this practice in sketching, make one more explaining how we see the setting sun above the horizon when it is actually already below the horizon.
5. Given a sequence of three optically transparent layers in air, with the refractive index sequence $n_{air}/n_1/n_2/n_3/n_{air}$ (where $n_1 < n_2 < n_3$), for a beam incident at an angle θ_i onto the layer n_1 , calculate the angle at which the beam exits the last layer (the one with n_3).
6. Re-write the Fresnel expression for r_{TE} (Equation 4.25) as a function of only n_i , n_t and θ_i .
7. Show that for a plate of glass at θ_{pe} , the TM polarization is really without any reflection losses (i.e., also at the internal interface of the plate).
8. Derive the expressions for θ_{pe} and θ_{pi} .
9. Show that $\theta_{pe} + \theta_{pi} = 90^\circ$.
10. Plot R_{TE} and R_{TM} for incidence from inside GaAs to an air interface as a function θ_i . What is the value of the critical angle? Can you predict which polarization will preferentially lase?
11. Plot r_{TE} and r_{TM} for internal incidence, showing that the phase shifts are the negative of those for external incidence, for $\theta_i < \theta_c$.

12. You couple white light into a glass prism; the beam exits the prism by striking the glass/air interface at 30° . Plot the transmission angle as a function of wavelength for a few values of λ . The data in Figure 3.5 or the Sellmeier formula may be helpful.
13. Given a two-layer structure surrounded on both sides by air, with the index sequence $n_{air}/n_1/n_2/n_{air}$ and with $n_1 = 1.45$ and $n_2 = 2.05$, we couple TM-polarized light into the n_1 surface at an angle of $\theta_i = NN^\circ$. Determine the fraction of the power transmitted through the two layers.
14. Repeat the previous problem, this time coupling into the n_2 surface.
15. Is it possible to couple into the surface of a slab of material and achieve total internal reflection at the other interface?
16. You are given a box full of Pyrex glass ($n = 1.472$) slides, which you can stack one on top of the other. Realizing that there is always a little air gap between slides in the stack, calculate the percentage of optical intensity transmitted at normal incidence through a stack of four slides. How many slides need to be stacked to reduce the power transmission below 50%? Ignore multiple internal reflections.
17. For a single slab of transparent material with perfectly parallel surfaces, derive an expression for the total amount of light transmitted, this time *not* ignoring multiple internal reflections. For a slab of BK7 glass, calculate the total transmission and the percent error which would result if internal reflections were ignored. Repeat for a slab of germanium.
18. For normal incidence from air, what must be the refractive index of a material be so that $R = T$?
19. Given a 2.5 mm thick piece of dyed lead-silicate glass ($n = 1.805$) with an absorption of $\alpha_{dB} = 1.2$ dB/cm which we would like to use as an attenuation filter, calculate the transmittance of the filter when used under normal incidence, considering all sources of optical loss. What would the dimensions of a filter using this material be if $T = 10\%$ were required? Is this practical? What would you change to allow for a more reasonable filter size?
20. Show that $(n_t^2 - n_i^2 \sin^2 \theta_i) < 0$ for $\theta_i > \theta_c$ and thus exponentially decaying solutions are obtained from Equation 4.55 for internal incidence from n_i to n_t .
21. Plot the variation of penetration depth as a function of incident angle (for $\theta_i > \theta_c$) for the evanescent field of a HeNe laser extending from PMMA into water.

5 Interferometry

In my model, important interference phenomena arise when individual strata come into contact. These chaotic fluctuations are, I suppose, what my music is really 'about'.¹

B. Ferneyhough, 1996

Interference is one of the most important phenomena in optics, and forms the basis for diffractive optics, one of the most relevant sub-disciplines of micro-optics. Interference effects result from the superposition of electromagnetic fields with closely matched frequency and explain many naturally-observed phenomena, such as the colorful rings generated by thin oil films on wet surfaces. For the optician, interference is a tool useful for both the development of novel optical components and for high-accuracy measurement tasks.

This chapter begins by considering coherence and interference in general, and then focuses on interferometers, optical systems which take advantage of interference effects for determination of distances, wavelengths, refractive indices and a host of related parameters. Interferometers are not only essential characterization tools for micro-optics, but are also fabricated using the technologies of micro- or integrated-optics to realize a wide variety of measurement tasks. We then turn to the optics of thin-film multilayers, optical coatings on surfaces which rely on interference to generate customizable transmission or reflection properties. By depositing layers of precisely controlled refractive index in layers at the correct thicknesses, the interference of partially transmitted and partially reflected fields can generate high or low reflection, or, for example, filters which only transmit a certain band of wavelengths. The design and fabrication of anti-reflection coatings is an important process step in the manufacture of both macro- and micro-optical components.

5.1 Coherence and interference

Interference occurs when two electromagnetic fields are superimposed. For interference effects to be observed, the two waves must be coherent, such that the coherence of an optical field is an important parameter to be considered in the ensuing analysis. We thus first consider the definition and characterization of coherence and then treat interference,

¹ Brian Ferneyhough (b. 1943), English composer.

resulting in relatively simple relationships for the intensities which result when two coherent fields meet.

5.1.1 Coherence

Coherence is a broad topic, and has been extensively considered in the literature; see for example (Saleh and Teich, 1991, Chapter 10) or (Hecht, 2002, Chapter 12). We can summarize here by stating that two waves are said to be coherent if

- their wavelength is stable;
- their relative phase is fixed; and
- their polarization is identical

over a given period of time. Whereas the wavelengths of two interfering waves need not be identical, they should be closely matched; coherence implies, however, that the wavelength difference is fixed.

Most light sources are incoherent. Thermal emitters, such as the incandescent light bulb, emit unpolarized photons in a wide spectrum of frequencies with no fixed phase relationship between them. As a result, the emitted fields are incoherent and interference effects are only visible in a very limited regime.

Coherence length and time

An ideally coherent light source would be monochromatic and have a fixed phase relationship between all parts of the wave train for $-\infty \leq t \leq \infty$. Since no light source is truly monochromatic and no waves exist for an infinite time, there is no such thing as absolute coherence: we can only speak of relative levels of coherence and the transition between a “coherent” and an “incoherent” source is a continuous one. Useful for this characterization are the coherence time, t_c [s], and coherence length, L_c [m], whereby the two are related simply as

$$t_c = \frac{L_c}{c} \quad (5.1)$$

for c the speed of light. t_c defines a time interval over which two waves are coherent and L_c the corresponding spatial distance. For (non-existent) ideal perfect coherence, t_c and L_c would both be infinite.

The temporal coherence, given by t_c , defines the time over which the frequency is stable and for which the phase is known at any point along the wave train. The spatial coherence, as defined by L_c , gives a measure for the path length difference two coherent waves can travel and still interfere, or, alternatively, the allowed spacing between two coherent sources that still gives rise to interference.

Why are coherence length and time not infinite? Electromagnetic fields are ultimately generated by atomic transitions. The relevant atomic states do not have zero lifetime, such that the finite transition time constants (typically in the nanosecond to picosecond

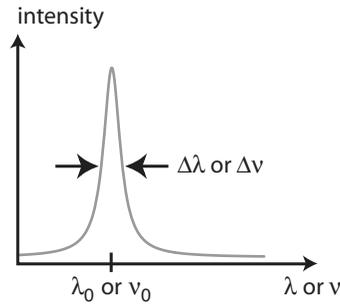


Figure 5.1 Linewidth of a monochromatic electromagnetic wave. The abscissa can be given in frequency (ν) or wavelength (λ), such that the wave has a center frequency or wavelength ν_0 or λ_0 and the corresponding linewidth is $\Delta\nu$ or $\Delta\lambda$.

range) imply, through the Fourier transform², that there is a finite spread in emission frequency. In addition, thermal vibrations and movement of the emitter may give rise to a Doppler shift in emission frequency. Finally, pulses are of finite length and we know, again through the Fourier transform, that a pulse length Δt gives rise to a non-zero frequency bandwidth $\Delta f \propto \Delta t^{-1}$.

Linewidth

All of these effects combine to result in an emission spectrum, even for a “monochromatic” source, which can be shown to have a Lorentzian intensity distribution (Zappe, 2004, Chapter 2.3), given in frequency ν by

$$I(\nu) \propto \frac{\Delta\nu}{(\nu - \nu_0)^2 + \Delta\nu^2}, \quad (5.2)$$

where ν_0 is the center emission frequency and $\Delta\nu$ the linewidth; this relationship is shown in Figure 5.1. We may plot the emission intensity either as a function of wavelength λ or frequency ν , yet in either case we see that every optical field contains a spectrum of wavelengths to an extent given by the linewidth $\Delta\lambda$ [m] or $\Delta\nu$ [Hz], which define the width (usually full-width at half maximum, FWHM) of the Lorentzian line shape. Linewidths in the two units are easily related as

$$\Delta\lambda = \frac{\lambda_0^2 \Delta\nu}{c} \quad (5.3)$$

for center wavelength λ_0 .

The linewidth is related to the coherence length of the optical wave by

$$L_c = \frac{\lambda_0^2}{\Delta\lambda} \quad (5.4)$$

² Jean Baptiste Joseph Fourier (1768–1830), the French mathematician who discovered the series and integral theorems which today bear his name, was orphaned as a young boy and just barely managed to keep his head during the revolution. His mathematical tools were developed to aid in the solution of boundary value problems arising in the analysis of heat transfer in arbitrarily shaped bodies.

Table 5.1. Linewidth $\Delta\lambda$, coherence length L_c and coherence time t_c of several popular light sources

light source	linewidth ($\Delta\lambda$)	L_c	t_c
Hg arc lamp at $\lambda = 546$ nm	1.6 nm	186 μm	0.6 ps
LED at $\lambda = 690$ nm	0.5 nm	952 μm	3.2 ps
diode laser at $\lambda = 780$ nm	0.02 pm	30.4 m	101 ns

such that a narrow linewidth implies a large coherence length and a long coherence time. We see thus that the concept of coherence is subject to some interpretation. An incandescent light bulb, with $\Delta\nu$ in the range of 100s of THz, is termed to be incoherent. A laser, which might have $\Delta\nu \approx 10$ MHz, is considered coherent. A Hg arc lamp, finally, with $\Delta\nu \approx 500$ GHz, is at best “semi-coherent,” an ill-defined term. A few examples of linewidths and the associated coherence lengths are given in Table 5.1.

Example 5.1: Let us examine the emission of one HeNe laser with $\lambda = 632.816$ nm and another with a blue shift of 2 pm; both have a coherence length of 30 cm. Plot the emission linewidth as a function of frequency and determine the linewidth in Hz.

We can determine the linewidth directly from

$$\Delta\nu = \frac{c}{L_c} = 1 \text{ GHz.}$$

The emission intensity as a function of frequency may be plotted by applying Equation 5.2 directly, converting our wavelengths to frequency by $\nu = c/\lambda$. We have $\lambda_1 = 632.816$ nm and the blue-shifted version $\lambda_2 = 632.814$ nm. From these we determine $\nu_1 = 474.071$ THz and $\nu_2 = 474.073$ THz.

Plugging these values into Equation 5.2, we obtain the two emission spectra of Figure 5.2. We see that the small blue-shift of 2 pm results in a spectrum which is shifted to higher frequencies by an amount less than the linewidth of the emission.

5.1.2 Interference

If two or more electromagnetic waves meet at a point in space, the linearity of the wave equation implies that the total field at that point is equal to the sum of the individual fields, namely

$$E_{tot} = \sum_i E_i(\mathbf{r}, t). \quad (5.5)$$

Thus we can simply add the fields; adding the fields of coherent waves leads to the phenomenon of interference (Hecht, 2002, Chapter 9).

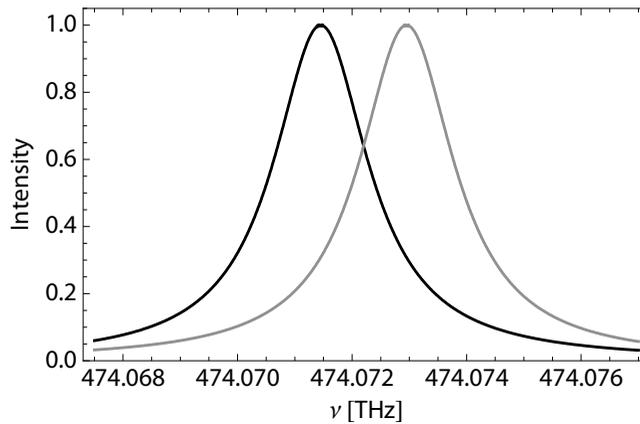


Figure 5.2 The emission spectra of two closely spaced wavelengths, $\lambda_1 = 632.816$ nm *black* and $\lambda_2 = 632.814$ nm *gray*. Both have a linewidth of 1 GHz and the spectra are plotted as a function of frequency, ν .

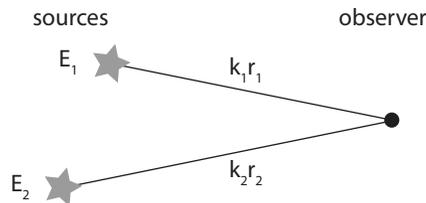


Figure 5.3 The superposition of two coherent sources with differing paths kr which meet at an observation point.

If two incoherent waves meet in space, the fields will add to generate an intensity pattern which averages to zero over time scales longer than the period of the waves, since there is no fixed phase or wavelength relationship between them. In this case no interference takes place and the total intensity may be easily calculated as the sum of the individual intensities of the waves.

For two waves of the same polarization which do have a fixed phase and frequency relationship, and are thus coherent, the fields add to form maxima or minima, stable in space and or time. These interference effects may not be calculated by adding the intensities, but require a coherent addition of the fields.

Sum of fields

We can see what happens if we add two one-dimensional fields with the same frequency ω but different values for kr , perhaps due to differing paths taken to get to the observation point, as schematically suggested in Figure 5.3. We thus have

$$E_1 = E_{10} \sin(\omega t + k_1 r_1) \quad (5.6)$$

$$E_2 = E_{10} \sin(\omega t + k_2 r_2) \quad (5.7)$$

and

$$E_t = E_1 + E_2. \quad (5.8)$$

From the trigonometric identity

$$\sin(\alpha + \beta) = \sin \alpha \cos \beta + \cos \alpha \sin \beta \quad (5.9)$$

we can write the total field as

$$E_T = E_{10}(\sin \omega t \cos k_1 r_1 + \cos \omega t \sin k_1 r_1) + E_{20}(\sin \omega t \cos k_2 r_2 + \cos \omega t \sin k_2 r_2) \quad (5.10)$$

$$= (E_{10} \cos k_1 r_1 + E_{20} \cos k_2 r_2) \sin \omega t + (E_{10} \sin k_1 r_1 + E_{20} \sin k_2 r_2) \cos \omega t \quad (5.11)$$

$$= E_0 \cos kr \sin \omega t + E_0 \sin kr \cos \omega t \quad (5.12)$$

$$= E_0 \sin(kr + \omega t). \quad (5.13)$$

We thus see that a single sinusoid results, but with a magnitude E_0 still unknown. To evaluate this parameter, we realize that we had implicitly defined

$$E_0 \cos kr = E_{10} \cos k_1 r_1 + E_{20} \cos k_2 r_2 \quad (5.14)$$

$$E_0 \sin kr = E_{10} \sin k_1 r_1 + E_{20} \sin k_2 r_2; \quad (5.15)$$

we can square and add the left- and right-hand sides to yield

$$E_0^2(\cos^2 kr + \sin^2 kr) = E_{10}^2 + E_{20}^2 + 2E_{10}E_{20} \cos k_1 r_1 \cos k_2 r_2 + 2E_{10}E_{20} \sin k_1 r_1 \sin k_2 r_2 \quad (5.16)$$

$$E_0^2 = E_{10}^2 + E_{20}^2 + 2E_{10}E_{20} \cos(k_1 r_1 - k_2 r_2) \quad (5.17)$$

$$E_0^2 = E_{10}^2 + E_{20}^2 + 2E_{10}E_{20} \cos \Delta\phi \quad (5.18)$$

for the phase difference between the two waves, $\Delta\phi$ [rad], given by

$$\Delta\phi = k_1 r_1 - k_2 r_2. \quad (5.19)$$

We thus finally have an expression for the total field, comprising the superposition of the two interfering fields, given as

$$E_T = \sqrt{E_{10}^2 + E_{20}^2 + 2E_{10}E_{20} \cos \Delta\phi} \sin(kr + \omega t). \quad (5.20)$$

The sum of two coherent electric fields, phase shifted by $\Delta\phi$ with respect to each other, therefore yields a harmonic function (a sinusoid) of the same temporal frequency ω , with a spatial frequency k which may be found from Equations 5.14 or 5.15 and a magnitude, given by Equation 5.20, which is a function of the original field magnitudes and also, critically, as we shall see presently, of the phase shift between the two interfering waves.

Sum of intensities

Recall that optical intensities are typically measured in an experiment, and that $I \propto E^2$. By squaring the field given by Equation 5.20, we see that the intensity of the superposed fields, I_T , takes the form

$$I_T = I_1 + I_2 + I_{12} \quad (5.21)$$

$$= I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \Delta\phi. \quad (5.22)$$

The intensity of the field generated by interference is thus not only a function of the field intensities, but also of the phase shift between the fields.

We can usefully distinguish between two special cases. Constructive interference occurs for $\Delta\phi = 0, \pm 2\pi, \pm 4\pi, \dots$, and yields maximum intensity

$$I_T|_{max} = I_1 + I_2 + 2\sqrt{I_1 I_2} \quad (5.23)$$

whereas destructive interference, which occurs for $\Delta\phi = \pm\pi, \pm 3\pi, \pm 5\pi, \dots$, gives rise to minimum intensity

$$I_T|_{min} = I_1 + I_2 - 2\sqrt{I_1 I_2}. \quad (5.24)$$

For the case where the interfering fields have the same magnitude, $E_{10} = E_{20}$ such that $I_1 = I_2 \equiv I_0$, we have a useful simplified form for the interference intensity

$$I_T = 2I_0(1 + \cos \Delta\phi) \quad (5.25)$$

$$= 4I_0 \cos^2 \left(\frac{\Delta\phi}{2} \right) \quad (5.26)$$

in which the role of the phase shift becomes prominent.

A plot of the interference intensity as a function of the phase shift $\Delta\phi$ is given in Figure 5.4. We see clearly the constructive interference for even multiples of π and destructive interference for odd multiples of π . This \cos^2 relationship represents the classical interference pattern for two coherent waves of identical frequency with variable phase shift between them.

From the definition of the phase shift, Equation 5.19, we see that $\Delta\phi \neq 0$ can have several origins. If two waves with identical spatial and temporal frequencies take differing paths to reach the point of interference, we have $\Delta\phi = k(r_1 - r_2)$, such that it is the path length difference which gives rise to $\Delta\phi$. Alternatively, if the same physical path lengths are traversed through differing materials, leading to different optical path lengths, we have $\Delta\phi = kr(n_1 - n_2)$.

Example 5.2: If we take the two HeNe laser lines we examined in the previous example, $\lambda_1 = 632.816$ nm and the one blue shifted by 2 pm at $\lambda_2 = 632.812$ nm, how does the interference of these two emission lines vary as a function of time? How about the interference of $\lambda_1 = 632.816$ nm with one at $\lambda_3 = 660.0$ nm?

We determine the temporal frequency from the wavelength through

$$\omega = \frac{2\pi c}{\lambda}$$

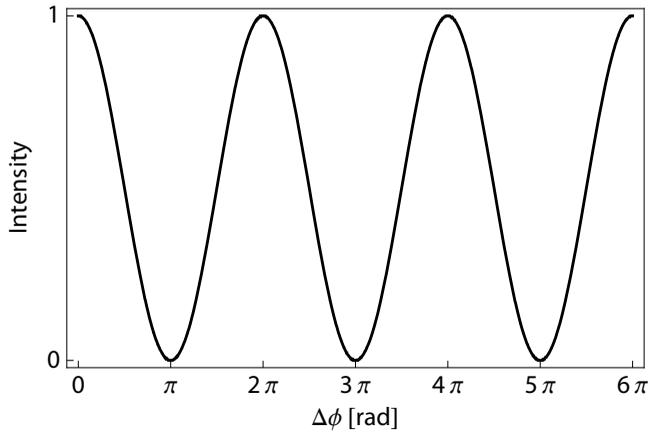


Figure 5.4 Variation of the interference intensity as a function of the phase shift $\Delta\phi$, as given by Equation 5.26. This example has visibility $V = 1$.

and then add the field quantities (assuming identical magnitudes) as

$$E_T = E_0 \cos(\omega_1 t) + E_0 \cos(\omega_2 t).$$

It can easily be shown by trigonometry that this expression is equivalent to

$$E_T = 2E_0 \cos \left[\frac{1}{2}(\omega_1 + \omega_2)t \right] \cos \left[\frac{1}{2}(\omega_1 - \omega_2)t \right].$$

We see that this expression contains sum frequency ($\omega_1 + \omega_2$) and difference frequency ($\omega_1 - \omega_2$) terms.

For the closely spaced wavelengths ($\lambda_1 = 632.816$ nm and $\lambda_2 = 632.814$ nm), we obtain the interference characteristic (normalized to unity) as a function of time as shown in Figure 5.5(a). The frequency sum term is much larger than the difference term in the above equation, so that, whereas the envelope is clearly visible, the sum frequency oscillations are at such high a frequency that the curve appears to be solidly filled. We see that there is an interference maximum roughly every 0.65 ns, such that this “beat” frequency is relatively easy to measure.

For the more widely spaced wavelengths ($\lambda_1 = 632.816$ nm and $\lambda_3 = 660.0$ nm), the interference characteristic of Figure 5.5(b) reveals both the sum and difference frequencies; the envelope and higher frequency oscillations are both clearly seen. The maxima are spaced at about 50 fs, a much more difficult value to measure than the previous one. When the wavelength spacing becomes even larger, the envelope function becomes increasingly difficult to see, such that the interference phenomenon begins to disappear. Measurement of the beat frequency is often easier than that of optical frequencies, so that taking the interference between two waves, one of known frequency, in this manner is a popular means to measure an unknown frequency.

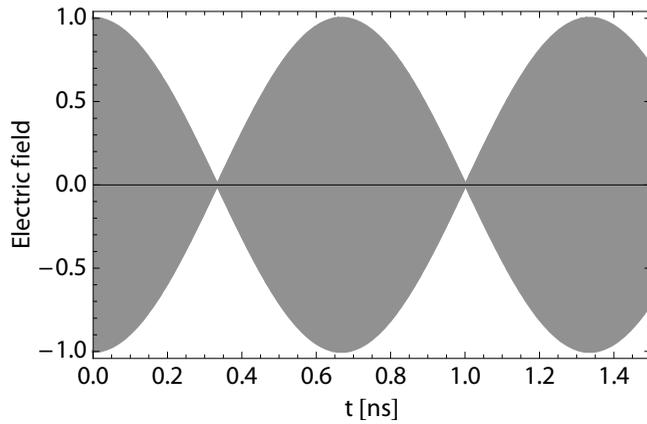
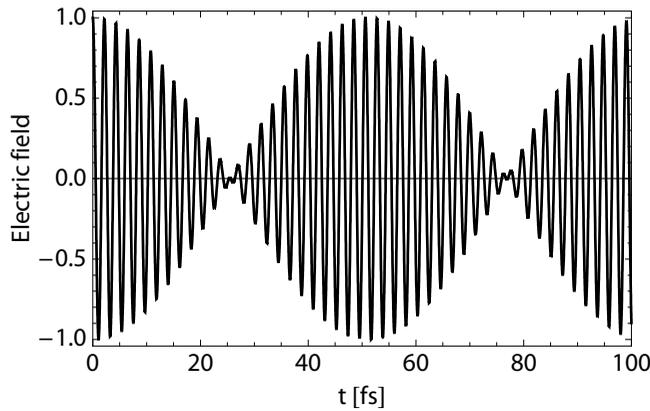
(a) $\lambda_1 = 632.816$ nm and $\lambda_2 = 632.814$ nm(b) $\lambda_1 = 632.816$ nm and $\lambda_3 = 660.0$ nm

Figure 5.5 Interference of two fields, normalized to unity, with slightly different wavelengths, plotted as a function of time. The sum frequency is too high to be resolved in (a).

Example 5.3: Let us take an optical field of frequency ω and split it into two paths, both with the same propagation constant k but taking two different routes to reach the point where they interfere, one with length z , the other with length $z + \Delta z$. What does the interference look like as a function of path length difference, Δz ?

We have two fields,

$$E_1 = E_0 \sin[\omega t + kz]$$

$$E_2 = E_0 \sin[\omega t + k(z + \Delta z)]$$

whose sum is given by

$$E_T = E_1 + E_2 = 2E_0 \sin \left[\omega t + kz + \frac{k\Delta z}{2} \right] \cos \left[\frac{k\Delta z}{2} \right].$$

The sine defines the normal oscillation of the field in time, but the cosine term varies only with the path length difference, $\frac{k\Delta z}{2}$. Remembering that $k = \frac{2\pi}{\lambda}$, we see that the intensity varies as

$$I \propto \cos^2 \left[\frac{\pi\Delta z}{\lambda} \right]$$

which has a maximum for path length differences which lead to integer π arguments for the cosine. Therefore, interference maxima (constructive interference) occur for $\Delta z = \lambda, 2\lambda, 3\lambda, \dots$ and interference minima (destructive interference) for $\Delta z = \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \dots$. Thus path length differences of a whole wavelength are required to generate interference maxima.

Visibility

Equation 5.22 shows that the interference at any given point in space gives rise to an intensity variation as a function of relative phase as well as the intensities of the interfering fields. A useful parameter for characterizing the quality of the interference is the visibility, V [], which relates the maximum (constructive) to minimum (destructive) interference values as

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}, \quad (5.27)$$

and may be expressed as a function of the input intensities I_1 and I_2 as

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}. \quad (5.28)$$

Visibility may be interpreted as contrast: a maximum value of $V = 1$ implies that destructive interference is complete (yielding intensity zero) and that contrast between minimum and maximum intensities is maximized. On the other hand, a value of V approaching zero means that the interference pattern becomes increasingly invisible since the minimum and maximum values approach each other.

Implications

Interference is an easily seen and exceedingly useful phenomenon. Although some experimental configurations employ the interference intensity at a single point, in many properly aligned and calibrated experimental arrangements, interference will give rise to two- or three-dimensional patterns of intensity maxima and minima, called fringes. These fringes, an example of which is shown in Figure 5.6, may be used as a sensitive measure for surface profiles, refractive index distributions, vibrations, thickness variations and a host of other parameters. We will see a variety of these in the following section.

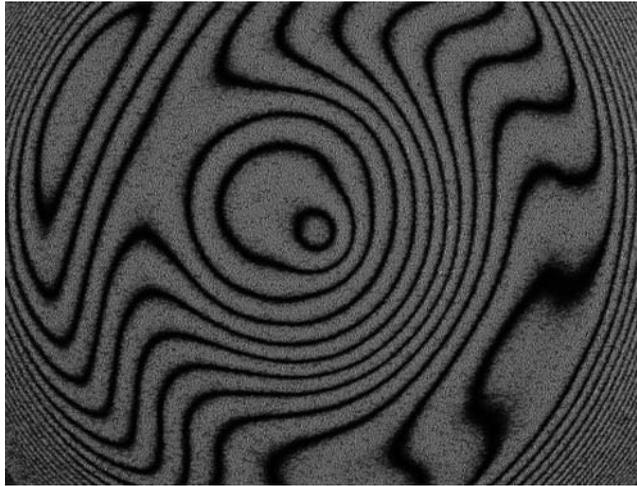


Figure 5.6 Interference fringes generated by coherent illumination of a 4-inch diameter Pyrex glass plate, as is often used as a substrate for micro-optical devices. Illumination is with a HeNe laser ($\lambda = 633 \text{ nm}$) so each fringe corresponds to a height change of λ ; the plate is clearly thicker in the center, with a circularly symmetric profile. The measurement was done using a shear plate arrangement. Measurement courtesy of Bernd Aatz.

As an aside, the randomly-moving intensity patterns which are observed (using appropriate eye protection) when an optical setup is illuminated with an expanded laser beam are perhaps the first place that optics students encounter interference. These so-called speckles are observed due to reflections of a coherent light source at different positions, so that in the eye, interference takes place between waves which have variable optical path lengths. Since, in the laboratory situation, the fields are transmitted through air, which is an optically inhomogeneous medium due to tiny temperature, density and velocity variations, the path lengths change on time scales of fractions of a second; thus the interference conditions change, and the intensity seems to “move” with time. Tiny movements of the detector (our eyes) also give rise to changes in the interference conditions, hence these speckle patterns which seem to dance in front of us in the laboratory.

Speckles are not only entertaining, they are also the basis for interferometry for measuring movement or deformations of a surface. In general, interference provides the basis for interferometers and thin-film interference filters, the subjects of the next two sections, as well as diffractive optics (Chapter 8), many types of integrated optical structures (Chapter 9) and explains the resonance in laser cavities (Chapter 10). Coherent optics, which uses interference of coherent waves to generate intensity minima and maxima as desired, is an important subset of micro-optics, since many micro-optical structures rely on interference for their function.

5.2 Interferometers

Interference effects form the basis for a large family of optical systems, namely the interferometers (Hariharan, 2003) (Yu and Yang, 1997, Chapter 9). Interferometers have been implemented at all levels of optics, macroscopically, using micro-optics and in integrated optics, and are used in a wide variety of applications, particularly for sensing and measurement tasks. The very compact and cost-effective interferometer systems which may now be conceived using the technologies of micro-optics have led to their use in areas in which their size and cost would previously have been prohibitive. Whereas we will focus on free-space micro-optical configurations here, in Section 9.7.2 we will examine waveguide-based integrated optical implementations.

The high sensitivity of interferometric systems to a host of measurement quantities, from displacement to wavelength to chemical concentrations, is the primary reason that they represent indispensable tools not only for optics but also for a broad range of sensor and metrology tasks. We will examine some of these applications below and in the problems at the end of the chapter, and thereby show how the different interferometer configurations have been optimized for a wide variety of measurements.

5.2.1 Young interferometer

Thomas Young³ was one of the first to demonstrate interference experimentally, performing his famous double-slit experiment in 1801. The configuration is shown in Figure 5.7 and should be familiar to any student of physics. The Young interferometer is essentially based on the double-slit arrangement, whereby an interference pattern is generated due to the superposition of coherent waves emitted from the apertures.

The interference pattern generated by the Young interferometer consists of a series of lines on the screen or detector; defining a normal extending from the aperture to the screen, maxima are seen at angles

$$\theta_{max} = \frac{m\lambda_0}{a} \quad (5.29)$$

from the normal for illumination wavelength λ_0 , slit spacing a [m] and integer m . For a spacing L [m] between aperture and screen, the spacing between fringes (interference maxima) Δs [m] is given by

$$\Delta s = \frac{L}{a} \lambda_0, \quad (5.30)$$

such that increasing the spacing L leads to larger fringe spacing.

³ Thomas Young (1773–1829), the English physicist, physician and Egyptologist, one of the great minds of his time, followed a truly phenomenal spectrum of intellectual pursuits. Capable in 13 languages by the age of 14, he not only analyzed the physics of the eye and helped translate the Rosetta stone, he also provided the first elegant experimental evidence for the wave nature of light.

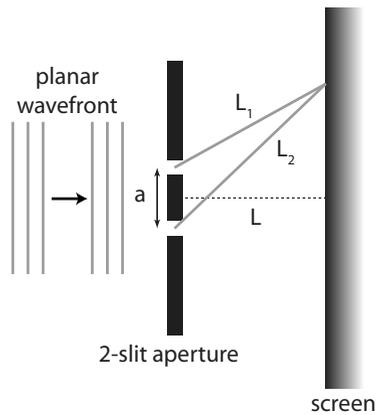


Figure 5.7 The Young interferometer, based on interference of waves emitted from two closely-spaced apertures illuminated by a planar wavefront incident from the left. The interference pattern is seen on the screen placed a distance L from the apertures.

The intensity pattern on the screen may be derived from consideration of the superposition of the coherent fields emitted from the slits and varies as

$$I = 4I_0 \cos^2 \left[\frac{k_0(L_2 - L_1)}{2} \right] \quad (5.31)$$

for spacing between the two slits and the screen L_1 [m] and L_2 [m] as shown in the figure; L , L_1 , L_2 and a may all be determined from one another by simple geometry.

The paths L_1 and L_2 are optical path lengths. If the media in or beyond the two slits have different refractive indices, or these change in response to a measurement variable, the interference conditions change and the fringe pattern will shift. This effect is taken advantage of when using the Young interferometer as a sensor: the two-dimensional fringe pattern is a sensitive function of the refractive index conditions in the optical paths through the slits and has thus been used as a refractive index sensor (Brandenburg, 1997).

5.2.2 Michelson and related interferometers

The Michelson⁴ interferometer belongs to the family of multiple beam interferometers, in which two or more beams are separated, traverse different optical paths and recombine at a screen or detector at which point interference takes place. In the Michelson configuration, shown schematically in Figure 5.8, a focused optical beam is divided into two using a beam splitter. One of the beams impinges on a fixed reference mirror and is reflected back into the beam splitter. The second beam is directed onto an

⁴ Albert Michelson (1852–1931), Polish-born American physicist, spent most of his career on the development of high-precision interferometric measurement techniques, by means of which he accurately measured the speed of light and dispensed with the aether. In 1907, he became the first American to be awarded a Nobel prize.

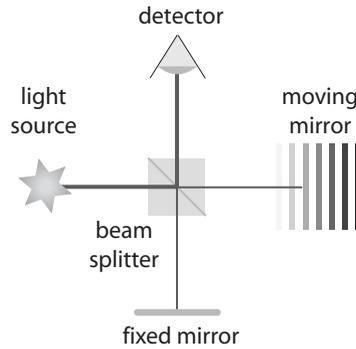


Figure 5.8 The Michelson interferometer: a focused optical beam is split into two paths by a beam splitter, one of which is reflected from a reference mirror, the other, typically, from a moving mirror. Interference between the two beams takes place at the detector.

unknown mirror, typically moving but possibly also at an unknown position, and this second reflection is superimposed on the first at a screen or detector.

The Michelson interferometer takes advantage of shifts in interference due to a change in optical path length, as we discussed in Example 5.3 of Section 5.1.2. The fringe pattern at the detector, or on a screen at the detector position, is one of concentric rings around the observer. Given a displacement of the moving mirror Δs [m], where $\Delta s = 0$ represents the balanced system (both reference and measurement arms the same length), dark rings are found at angular positions θ from the optical axis given by

$$\theta = \arccos \left[\frac{m\lambda_0}{2\Delta s} \right] \quad (5.32)$$

for operating wavelength λ in air and integer m . Essential for the observation of the fringes (or any interference effects at all) is that the displacement is smaller than the coherence length of the source, so we require

$$\Delta s \ll L_c. \quad (5.33)$$

As the magnitude of the mirror displacement approaches or exceeds the coherence length, the fringe visibility goes to zero. As a result, the difference in arm lengths of the Michelson interferometer must always be significantly smaller than L_c .

The Michelson interferometer is frequently used in a point-measurement mode, meaning that the two-dimensional fringe pattern is ignored and only the intensity of the center maximum is measured by a point detector; this configuration is implied in Figure 5.8. In this case, the interference intensity at the detector I is given by

$$I = 1 + \cos \left(\frac{4\pi n \Delta s}{\lambda_0} \right) \quad (5.34)$$

where n is the refractive index of the medium in the measurement arm. A “fringe” in this case is a transition from intensity minimum to maximum, and the number of fringes

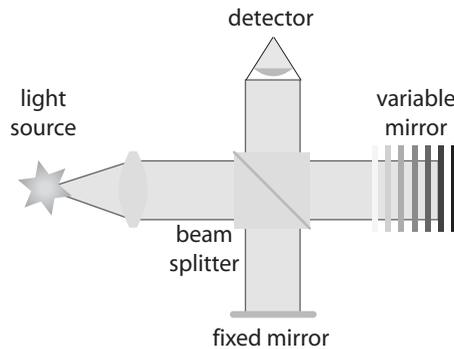


Figure 5.9 With the same basic structure as the Michelson interferometer, the Twyman-Green interferometer uses expanded beams and imaging optics.

N measured as the moving mirror is displaced from zero to Δs is

$$N = \frac{2n\Delta s}{\lambda_0}. \quad (5.35)$$

From this last expression, we see that the number of fringes observed is a function of the movement of the mirror as well as the refractive index of the medium in that arm.

As a result, counting fringes allows one to use the Michelson interferometer as a sensitive measurement tool for displacement or movement. An integrated optical implementation based on III-V semiconductors (Hofstetter et al., 1995, 1996, 1997) demonstrated a displacement resolution of about 20 nm at a measurement distance of up to 0.4 m; similar structures have also been demonstrated in silicon (Valette et al., 1990) and in glass (Lang et al., 1998). It may also be used for monitoring changes in refractive index, for two fixed mirrors. The only limitation is that the optical path-length difference ($n\Delta s$) must be significantly smaller than the coherence length L_c , as mentioned above.

Twyman-Green interferometer

There are two important variations of the Michelson interferometer arrangement which have proven to be very useful, particularly for the measurement of surface profiles and irregularities. Whereas the Michelson interferometer generally uses a collimated beam, the Twyman-Green⁵ interferometer uses the same configuration but with expanded beams and imaging optics, as shown in Figure 5.9. As a result, the optical paths have planar wavefronts and a perfectly balanced Twyman-Green interferometer shows no fringes at the detector. Since alignment of the interferometer is critical, the mirrors in macroscopic arrangements are often purposely tilted to generate visible fringe patterns, allowing an easier system calibration.

⁵ Frank Twyman (1876–1959), an English engineer known for his design of optical instruments; and Alfred Green, foreman in the optical shop of Hilger, the optical instrumentation company Twyman directed. They developed their eponymous interferometer in 1916, primarily for use in the spectrophotometry of materials.

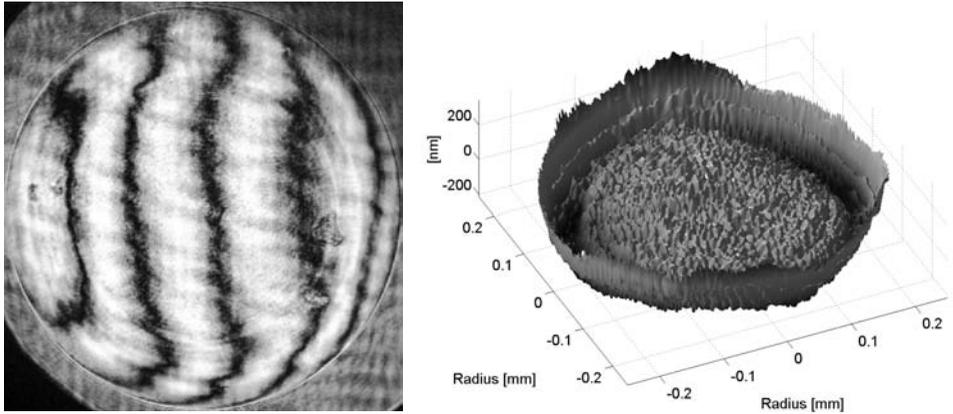


Figure 5.10 Measurement of an embossed glass lens with diameter $500\ \mu\text{m}$ and radius of curvature $950\ \mu\text{m}$ using a Twyman-Green interferometer. The lens is measured in reflection and the measured fringe pattern is on the left, the calculated surface profile of the lens on the right. Compare with Figure 5.13.

This interferometer arrangement is particularly useful for measuring deformation of surfaces or the homogeneity of materials. The Twyman-Green has been of considerable value, for example, in the analysis of microlens profiles, based on the reflection from the lens surface (Lindlein et al., 2004). In Figure 5.10 is shown the measurement of an embossed glass microlens using reflection from the lens surface in a Twyman-Green arrangement (Reichelt and Zappe, 2005). The lens, with a radius of $500\ \mu\text{m}$ and a radius of curvature of $950\ \mu\text{m}$, generates the fringe pattern seen at the left of the figure, where the fringe spacing corresponds to $\lambda_0/2$; illumination was with a HeNe laser, so $\lambda_0 = 633\ \text{nm}$. At the right of the figure is the surface profile of the lens, calculated from the fringe pattern.

Fizeau interferometer

A second variation on the Michelson arrangement is the Fizeau⁶ interferometer, shown schematically in Figure 5.11. In this case, the reference and measurement arms overlap, so that the arrangement looks a bit like a “single-armed” Michelson interferometer.

In the arrangement shown in the figure, the test surface is a transparent wedge. The back surface of the wedge, normal to the optical axis, provides a reference reflection, whereas the tilted front surface is the measurement object. For two perfect planes, the Fizeau interferometer generates parallel fringes with spacing a function of the relative angles between the planes. If there is a distortion or roughness on the tilted front surface,

⁶ Armand Hippolyte Louis Fizeau (1819–1896), scion of a wealthy French family, made the first accurate measurements of the speed of light and demonstrated that light travels more slowly in water than in air. He applied interferometry for measurement of distances and the apparent diameter of stars.

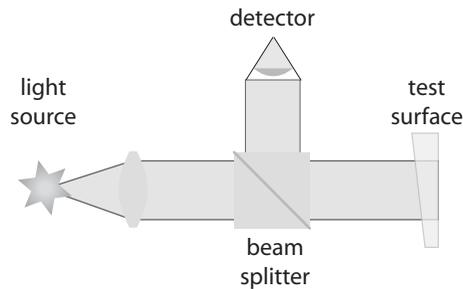


Figure 5.11 The Fizeau interferometer uses a single arm, in which both the reference and measurement surfaces are found. In this schematic, the plane of the test surface normal to the optical axis provides the reference reflection, the tilted plane is the one under test.

or a non-uniformity in the refractive index of the transparent block, these effects will manifest themselves as a distortion of the fringe pattern. Although the Fizeau interferometer is tricky to align, its simple configuration and sensitivity to surface undulations with respect to the reference plane makes it an important analysis tool, and is one of the standard means for optical testing (Malacara and DeVore, 1992; Malacara et al., 2005), also for the analysis and characterization of aspheric lenses (Asfour and Poleshchuk, 2006).

5.2.3 Mach-Zehnder interferometer

The Mach-Zehnder⁷ interferometer represents a slightly different two-beam arrangement and is shown in Figure 5.12. The input optical field is split into two arms, such that the light takes two different paths to a second beam splitter which recombines the beams and directs them onto a detector, where again interference takes place.

The two separated optical paths imply that the Mach-Zehnder interferometer is sensitive to propagation conditions in the two arms. The configuration is generally employed with one reference arm, insulated from any environmental changes, and the other a measurement arm. A change in the refractive index of the transmission medium Δn in the measurement arm, for example, gives rise to a phase shift $\Delta\phi$ in that arm, with magnitude

$$\Delta\phi = 2\pi \frac{L\Delta n}{\lambda_0} \quad (5.36)$$

for arm length L [m] (over which the index change Δn takes place) and vacuum wavelength λ_0 .

⁷ Ludwig Mach (1868–1951), son of the reputable Ernst Mach, was apparently a physicist of seriously limited talent, not to mention a drug addict, who sought to regain financial losses incurred in World War I by attacking Einstein's relativity theory and thereby forming unsavory social and political alliances in 1930s Germany (Wolters, 2005); Ludwig Zehnder (1854–1949), a Swiss professor of physics in Basel, was a student of Wilhelm Röntgen and, while at the University of Freiburg, made one of the first full-body x-ray images, a 1.84 m long radiograph of a dead soldier.

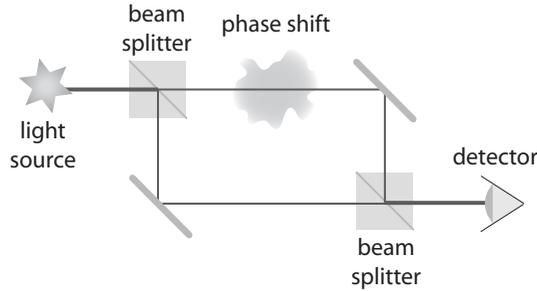


Figure 5.12 In the Mach-Zehnder interferometer, the excitation field is split into two arms, which take two different paths to a detector, where interference takes place. A change in the optical path length of one arm leads to a shift in the interference fringes.

The induced phase shift gives rise to a modulation in the interference intensity I of the form

$$I = \frac{I_0}{2}(1 + \cos \Delta\phi) = \frac{I_0}{2} \left[1 + \cos \left(\frac{2\pi L \Delta n}{\lambda_0} \right) \right] \quad (5.37)$$

such that the intensity is a sensitive function of optical path length, i.e. refractive index. We also see that the intensity shift scales with the path length through the variable medium L , so sensitive Mach-Zehnder arrangements attempt to maximize measurement arm length. As with the Michelson interferometer, however, the optical path length *difference* between measurement and reference arms must be less than L_c .

Mach-Zehnder interferometers have a wide range of applications and have been implemented macro- and micro-optically as well as in waveguide optics. Due to their sensitivity to refractive index variations, they have been used as waveguide-based chemical and biological sensors, with a refractive index resolution of $\Delta n/L = 1.3 \times 10^{-5} / \text{mm}$ (Maisenholder et al., 1997), and for pressure sensing based on deformation-induced change in refractive index of a waveguide (Zappe, 1999, Chapter 1.3.4); we will look at the former implementation in more detail in Section 9.7.2. For optical characterization, the Mach-Zehnder is useful for monitoring the shape evolution of photoresist-based microlenses (Mancebo and Bará, 1998) and provides an important alternative for the reflective measurement of microlenses using the Twyman-Green configuration; the profile of the same embossed glass microlens shown in Figure 5.10, but now measured in transmission using a Mach-Zehnder interferometer, is shown in Figure 5.13.

Other micro-optical implementations, also primarily waveguide-based, use Mach-Zehnder interferometers as high-speed optical switches (Schubert et al., 2002; Gan and Kärtner, 2005; Gu et al., 2007; Liu et al., 2007). By modulating the relative phase shift in one arm between 0 and π using an electro-optic modulator, the output intensity can be varied between maximum and minimum which, given good visibility, provides digital optical modulation. Such interferometric modulators are widely used in optical telecommunications systems.

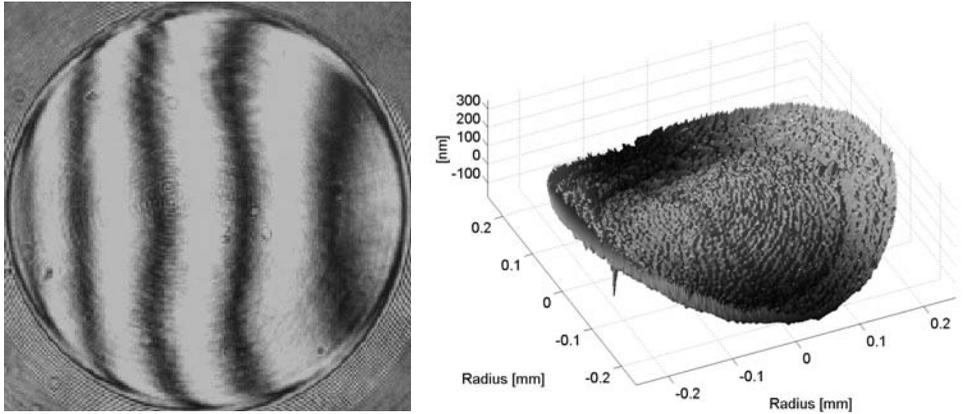


Figure 5.13 Measurement of an embossed glass lens with diameter $500\ \mu\text{m}$ and radius of curvature $950\ \mu\text{m}$ using a Mach-Zehnder interferometer. The lens is measured in transmission and the measured fringe pattern is on the left, the calculated surface profile of the lens on the right. Compare with Figure 5.10.

5.2.4 Sagnac interferometer

Whereas the interferometers discussed in the previous sections allowed determination of rectilinear motion or linear displacement, the Sagnac⁸ interferometer is sensitive to rotation. Configured as shown in Figure 5.14, the excitation field is split into two by a beam splitter and one field is directed clockwise, the other counter-clockwise around the interferometer until the two recombine again at the beam splitter and are directed to the detector.

The Sagnac arrangement rotates with angular velocity ω [rad/s], clockwise in the example shown. The two counter-propagating beams thus experience different times for one transmission around the interferometer. Using the example of a square loop, with radial dimension R as indicated in Figure 5.14, each side of the interferometer has length $\sqrt{2}R$ and the time taken for light to complete one clockwise rotation is

$$t_{cw} = 4 \left(\frac{\sqrt{2}R}{c - \frac{v}{\sqrt{2}}} \right) = \frac{8R}{\sqrt{2}c - \omega R} \quad (5.38)$$

for the movement of the interferometer corners at velocity v

$$v = \omega R. \quad (5.39)$$

⁸ Georges Sagnac (1869–1926), a French physicist and professor in Paris, demonstrated the effect that now bears his name in 1913. He was an ardent opponent of relativity, despite the fact that the Sagnac effect is consistent with it; alas, the effect is also consistent with a hypothetical aether.

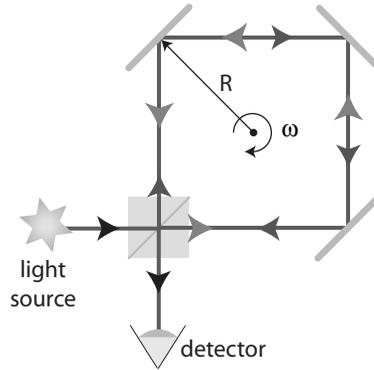


Figure 5.14 A Sagnac interferometer relies on two counter-rotating beams; these are generated by the beam splitter, which recombines the beams at the detector. The entire arrangement rotates (in this case) clockwise at angular velocity ω .

Likewise, light moving counter-clockwise requires a time

$$t_{ccw} = \frac{8R}{\sqrt{2c + \omega R}} \quad (5.40)$$

so that the time difference between the two fields at the detector is found from

$$\Delta t = t_{cw} - t_{ccw} = \frac{R}{\sqrt{2c + \omega R}} - \frac{R}{\sqrt{2c - \omega R}}. \quad (5.41)$$

If the Sagnac interferometer rotates such that its linear velocity is small compared to c , or

$$\omega R \ll c, \quad (5.42)$$

the time shift may be approximated as

$$\Delta t \approx \frac{8R^2\omega}{c^2}. \quad (5.43)$$

Since the period of an optical wave T is related to velocity and wavelength by

$$T = \frac{\lambda}{c}, \quad (5.44)$$

we may define a normalized phase shift as

$$\frac{\Delta t}{T} = \frac{8R^2\omega}{\lambda c}. \quad (5.45)$$

This last expression, a dimensionless quantity, gives the number of fringes which result when the Sagnac interferometer rotates at angular velocity ω , with respect to the loop at rest. This arrangement is thus used as a sensitive interferometric means to measure rotation.

Sagnac interferometers are widely used for high-resolution rotation measurement, and have replaced macroscopic mechanical gyroscopes for the most demanding applications. Typically using fiber rather than a free-space arrangement, fiber gyroscopes

(typically referred to as fiber-optic gyros, or FOGs) or ring lasers are the most common configurations (Auch, 1994; Barbour and Schmidt, 2001; Rasch et al., 2004). The former generally uses a fiber loop with a diameter on the order of $R \sim 10$ cm and a total fiber length of several kilometers; since optical fiber (the subject of Chapter 9) has a small diameter, such systems are compact and benefit from a long optical path length and thus enhanced fringe shift for a given rotation. State-of-the-art FOG designs employ integrated optical chips, chiefly using LiNbO_3 , for wave splitting, phase modulation and polarization control.

Optical fiber gyroscopes based on the Sagnac interferometer are generally employed in applications with significant demands on performance and stability for rotational measurements. The accuracy ranges typically required are on the order of $100^\circ/\text{hr}$ for automotive applications, $10^\circ/\text{hr}$ for ships or robotics, $0.1^\circ/\text{hr}$ for aircraft and $10^{-3}^\circ/\text{hr}$ for use in space; demands on reliability are very high for the latter applications and performance is typically limited by temperature drift and laser noise. Optical gyroscopes have been used in automotive systems by Japanese manufacturers since 1992, and are now becoming particularly widespread in aerospace applications; the navigation system of the Airbus A380 relies on them. For low-resolution rotational sensing, MEMS-based (non-optical) accelerometers are becoming increasingly popular, but micro-optical technology is allowing novel multi-dimensional rotational sensor configurations (Cole et al., 2000) as well as high-performance diffractive beam splitters for advanced Sagnac interferometers (Fahr et al., 2007).

5.2.5 Fabry-Perot interferometer

It is likely that the interferometer manufactured by far in the largest quantities is the Fabry-Perot⁹ interferometer. Structurally it is perhaps the simplest of the interferometer types we consider here, although it is probably simpler to fabricate using the technologies of microfabrication rather than by classical optical assembly techniques.

The Fabry-Perot interferometer, also frequently referred to as a Fabry-Perot cavity, resonator or etalon, was originally developed for spectroscopy (Fabry and Pérot, 1901; Mulligan, 1998) and consists of two parallel reflecting surfaces, as shown in Figure 5.15. The mirrors are often only partially reflecting, with reflectivity r_L and r_R , and transmissivity t_L and t_R , for the left and right surfaces, respectively. For a light field E_0 coupled into the cavity, wavelength-dependent phase matching of the field upon multiple reflections allows establishment of resonance, and therefore transmission through the cavity, whereas wavelengths which are far from resonance decay due to destructive interference.

⁹ Charles Fabry (1867–1945) and Alfred Pérot (1863–1925), both at the University of Marseille, France, conceived and demonstrated their interferometer together in 1897. Fabry was intensely interested in spectroscopy, particularly for astronomy, and discovered, with H. Buisson, the ozone layer in 1913. Pérot was exceedingly talented at mechanical construction, and it is likely that their first interferometer only worked because of Pérot's skill in equipment design.

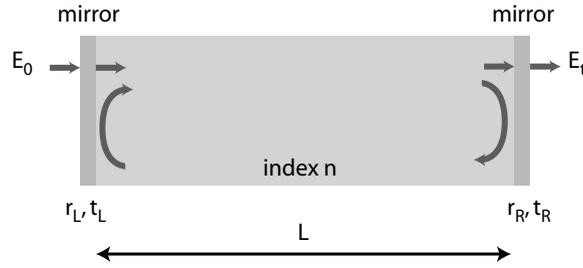


Figure 5.15 The Fabry-Perot interferometer consists of two parallel reflecting surfaces, of which at least one typically has $R < 1$. Light coupled in at wavelengths which lead to resonance, for which the cavity length is an integer number of wavelengths, is transmitted, whereas other wavelengths decay due to destructive interference.

Cavity resonance

We can analyze the phase matching which leads to resonance by considering the propagation of the field E through the cavity when subject to multiple reflections (Lipson et al., 1995) (Iizuka, 2002a, Chapter 3). For cavity length L and wave vector (or propagation constant) k , the phase shift ϕ the wave experiences after one round trip through the cavity is given by

$$\phi = 2kL = \frac{4\pi nL}{\lambda_0} \quad (5.46)$$

for a cavity filled with material of refractive index n . Resonance occurs if the phase shift is an integer multiple of 2π , or $\phi = 2m\pi$ for integer m .

The nature of the transmission through the cavity can be determined by calculating what happens when a field E_0 with wavelength λ_0 is incident from the left and is subject to multiple reflections inside the cavity, each round trip increasing the phase shift by ϕ . The transmitted optical field E_t may be seen to be

$$E_t = E_0 t_L t_R + E_0 (t_L r_R r_L t_R) e^{j\phi} + E_0 t_L (r_R r_L)^2 t_R e^{j2\phi} + \dots \quad (5.47)$$

which, by setting $r_L = r_R = r$ and $t_L = t_R = t$ for the special but realistic case of a symmetric cavity, and remembering that $r^2 = R$ and $t^2 = T$, can be written as a function of reflectivity and transmittivity as

$$E_t = E_0 T (1 + R e^{j\phi} + R^2 e^{j2\phi} + \dots) \quad (5.48)$$

By summing the series, we can then determine the ratio of the transmitted to incident fields to be

$$\frac{E_t}{E_0} = \frac{T}{1 - R e^{j\phi}} \quad (5.49)$$

We see thus that the transmission through the Fabry-Perot resonator is a critical function of the phase, which is in turn, by Equation 5.46, directly related to the ratio of wavelength to optical cavity length. For a fixed nL , wavelengths for which the optical cavity length is an integer multiple of λ , resonance occurs and transmission is maximum.

The effect of resonance is best seen by considering the transmission of optical intensity (Zappe, 2004, Chapter 2.5), which is proportional to the (complex) square of the fields,

$$\frac{I_T}{I_0} = \frac{E_T E_T^*}{E_0 E_0^*} = \frac{T^2}{(1 - Re^{j\phi})(1 - Re^{-j\phi})}. \quad (5.50)$$

This expression may be expanded using the Euler¹⁰ relation to yield

$$\frac{I_T}{I_0} = \frac{T^2}{(1 - R \cos \phi - jR \sin \phi)(1 - R \cos \phi + jR \sin \phi)} \quad (5.51)$$

$$= \frac{T^2}{1 + R^2 - 2R \cos \phi}. \quad (5.52)$$

which may be written, remembering that $T = 1 - R$ and by using a bit of trigonometry ($1 - \cos \theta = 2 \sin^2(\theta/2)$), as

$$\frac{I_T}{I_0} = \frac{(1 - R)^2}{(1 - R)^2 + 4R \sin^2\left(\frac{\phi}{2}\right)}. \quad (5.53)$$

We thus have an expression for the transmission through the Fabry-Perot resonator as a function of mirror reflectance and the phase, which yields a characteristic as shown in Figure 5.16. The resonances (maxima) in transmission are clearly seen for $\phi = m2\pi$ for integer m . From this consideration, and Equation 5.46, we see that resonance occurs for a cavity length (for a fixed wavelength)

$$L = \frac{m\lambda_0}{2n} \quad (5.54)$$

or a wavelength (for a fixed cavity length)

$$\lambda_0 = \frac{2nL}{m}. \quad (5.55)$$

Recalling that $\nu = c/\lambda$, the corresponding frequencies at which resonance occurs for fixed cavity length are given by

$$\nu_0 = \frac{mc}{2nL}. \quad (5.56)$$

Finesse

It is useful to define the finesse F [], a dimensionless parameter which is only a function of mirror reflectance,

$$F = \frac{\pi\sqrt{R}}{1 - R} \quad (5.57)$$

¹⁰ Leonhard Euler (1707-1783), an exceedingly prolific Swiss mathematician, was too young (16) upon graduation from the University of Basel to accede to a professorship there and subsequently spent most of his adult life in Russia. Said to have been blinded in one eye by un-protected solar observations (students working with lasers take note), his work formed the basis for much of modern calculus, trigonometry and analysis, which he applied to a wide spectrum of real-world problems.

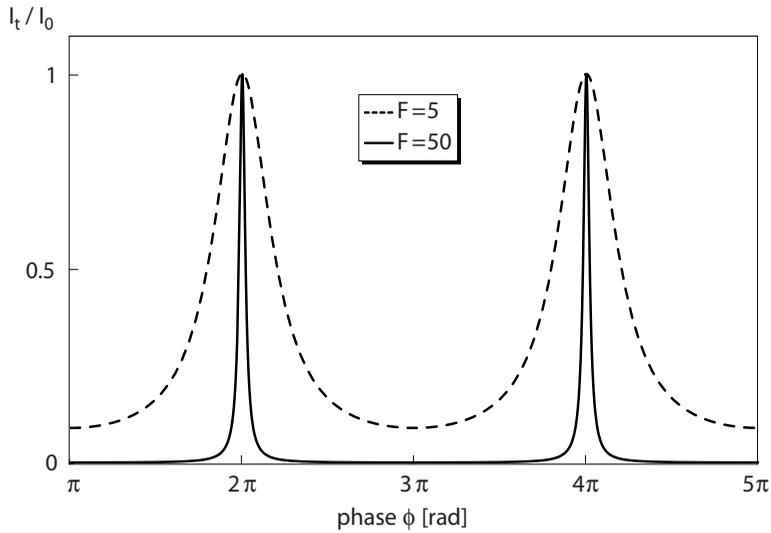


Figure 5.16 Transmission through a Fabry-Perot interferometer as a function of the phase ϕ for two different values of finesse, $F = 5$ dashed and $F = 50$ solid.

and defines the sharpness of the resonance. Using this definition, we can write the transmission characteristic of the Fabry-Perot resonator as

$$\frac{I_T}{I_0} = \frac{1}{1 + \left(\frac{2F}{\pi}\right)^2 \sin^2\left(\frac{\phi}{2}\right)}. \quad (5.58)$$

For the transmission characteristics of Figure 5.16, the resonance for two values of finesse are shown: $F = 5$, obtained for $R = 0.54$ and $F = 50$, for $R = 0.94$. We see that higher values of F lead to sharper resonance characteristics and narrower transmission maxima. Fabry-Perot cavities made using mirrors of higher reflectance thus have more sharply defined maxima. Finesse values of several hundred are routinely available and cavities with F in excess of 10 000 are used for specialized spectroscopic applications.

Free spectral range

We see from the Fabry-Perot transmission characteristic of Figure 5.16 that a transmission maximum occurs periodically, for every wavelength/optical cavity length combination that leads to a total phase shift of an integer multiple of 2π . From Equation 5.56, the spacing between resonance frequencies is given by

$$\Delta\nu_{FP} = \frac{c}{2nL} \quad (5.59)$$

where $\Delta\nu_{FP}$ [Hz] is known as the free spectral range (FSR), the symbol for which should not be confused with that for the linewidth, $\Delta\nu$. We may also express FSR in terms of

wavelength, $\Delta\lambda_{FP}$ [m], namely

$$\Delta\lambda_{FP} = \left(\frac{\lambda^2}{c}\right) \Delta\nu_{FP}. \quad (5.60)$$

The free spectral range thus provides a measure for the spacing between the resonance peaks. If the Fabry-Perot resonator is to be used to either to distinguish between two closely spaced wavelengths or for the establishment of resonance at a single wavelength, the FSR defines the wavelength range for which the resonance is uniquely defined. For wavelength shifts or differences larger than the FSR, the resonance is no longer single-valued: we cannot distinguish between the desired resonance and the adjacent one shifted by 2π from it.

Spectral width

A final useful parameter for characterizing the resonance of a Fabry-Perot resonator is that of the width of the resonance. Typically, the full-width half-maximum (FWHM) of the transmission characteristic is used, such that the resonance width is taken for $I_T/I_0 = 1/2$. Since, at resonance, the sine term is close to zero, so that $\sin \phi \approx \phi$, we have, at half-maximum,

$$\frac{F\phi}{\pi} = 1 \quad (5.61)$$

so that the FWHM in frequency is

$$FWHM_\nu = 2\phi = \frac{2\pi}{F} = \frac{\Delta\nu_{FP}}{F} \quad (5.62)$$

where the last equivalence derives from the fact that a phase shift of 2π corresponds to the FSR. We may also express spectral width in wavelength, $FWHM_\lambda$, related to $FWHM_\nu$ as

$$FWHM_\lambda = \left(\frac{\lambda^2}{c}\right) FWHM_\nu. \quad (5.63)$$

Implementations

The classically-assembled Fabry-Perot interferometer places extreme demands on optomechanical assembly. Alignment of the mirrors is critical as these must be positioned absolutely parallel to each other. Systems such as this are commercially available, primarily for use in spectroscopy, but are expensive and often bulky. The manufacturing techniques of microsystems engineering, particularly the micromachining of semiconductor substrates, has however allowed the relatively easy fabrication of micro-Fabry-Perot cavities, essentially “for free.” When a III-V semiconductor is cleaved, it breaks along a crystalline plane, with virtually atomic flatness; two such cleaved planes made in the same crystal are highly parallel, due to the quality of the crystal structure. Since such planes act as partially reflecting mirrors ($R \approx 0.3$ for GaAs to air), two parallel cleaved planes form an excellent Fabry-Perot cavity and indeed, this cavity forms the basis for the resonator of semiconductor laser diodes, as we will see in greater detail in Chapter 10. Since Fabry-Perot laser diodes, the simplest laser diode structure, are

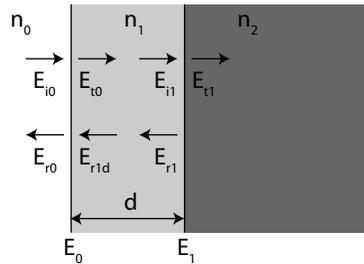


Figure 5.17 An optical multilayer stack of three materials with refractive indices n_0 , n_1 , and n_2 . An electromagnetic wave with electric field magnitude E_{i0} is incident from the left; a portion E_{t0} is transmitted and E_{r0} reflected. The former results in a field E_{i1} incident onto the second boundary, of which a portion E_{t1} is transmitted and E_{r1} reflected. The reflected part returns to the first interface and leads to a field E_{r1d} . The arrows show the propagation directions. The fields E_0 and E_1 are the total fields at the n_0/n_1 and n_1/n_2 boundaries, respectively.

manufactured in volumes of tens of millions, this interferometer is indeed the most widespread and widely applied of the configurations we have discussed, even if most users are completely unaware of it.

5.3 Optical multilayers

Interference phenomena are not only used in interferometers, they also form the basis for an important family of optical coatings, including anti-reflection and high-reflection multilayers and the broad yet important field of interference filters. Most macroscopic optical elements are coated with precisely-defined thicknesses of materials with precisely-defined refractive indices, and the optical performance of these layers is frequently a measure for the quality of the component. The use of optical coatings for micro-optical components is still in its infancy, but as the performance of microlenses improves, the required optical coating technology will become essential.

Optical coatings, thin-film filters and interference filters are a large topic, with a considerable literature; see for example (Macleod, 2001). Many processes are industrially standard, since the macroscopic optics field relies heavily on them. We will examine the basics of the topic here, showing how the optical effects of multilayer structures may be calculated.

5.3.1 Single thin film

We begin the analysis by examining the effect of a single thin transparent layer on an optical substrate, as shown schematically in Figure 5.17. We let an electromagnetic wave be incident on the system from the left and examine what happens to the fields by considering the boundary conditions at each interface. To simplify our lives, we will assume normal incidence although the general case for oblique incidence is not significantly more complicated (Hecht, 2002, Chapter 9.7).

Boundary conditions for fields

We recall from Equation 2.56 that the Poynting vector, which is parallel to the wave vector \mathbf{k} and thus defines the direction of energy flow, is related to the magnitudes and directions of the electric and magnetic fields, \mathbf{E} and \mathbf{H} ; the vector relationships, derived from the wave equation, take the form

$$\mathbf{k} \times \mathbf{E} = k_0 \mu \mathbf{H} \quad (5.64)$$

$$\mathbf{k} \times \mathbf{H} = -k_0 \epsilon \mathbf{E} \quad (5.65)$$

where k_0 is the scalar quantity $2\pi/\lambda_0$. Considering the propagation directions indicated in Figure 5.17, we can write, for the scalar electric fields E_0 and E_1 ,

$$E_0 = E_{i0} + E_{r0} = E_{t0} + E_{r1d} \quad (5.66)$$

$$E_1 = E_{i1} + E_{r1} = E_{t1} \quad (5.67)$$

for the n_0/n_1 and n_1/n_2 boundaries, respectively. Recalling the definition of impedance Z (Equation 2.57), for which

$$Z = \sqrt{\frac{\mu_m}{\epsilon_m}} = \frac{1}{n} Z_0 \quad (5.68)$$

and thus

$$H = \frac{1}{Z} E \quad (5.69)$$

for refractive index n and impedance of free space Z_0 , we can similarly write the sums of the magnetic fields as

$$H_0 = \frac{1}{Z_0} (E_{i0} - E_{r0}) = \frac{1}{Z_1} (E_{t0} - E_{r1d}) \quad (5.70)$$

$$H_1 = \frac{1}{Z_1} (E_{i1} - E_{r1}) = \frac{1}{Z_2} E_{t2}. \quad (5.71)$$

The minus signs result from the asymmetry in the vector relationships of Equations 5.64 and 5.65; apply the right-hand rule to see why.

The relationship between the transmitted or reflected fields and their value on the other side of the film, with thickness d and refractive index n_1 , is given by the simple propagation formula, namely

$$E_{i1} = E_{t0} e^{-jk_1 d} \quad (5.72)$$

and

$$E_{r1d} = E_{r1} e^{-jk_1 d}, \quad (5.73)$$

the latter implying

$$E_{r1} = E_{r1d} e^{+jk_1 d} \quad (5.74)$$

where $k_1 = (2n_1\pi) / \lambda_0$. Hence, at the n_1/n_2 boundary, we have

$$E_1 = E_{t0} e^{-jk_1d} + E_{r1d} e^{jk_1d} \quad (5.75)$$

$$H_1 = \frac{1}{Z_1} (E_{t0} e^{-jk_1d} - E_{r1d} e^{jk_1d}). \quad (5.76)$$

Solving these last two equations for E_{t0} and E_{r1d} and plugging the results into Equations 5.66 and 5.70, we obtain

$$E_0 = E_1 \cos k_1d + jH_1 Z_1 \sin k_1d \quad (5.77)$$

$$H_0 = j \frac{E_1}{Z_1} \sin k_1d + H_1 \cos k_1d \quad (5.78)$$

which can most practically be written in matrix form as

$$\begin{bmatrix} E_0 \\ H_0 \end{bmatrix} = \begin{bmatrix} \cos k_1d & jZ_1 \sin k_1d \\ j\frac{1}{Z_1} \sin k_1d & \cos k_1d \end{bmatrix} \begin{bmatrix} E_1 \\ H_1 \end{bmatrix} \quad (5.79)$$

or

$$\begin{bmatrix} E_0 \\ H_0 \end{bmatrix} = F_i \begin{bmatrix} E_1 \\ H_1 \end{bmatrix}. \quad (5.80)$$

The matrix F_i contains all the relevant information about the film i : its thickness, refractive index and impedance. If there were second film to the right of the E_1 , with fields E_2 and H_2 at the surface of the substrate, we could then simply multiply a second matrix to obtain the fields, namely

$$\begin{bmatrix} E_0 \\ H_0 \end{bmatrix} = F_1 F_2 \begin{bmatrix} E_2 \\ H_2 \end{bmatrix} \quad (5.81)$$

such that for p films, we have the form

$$\begin{bmatrix} E_0 \\ H_0 \end{bmatrix} = F_1 F_2 F_3 \dots \begin{bmatrix} E_p \\ H_p \end{bmatrix} = F_T \begin{bmatrix} E_p \\ H_p \end{bmatrix} \quad (5.82)$$

where

$$F_T = \prod_{i=0}^p F_p. \quad (5.83)$$

The matrix form thus provides a very simple means to express the effects of optical multilayers: we can multiply the characteristic matrices of each layer to determine the effect of the layer stack as a whole.

Reflectivity

More useful than an evaluation of the fields at the boundaries of the optical layers is the determination of the reflectance which results. If we replace the field quantities in Equation 5.80 with their equivalents, we obtain

$$\begin{bmatrix} E_{i0} + E_{r0} \\ \frac{1}{Z_0} (E_{i0} - E_{r0}) \end{bmatrix} = F \begin{bmatrix} E_{t1} \\ \frac{1}{Z_2} E_{t1} \end{bmatrix}. \quad (5.84)$$

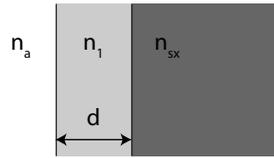


Figure 5.18 A single optical layer with thickness d and refractive index n_1 , with light incident from air (n_a) and transmitted into a substrate with n_{sx} .

Recalling our definitions for reflectivity r and transmittivity t ,

$$r = \frac{E_{r0}}{E_{i0}} \quad (5.85)$$

$$t = \frac{E_{t1}}{E_{i0}}, \quad (5.86)$$

we can then re-write our matrix expression as

$$\begin{bmatrix} 1 + r \\ \frac{1}{Z_0}(1 - r) \end{bmatrix} = F \begin{bmatrix} t \\ \frac{1}{Z_2}t \end{bmatrix}. \quad (5.87)$$

Simplifying the layer matrix F from Equation 5.79 by adopting the following notation,

$$F = \begin{bmatrix} \cos k_1 d & jZ_1 \sin k_1 d \\ j\frac{1}{Z_1} \sin k_1 d & \cos k_1 d \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}, \quad (5.88)$$

we can solve Equation 5.87 for the reflectivity r to yield

$$r = \frac{\frac{1}{Z_0} \left(F_{11} + \frac{1}{Z_2} F_{12} \right) - \left(F_{21} + \frac{1}{Z_2} F_{22} \right)}{\frac{1}{Z_0} \left(F_{11} + \frac{1}{Z_2} F_{12} \right) + \left(F_{21} + \frac{1}{Z_2} F_{22} \right)}. \quad (5.89)$$

This last expression allows us to calculate the reflectivity (or reflectance $R = rr^*$) for an arbitrary thin film of thickness d and with refractive index n_1 . The refractive indices of the incident medium and the substrate are present in the impedances Z_0 and Z_2 , respectively. We will now examine a few special cases and see how all of this can be exceedingly useful.

5.3.2 Quarter- and half-wave films

Consider the single-film structure of Figure 5.18. For a thin film with refractive index n_1 , and thus $k_1 = (2n_1\pi)/\lambda_0$ for operating wavelength λ_0 , and thickness d , examination of the transmission matrix, Equation 5.88, reveals that we expect interesting things to happen for $k_1 d = \pi/2$ and $k_1 d = \pi$.

Quarter wave film

Let us consider the case for $k_1 d = \pi/2$, which, since a phase shift of $\pi/2$ implies a quarter of a wavelength, is known as a quarter-wave film. The transmission matrix (Equa-

tion 5.88) simplifies to

$$F_1 = \begin{bmatrix} 0 & jZ_1 \\ j\frac{1}{Z_1} & 0 \end{bmatrix} \quad (5.90)$$

so that the reflectivity, Equation 5.89, becomes

$$r_{QW} = \frac{Z_1^2 - Z_a Z_{sx}}{Z_1^2 + Z_a Z_{sx}} \quad (5.91)$$

for impedances Z_a and Z_{sx} of the air and the substrate, respectively. Recalling from Equation 5.68 that $Z_i \propto (1/n_i)Z_0$, this last expression can be re-written in the convenient form

$$r_{QW} = \frac{n_a n_{sx} - n_1^2}{n_a n_{sx} + n_1^2} \quad (5.92)$$

implying a reflectance, corresponding to what would be measured,

$$R_{QW} = \left(\frac{n_a n_{sx} - n_1^2}{n_a n_{sx} + n_1^2} \right)^2. \quad (5.93)$$

This last relation gives the reflectance of a single quarter-wave film as a function of the refractive indices of the materials, and we see that $R_{QW} = 0$ for

$$n_1 = \sqrt{n_a n_{sx}}. \quad (5.94)$$

Thus if a material with a refractive index corresponding to the geometric mean of the indices of the incident medium (air in this case, but n_a may be arbitrary) and the substrate is deposited on the substrate with a quarter-wave thickness, the film serves as a perfect anti-reflection (AR) coating for the design wavelength λ_0 . This effect is a very important one and implies that simple coatings may be employed to reduce unwanted reflections from optical surfaces.

Example 5.4: Let us examine the transmission characteristic of a single thin film of the type shown in Figure 5.18 as a function of the phase $\phi = k_1 d$. We consider a BK7 substrate with $n_{sx} = 1.5168$ in air and use a HeNe laser, for which $\lambda_0 = 633$ nm, for illumination.

Using Equation 5.94, we find that an optimal index for a single-layer AR coating would be $n_1 = 1.232$. A quarter wave film would then result for $d = \pi/(2k_1) = \lambda_0/4n_1 = 128$ nm. In Figure 5.19, we plot the variation of R for such a film as a function of phase by multiplying Equation 5.89 by its complex conjugate, yielding $rr^* = R$, and we see that indeed $R \rightarrow 0$ for $\phi = \pi/2$, the quarter wave condition. The thin solid line at $R \approx 0.04$ is the reflectance which would result from a bare BK7 substrate with no coating.

As we will presently see in the next section, no optical material with a refractive index of 1.232 is known; the closest we can get is MgF_2 , which gives $n_1 = 1.38$. Figure 5.19 also plots R for this film material, and we see that there is a reduction in reflectance for $\phi = \pi/2$, but the value does not reach zero, since the indices are not optimally matched.

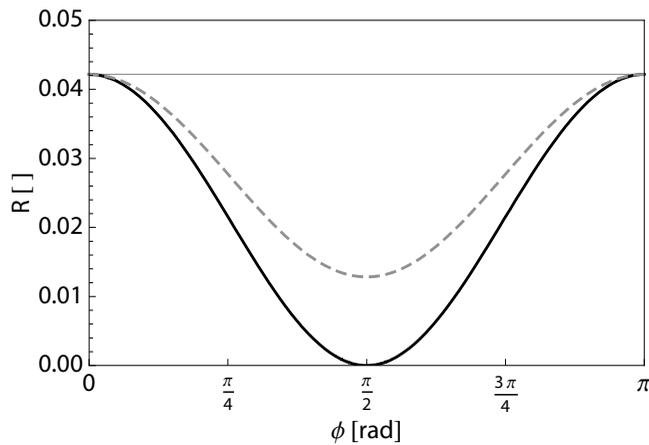


Figure 5.19 Reflectance at normal incidence of a single optical film in air on a BK7 substrate ($n_{sx} = 1.5168$) as a function of the phase $\phi = k_1d$. The thin solid line at $R \approx 0.04$ is the reflectance of the glass with no optical coating; shown is the reflectance for a hypothetical film with $n_1 = 1.232$ solid line and for MgF_2 with $n_1 = 1.38$ dotted line.

Half-wave film

From the characteristic of Figure 5.19, we see that for $\phi = 0$ or $\phi = \pi$, the value for reflectance is the same as that of the uncoated substrate. The thickness for which $\phi = \pi$ is termed a half-wave layer, and the result is the native reflectance of the substrate, as if the film were not there.

Materials

Key to the use of a single layer as an AR coating is the availability of a material with the correct refractive index. We saw in the example above, however, that the optimal refractive index for single-layer coatings to be used in air with typical glasses (where $n \approx 1.5$) is on the order of 1.2 and the material which comes closest to this value is MgF_2 which has $n = 1.38$. Also sometimes used is cryolite, Na_3AlF_6 with $n = 1.35$, but this material is soft and hygroscopic, limiting its utility as a durable optical coating. Since SiO_2 ($n = 1.46$) is widely used in microsystems engineering and micro-optics, it is also often used as an optical thin film, its relatively high index compensated for by its generally uncomplicated deposition.

As we will see in Section 5.3.4 below, low-index optical films are often paired with high-index films in multilayer stacks; a maximum index difference between the two is typically desirable. For high-index films, TiO_2 ($n_o = 2.621$, $n_e = 2.919$) or Ta_2O_5 ($n = 2.2$) are often used; the former is birefringent as a crystal, which is not a problem when deposited as an amorphous thin film, for which the index value is typically quoted as $n = 2.34$. In addition, ZnS ($n = 2.35$) is sometime used as a high-index film, but also suffers from poor hardness and susceptibility to water absorption.

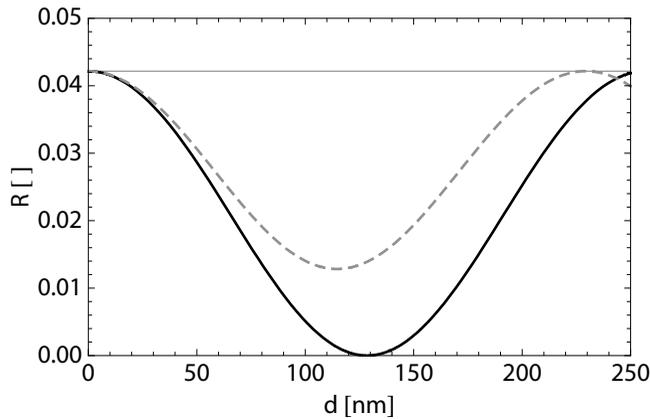


Figure 5.20 Reflectance at normal incidence of a single optical film in air on a BK7 substrate ($n_{sx} = 1.5168$) optimized for $\lambda_0 = 633$ nm as a function of the film thickness d . The reflectance for a hypothetical film with $n_1 = 1.232$ *solid line* and for MgF_2 with $n_1 = 1.38$ *dotted line* are shown. The thin solid line at $R \approx 0.04$ is the reflectance of the glass with no optical coating.

Example 5.5: Film thickness is a critical parameter in determining the anti-reflection performance of an AR coating. Let us see how the reflectance varies as the film thickness varies from its optimal design value. We again take a single film structure on a BK7 substrate with $n_{sx} = 1.5168$ in air and use a HeNe laser ($\lambda_0 = 633$ nm) for illumination.

As we found in Example 5.4, the optimal index for a single-layer AR coating would be $n_1 = 1.232$ and a quarter wave film results for $d = 128$ nm. Using Equation 5.89 as before, Figure 5.20 shows the variation of R with d , again for a hypothetical film with $n_1 = 1.232$ and for MgF_2 ($n_1 = 1.38$).

We notice that the two curves are displaced from each other, unlike the case for Example 5.4; the reason is that the two curves represent different refractive indices and thus different *optical* thicknesses. As a result, the optimum thickness for the film with $n_1 = 1.232$ is the design value $d = 128$ nm, but the quarter wave thickness value for MgF_2 is $d = 115$ nm.

We see from this characteristic that film thickness is indeed a critical design and fabrication parameter. Variations of less than 50 nm can result in reflectance changes on the order of a factor of two. In addition, the importance of the absolute refractive index value should be clear: changing n changes the optical path length, so that the conditions for optimal anti-reflection (the quarter-wave optical thickness) also change with refractive index.

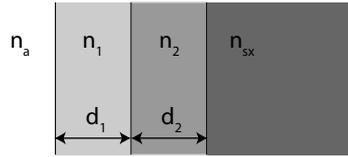


Figure 5.21 A dual optical layer structure with thicknesses d_1 and d_2 corresponding to refractive indices n_1 and n_2 , respectively. Light is incident from air (n_a) and transmitted into a substrate with n_{sx} .

5.3.3 Dual-layer coatings

The tools we developed above may easily allow us to consider optical coatings consisting of more than one layer. Multilayer coatings, as we will see presently, provide even greater flexibility in the design of anti-reflection or high-reflection coatings, as well as a broad spectrum of interference filters with custom-designed wavelength response.

Consider the dual-layer structure shown in Figure 5.21, two films with thicknesses d_1 and d_2 and refractive indices n_1 and n_2 , respectively. We can now calculate its reflection properties by using a product of transmission matrices as we saw in Equation 5.83. If we again consider the special case of quarter-wave films, then the transmission matrix of Equation 5.88 simplifies to the form of Equation 5.90 and our two layer structure is characterized by

$$F = F_1 F_2 = \begin{bmatrix} 0 & j\frac{1}{n_1} \\ jn_1 & 0 \end{bmatrix} \begin{bmatrix} 0 & j\frac{1}{n_2} \\ jn_2 & 0 \end{bmatrix} \quad (5.95)$$

$$= \begin{bmatrix} -\frac{n_2}{n_1} & 0 \\ 0 & -\frac{n_1}{n_2} \end{bmatrix} \quad (5.96)$$

from which we can determine the reflectivity, using the same procedure as in the last section, to be

$$R_{2QW} = \left(\frac{n_2^2 n_a - n_{sx} n_1^2}{n_2^2 n_a + n_{sx} n_1^2} \right)^2. \quad (5.97)$$

We see from the previous expression that reflectance can again be identically zero, for

$$\frac{n_2}{n_1} = \sqrt{\frac{n_{sx}}{n_a}} \quad (5.98)$$

where typically $n_2 > n_1$ for most coating sequences. We thus have a condition for the relationship between the film indices which will allow optimized anti-reflection, for a particular wavelength.

Example 5.6: Using the thin-film materials available, optical AR coatings are usually optimized for operation at green to yellow wavelengths, since these correspond roughly to the center of the visible wavelength range. How do the anti-reflection properties of

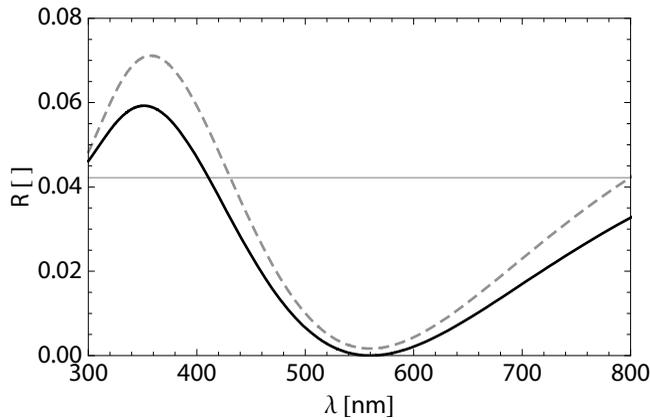


Figure 5.22 Reflectance at normal incidence of a dual optical film system in air on a BK7 substrate ($n_{sx} = 1.5168$) optimized for $\lambda_0 = 560$ nm as a function of wavelength. The reflectance is calculated for $n_1 = 1.38$ (MgF_2) and a hypothetical optimal film with $n_2 = 1.70$ solid line and for Al_2O_3 with $n_2 = 1.77$ dotted line, where all films are a quarter-wavelength thick for the design wavelength. The thin solid line at $R \approx 0.04$ is the reflectance of the glass with no optical coating.

a dual-layer coating vary with wavelength around the design value? Let us take a BK7 glass substrate and design a two-layer sequence optimized for $\lambda_0 = 560$ nm.

As with the single-layer case, the optimal refractive indices are one thing; the available materials are another. Remembering that $n_{sx} = 1.5168$ for BK7, and that $n_a = 1$, we can employ MgF_2 for layer 1 (so that $n_1 = 1.38$), which would optimally require $n_2 = n_1 \sqrt{n_{sx}/n_a} \approx 1.70$. Looking into our catalogue of materials, we find that aluminum oxide, Al_2O_3 , would provide $n_2 = 1.77$, which is reasonably close. We can do the calculation for both 1.70 and 1.77.

We employ Equation 5.88 to define transmission matrices F_1 and F_2 for the two films, using the correct refractive indices, and multiply these to obtain the transmission matrix for the dual-layer structure,

$$F_{dual} = F_1 F_2.$$

The matrix elements of F_{dual} are then fed into Equation 5.89, where we recall that Z_0 corresponds to the incident medium (air) and Z_2 the substrate (BK7 glass); we further recall that $Z_n = (1/n)Z_0$, so that we can replace $1/Z_0 \rightarrow n_a$ and $1/Z_2 \rightarrow n_{sx}$ in the expression for r which results. Finally, we determine the reflectance from $R = rr^*$.

The result is plotted in Figure 5.22, for the optimal hypothetical $n_2 = 1.70$ and the experimentally realizable $n_2 = 1.77$ using Al_2O_3 . We see qualitatively different behavior than that which was seen for the film thickness variation for the single-layer structure of Figure 5.20. We see, as expected, $R = 0$ at the design wavelength, $\lambda_0 = 560$ nm for the optimal film ($n = 1.70$), and a value $R > 0$ for Al_2O_3 ($n = 1.77$). For $\Delta\lambda \approx \pm 150$ nm, there is a reduction in reflectance as compared to the bare glass surface, but we see that for certain wavelengths (in the blue to violet part of the spectrum in this example), reflectance is actually higher than an uncoated glass substrate.

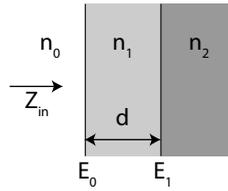


Figure 5.23 Two optical layers, of which the first has thickness d , with refractive indices n_1 and n_2 . The impedance seen from from the n_0 region to the left, looking to the right, is given by Z_{in} .

This type of spectral behavior is the reason that AR-coated photographic lenses or eyeglasses, for example, often have a purple cast to them; those shorter wavelengths are reflected more efficiently by the coatings which are optimized, as here, for the center of the visible spectrum.

5.3.4 Impedance concepts

We saw in the previous section, and the last example, that combining two or more films requires that we multiply the transmission matrices to allow calculation of the reflectivity or transmittivity. This mathematical process quickly becomes tedious if we combine a large number of films, which we will do in Section 5.3.5. Luckily, we can take an alternative approach employing the impedances of the films, with which complex layer stacks can be analyzed quickly (Brooker, 2003, Chapter 6).

Consider two optical films, as shown in Figure 5.23. At the n_0/n_1 boundary, we can write the impedance as seen looking to the right, using the field definitions of Equations 5.77 and 5.78, as

$$Z_{in} = \frac{E_0}{H_0} = \frac{E_1 \cos k_1 d + jH_1 Z_1 \sin k_1 d}{j \frac{E_1}{Z_1} \sin k_1 d + H_1 \cos k_1 d} \quad (5.99)$$

where Z_1 is the impedance (Equation 5.68) of film n_1 and $k_1 = (2\pi n_1/\lambda_0)$.

Quarter-wave films

For the case of quarter-wave film thicknesses, $k_1 d = \pi/2$ and the previous expression simplifies to

$$Z_{in} = \frac{H_1 Z_1}{\frac{E_1}{Z_1}} \quad (5.100)$$

and since E_1 and H_1 define the impedance looking to the right at the next boundary, the n_1/n_2 interface, so that

$$\frac{E_1}{H_1} = Z_2 \quad (5.101)$$

for Z_2 the impedance of film 2, such that we finally have

$$Z_{in} = \frac{Z_1^2}{Z_2}. \quad (5.102)$$

This last relationship thus provides a simple means to express the film impedance seen from the surface as a function of the layers deposited on a substrate.

If we use this approach for the classic single-film structure of Figure 5.18, we would find the simple input impedance

$$Z_{in} = \frac{Z_1^2}{Z_{sx}}. \quad (5.103)$$

Recalling the reflectivity for this structure as given in Equation 5.92, we can re-write this in terms of impedance, remembering that $Z \propto 1/n$, as

$$r = \frac{n_a n_{sx} - n_1^2}{n_a n_{sx} + n_1^2} = \frac{Z_1^2 - Z_a Z_{sx}}{Z_1^2 + Z_a Z_{sx}}. \quad (5.104)$$

By incorporating the definition of Z_{in} into this last expression, we obtain a very simple form for reflectivity,

$$r = \frac{Z_{in} - Z_a}{Z_{in} + Z_a}. \quad (5.105)$$

We see thus that the reflectivity is directly related to the difference between input impedance and the impedance of the incident medium, air in this case. The larger the difference between the two, the larger r becomes, an effect which we will use to advantage when we consider multilayer interference filters below.

Half-wave films

In a half-wave film, $k_1 d = \pi$ and the expression for input impedance becomes simply

$$Z_{in} = \frac{E_1}{H_1} = Z_2. \quad (5.106)$$

Thus a layer of half-wave thickness, as we saw previously in Section 5.3.2, has no optical effect and acts optically as if it were not there.

5.3.5 Multilayer interference filters

We have seen that adding more layers to an optical surface qualitatively and quantitatively changes the reflection characteristic; this effect was clearly seen in Figure 5.22. To see what happens as we add more optical layers to a surface, we can use either of the tools we developed above: multiplication of the transmission matrices or impedances. Whereas matrix multiplication is the more general approach, particularly as film thicknesses vary from their quarter-wave values or non-normal incidence is considered, we will use impedances to show most simply the effect of optical multilayers.

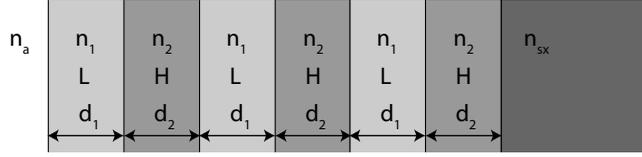


Figure 5.24 A sequence of multiple optical layers, alternating between low index (L), with n_1 and thickness d_1 , and high index (H), with n_2 and thickness d_2 , on a substrate with n_{sx} . Three LH pairs are shown; incidence is from air, n_a .

Bragg mirror reflectance

Consider the multilayer structure shown in Figure 5.24; a sequence of three high (H) and low (L) index pairs (with indices and thicknesses n_1, d_1 and n_2, d_2 , respectively) is shown. If we start from the right (the substrate n_{sx}) and calculate the impedance based on the considerations leading to Equation 5.102, we have, for the first (H) layer

$$Z_{in1} = \frac{Z_H^2}{Z_{sx}}. \quad (5.107)$$

This impedance acts as the load for the second (L) layer,

$$Z_{in2} = \frac{Z_L^2}{Z_{in1}} = \frac{Z_L^2}{Z_H^2} Z_{sx} \quad (5.108)$$

which is in turn the load for the third (H) layer

$$Z_{in3} = \frac{Z_H^2}{Z_{in2}} = \frac{Z_H^4}{Z_L^2 Z_{sx}} \quad (5.109)$$

such that for the fourth (L) layer, we have

$$Z_{in4} = \frac{Z_L^2}{Z_{in3}} = \frac{Z_L^4}{Z_H^4} Z_{sx}. \quad (5.110)$$

We see the pattern here: for p HL layer pairs, the impedance seen from the surface is thus given by

$$Z_{in} = \left(\frac{Z_L}{Z_H} \right)^{2p} Z_{sx}. \quad (5.111)$$

Recalling that $Z \propto 1/n$, we can then write the reflectivity from Equation 5.105 as

$$r = \frac{\left(\frac{Z_L}{Z_H} \right)^{2p} - Z_a}{\left(\frac{Z_L}{Z_H} \right)^{2p} + Z_a} \quad (5.112)$$

$$= \frac{n_a \left(\frac{n_H}{n_L} \right)^{2p} - 1}{n_a \left(\frac{n_H}{n_L} \right)^{2p} + 1} \quad (5.113)$$

and we recall that reflectance $R = r^2$. We thus see that increasing the number of mirror pairs p then results in a higher value for reflectance.

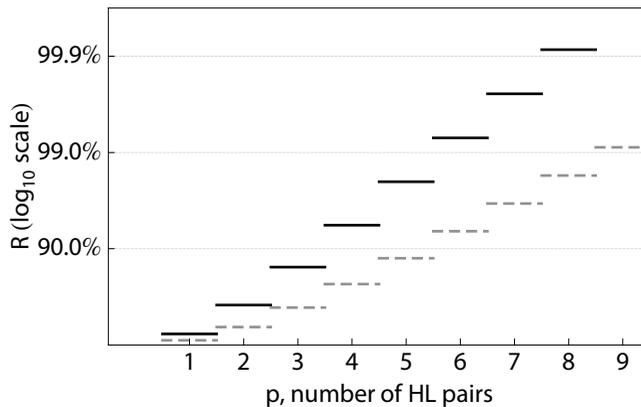


Figure 5.25 Reflectance of Bragg mirrors using p quarter-wave mirror pairs for two material systems: MgF_2 ($n = 1.38$) and TiO_2 ($n = 2.34$) *black solid lines*; and SiO_2 ($n = 1.46$) and Si_3N_4 ($n = 2.05$) *gray dashed lines*. Note that the R axis is a log scale and p takes on only integer values.

Mirrors such as this, made from alternating quarter-wave layers of material, are often called Bragg¹¹ mirrors. We see from inspection of the last expression above that the highest reflectance values result from material combinations with the highest possible difference in refractive indices between the H and L layers; n_H/n_L is thus large, and when multiplied by itself $2p$ times, even larger. This effect is especially useful in the fabrication of very high reflectance (HR) Bragg mirrors. Using a sequence of QW layers with large index difference, Bragg mirrors with $R > 0.999$ may be fabricated; compare this value with that for, say, aluminum ($R \approx 0.91$), which is the reflecting surface for many standard technical mirrors, such that Bragg mirrors are often the only means to achieve very high reflectance values.

Example 5.7: We can examine the performance of multi-layer Bragg mirrors using two possible material systems. Given, on the one hand, MgF_2 with $n = 1.38$ and TiO_2 with $n = 2.34$, and on the other hand, SiO_2 with $n = 1.46$ and Si_3N_4 with $n = 2.05$, how many mirror pairs p do we need using these two systems to reach $R = 99\%$?

We can apply Equation 5.113, take $R = r^2$, and plot R as a function of p (recalling that p is an integer); the increase in reflectance with increasing mirror pair number is shown in Figure 5.25. On this log plot, we see that, using the $\text{MgF}_2 / \text{TiO}_2$ system, $p = 4$ is sufficient to generate a mirror with $> 90\%$ reflectance. With $p = 6$, we have $R > 99\%$, and $p = 8$ results in $R \geq 99.9\%$.

¹¹ William Henry Bragg (1862–1942) and his son William Lawrence Bragg (1890–1971) formed a remarkable English father/son team and was the only parent/child pair to be awarded a joint Nobel Prize, in 1915, when William Lawrence was only 25. They founded the field of X-ray crystallography and demonstrated that X-rays were indeed TEM waves. Both were gifted teachers and did much to promote science for the general public.

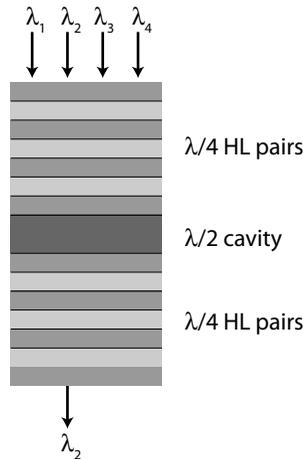


Figure 5.26 A resonant cavity filter comprising a $\lambda/2$ central cavity surrounded on either side by Bragg mirrors consisting of $\lambda/4$ -thick HL pairs. For broad-band input, all wavelengths except a single narrow-band transmission wavelength are reflected, λ_2 in this sketch.

For the $\text{SiO}_2 / \text{Si}_3\text{N}_4$ system, due to the lower refractive index contrast, we require more mirror pairs to reach a certain level of reflectance. We see from the plot that $p = 9$ is required for $R \geq 99\%$. Due to the higher refractive index contrast, semiconductor materials are thus usually more practical than, say, polymers, for fabricating Bragg mirrors.

Bragg multilayer HR mirrors do have two limitations, however: the reflectance is wavelength dependent, as we saw for a dual-layer film system in Figure 5.22, and also incidence-angle dependent. Both of these dependences may be quantified using a transmission matrix calculation of reflectance, but it can be noted that most multilayer mirrors work well even at incident angles up to 30° from the normal. The best Bragg mirror performance is achieved quarter wave stacks which begin and end with H material, and thus have the sequence [air/HLHL. . . HLHLH/substrate].

Resonant Bragg filters

Combining two Bragg mirrors with a half-wavelength thick central cavity, thus a sequence of the form [air/HLHL. . . HLLH. . . LHLH/air] forms a resonant transmission cavity, shown schematically in Figure 5.26. The, usually broad, reflection spectrum of the Bragg mirrors defines the so-called *stop-band*. Near the center of the stop-band is a sharp transmission maximum, at the wavelength defined by the half-wave cavity between the two Bragg mirrors; Figure 10.18 shows a typical reflectance characteristic. This structure may be used as a sharp bandpass filter (Strassner et al., 2000) and forms the basis for the cavity structure of the VCSEL, the vertical cavity surface emitting laser, which we will study in greater depth in Chapters 8 and 10.

Through a controlled change in material or structural properties, the center wavelength of the stop-band, the wavelength of maximal reflection, of a Bragg mirror or the position of the transmission peak in a Bragg resonant filter, can be tuned. Thermo-optic tuning is particularly attractive when using silicon-based materials (Domash et al., 2004) and MEMS fabrication technology can be used to manufacture free-standing Bragg filter membranes which can be thermally tuned in wavelength with millisecond time constants. In addition, a physical change in Bragg layer thickness itself can also be used as a tuning mechanism; polymers may be induced to swell, for example in the presence of solvents in the ambient atmosphere (Mönch et al., 2006a,b), and this swelling may result in thickness changes on the order of 100%. We will discuss these tunable optical filters in greater detail in Chapter 12.

Problems

1. Plot the interference characteristic generated by two waves of identical intensity in air with wavelength $\lambda = 780 \text{ nm}$ as a function of path-length difference, for the range $0 \leq \Delta L \leq 8 \text{ }\mu\text{m}$.
2. Given a laser beam with $\lambda_0 = 1.55 \text{ }\mu\text{m}$, which is split into two paths of identical length. In one path we place a $15 \text{ }\mu\text{m}$ thick piece of an exotic transparent material whose refractive index may be changed from 1.60 to 3.20. Ignoring any possible intensity difference between the beams, plot the interference characteristic as a function of refractive index.
3. The professor insists that you re-consider the material in the previous problem more carefully. You consult a data sheet and are shocked to find that the material has an absorption coefficient of $4 \times 10^3 \text{ dB/cm}$ at your measurement wavelength. Not forgetting Fresnel reflection, plot the interference characteristic as a function of refractive index again.
4. Show that the sum of the fields, E_T , in Example 5.2 is as given.
5. Show that the sum of the fields, E_T , in Example 5.3 is as given.
6. We examine the interference pattern generated from two beams, one of which is kept at a constant power of 1 mW. Calculate the visibility for an interfering beams with powers 1100 μW , 950 μW , 500 μW , 90 μW and 850 nW.
7. Plot the interference characteristic of a Young interferometer with slits $3 \text{ }\mu\text{m}$ apart illuminated by a HeNe laser in air as measured on a screen 10 cm from the illuminated slits; plot for a range $\pm 200 \text{ mm}$ from the pattern center.
8. For the Young interferometer in the previous example, a 5 mm thick transparent crystal with $n = 2.8$ is placed before one slit. Re-plot and compare the interference patterns.
9. Assuming that the electronics with which you evaluate the interference intensity of your Michelson interferometer (illuminated with a diode laser with $\lambda_0 = 850 \text{ nm}$) allows you to distinguish a phase shift of $2\pi/50$ in air, what movement resolution of the measurement mirror does your system provide?
10. A perfectly balanced Michelson interferometer has an output at the maximum of the interference characteristic. The mirror begins to move and the output signal

decreases. Is the measurement mirror moving closer to or farther away from the interferometer?

11. Using an integrated optical Michelson interferometer with a laser source whose linewidth is measured at the operating wavelength of $\lambda_0 = 780 \text{ nm}$ to be 2 pm and which has a fixed reference path length of $100 \text{ }\mu\text{m}$, you measure the movement of a test mirror 5 cm from the interferometer which is displaced by $20 \text{ }\mu\text{m}$. How many fringes do you measure? You move the mirror to the other end of the 1 m long optical table, and measure again. How many fringes do you measure now?
12. A classical, free-space Mach-Zehnder interferometer is set up in a temperature-controlled environment. The beam of one arm of the interferometer is directed through a temperature-variable air path 10 cm long. You know that the refractive index of dry air at standard atmospheric pressure (100 kPa) varies approximately as

$$n_{air}(T) = 1 + \frac{7.86 \times 10^{-2}}{T + 273}$$

for $\lambda_0 = 633 \text{ nm}$. Around room temperature, what is the phase shift in degrees per degree celsius of temperature change which you can expect?

13. The measurement beam of a free-space Mach-Zehnder interferometer is directed through a glass ampule which has a 3 mm space filled with deionized water. You know that the refractive index of water varies with temperature roughly as (Abbate et al., 1978)

$$n_{water}(T) = 1.34498 + B \left(T - T_k e^{\frac{T-T_0}{T_k}} \right)$$

for $B = 26.2 \times 10^{-5} \text{ K}^{-1}$, $T_k = 48.5 \text{ K}$ and $T_0 = 2 \text{ K}$ and a wavelength of 633 nm . Sketch the variation of n_{water} with temperature in the range $10 \leq T \leq 100 \text{ }^\circ\text{C}$. The water temperature now changes from 20° C to 35° C . How many fringe shifts result from this temperature change? Plot the fringe pattern you observe in the Mach-Zehnder as the temperature rises.

14. You are given an integrated optical Mach-Zehnder interferometer operating at $\lambda_0 = 650 \text{ nm}$ with a 3 mm long exposed measurement arm for which you know that the change in phase with respect to change in external refractive index per millimeter measurement arm length is $\frac{\Delta\phi}{\Delta n} = 8 \times 10^5 \text{ rad}$. The sensor is in a flow cell filled with water ($n = 1.3318$) and slowly the flow changes to 13% ethanol ($n = 1.3386$) in water (about the strength of wine). Plot the resulting interference characteristic.
15. The north/south railway connection through the Gotthard in the Swiss Alps is by means of an amazing series of tunnels, dug in the 19th century. Some of these make a full 360° rotation while underground, such that one can emerge from the mountain and see the tunnel entrance a few hundred meters directly below oneself at the same point in the valley, without having felt oneself going around a curve. Assuming the train moves at 80 km/h through such a helical tunnel with a length of 1476 m (the Pfaffensprung Kehrtunnel), how many fiber windings would a 10 cm diameter Sagnac interferometer operating with a 1550 nm laser diode require to indicate one fringe shift while in the tunnel and what is the total required fiber length?

16. Plot the Fabry-Perot interference as a function of phase generated by slab of BK7 glass. What is the visibility of the resulting intensity pattern?
17. What mirror reflectance is required to generate a Fabry-Perot cavity with finesse of 10 000?
18. Calculate F_T for a dual-layer system, with the first film of index n_1 and thickness d_1 and the second with n_2 and d_2 ; write as a function of k_1 and k_2 as well as Z_1 and Z_2 . Express in matrix form. Then calculate the matrix values numerically for layer 1 SiO_2 ($n_1 = 1.45$ with $d_1 = 134$ nm) and layer 2 Si_3N_4 ($n_2 = 2.05$ with $d_2 = 95$ nm), when used for a 780 nm laser diode.
19. For the layer structure of the previous example, calculate a numerical value for reflectance for the layers deposited on BK7 ($n = 1.5168$) in an air ambient. Verify that the expected value results for the case where there are no films on the substrate (i.e., native reflectance of the glass substrate).
20. Plot the variation of reflectance of a single quarter-wave film deposited on BK7 glass in air as a function of film refractive index, in the range $1 \leq n_1 \leq 1.4$; verify that the minimum is where we expect it to be.
21. Plot the variation of reflectance of a dual quarter-wave film system deposited on BK7 glass in air as a function of the refractive index of n_2 , given a top film of MgF_2 with $n = 1.38$. Sketch for the range $1.5 \leq n_2 \leq 1.9$ and again verify that the minimum is where we expect it to be.
22. Determine the variation of reflectance for a Bragg mirror composed of p HL pairs as a function of the index contrast of the pairs, n_H/n_L ; calculate and plot for $p = 3, 5, 8$. Indicate which contrast values are achievable with the $\text{SiO}_2/\text{Si}_3\text{N}_4$ and $\text{MgF}_2/\text{TiO}_2$ material systems.

Part II

Micro-optics

6 Reflective micro-optics

*Les miroirs feraient bien de réfléchir un peu plus avant de renvoyer les images.*¹

Jean Cocteau

Of all sub-disciplines of micro-optics, reflective micro-optics is perhaps the simplest from the optical point of view, yet the most successful commercially. Micromirrors are manufactured by the billions and their simple optical function, namely to reflect an incident light beam, is the basis for a significant sector in the home entertainment industry on the one hand, and forms the backbone for the fiber networks on which the planet relies on the other.

We will see how micromirrors have revolutionized video beamers and have become the mainstay for optical telecommunications in the sections below. Beginning with a basic analysis of reflection for planar, parabolic and spherical mirrors, we will subsequently look at the basic technology of micromirrors, their structure, optical behavior and means for actuation. A special case for reflective micro-optics is that of adaptive optics, which relies on mirror arrays or deformable single mirrors to correct a distorted wavefront, and has become an essential component in medical and astronomic imaging systems. Finally, as micromirrors represent advanced technological products, we will examine a few case studies, showing where you are likely to have already encountered one of these components.

6.1 Reflection

Reflective micro-optics relies on reflection of an optical field from a mirror surface, which is, at least in its basics, relatively easy to understand. In striving for high reflectivity for ever larger wavelength ranges, engineers continue to study reflection both theoretically and experimentally, and this seemingly well-developed topic is still the subject of a fair amount of research activity.

¹ “Mirrors would do well to reflect (think) a bit more before sending back images.” Jean Cocteau (1889-1963), French filmmaker. After seeing his 1950 film *Orphée*, mirrors never seem quite the same again.

6.1.1 Mirror surfaces

The function of a mirror is to reflect the maximum possible amount of incident electromagnetic radiation. Whereas our daily experience in front of the bathroom mirror takes place at visible wavelengths (at least for most of us), mirrors are vital optical components for electromagnetic waves ranging from radio to x-ray wavelengths. The technologies for different frequency bands vary considerably; x-ray mirrors, for example, are usually structured so that they only work for grazing incidence of the electromagnetic fields. Since our interest is in micro-optics, we will focus here on mirrors suitable for the visible and near infrared wavelengths, for which most micro-optical components are intended.

Most planar mirrors consist of a glass blank with a metallic coating; we will discuss suitable materials in the next section. Curved mirror surfaces, such as the parabolic or spherical structures we will see below, also have their shape defined by a rigid substrate, which may be glass or a ceramic but also a bulk metal, and are likewise coated with a highly-reflecting metal layer. Micro-optical mirrors, manufactured using the technologies of microsystems engineering, may consist of bulk metal films (thick enough to be mechanically stable) or use silicon and other semiconductors as the physical substrate.

Macroscopic consumer mirrors (e.g., the bathroom mirror) usually have the reflecting layer behind the glass, to protect the thin and fragile metal reflector, such that wiping off water condensation by rubbing with a towel is not a problem. Technical mirrors used in the laboratory, on the other hand, have the reflecting metal layer exposed on the surface, such that the point of reflection is well defined and transmission through the glass plate does not have to be considered. For that reason, technical mirrors are more delicate and are generally cleaned by a gentle nitrogen stream or, if necessary, using deionized water and a soft cloth suitable for cleaning optical surfaces. Cleaning by indiscriminate rubbing or using a paper towel results at best in a scratched surface, at worst, in an annoyed PhD advisor. Micromirrors, if cleanable at all, need to be pampered in the cleanroom. That cleaning is not always necessary can be seen in Figure 6.1

6.1.2 Metals for reflection

Key to the function of a mirror is its reflecting layer which, for most consumer, technical and micromirrors, is typically a metal film. For advanced high-reflectivity structures, optical multilayers, which we saw in Section 5.3, are frequently employed; as we learned there, these thin film structures are typically complex and have a strong wavelength and incidence-angle dependence. Reflecting metal films, in contrast, usually have a broad usable wavelength range and have no discernible variation in reflectance with incidence angle. In addition, metal films can be deposited using a variety of established evaporation technologies, and have a high tolerance for thickness variations.

Aluminum

The most popular metal for technical (and consumer) mirrors is aluminum, due to its adequately high reflectance, ease of deposition and relative immunity from oxidation.



Figure 6.1 The author peering up into one of the primary mirrors of a solar telescope buried in a mine shaft at the top of the Schauinsland, a mountain adjacent to Freiburg, Germany. Note that the mirror surface, especially near the top, looks like someone cleaned it with a damp dish towel, but for the optical system in question, this does not matter. However, there is something wrong with this picture; see Problem 1.

Usually deposited by evaporation on a mirror pre-form, Al layers have a reflectance in excess of 0.90 for most of the near-UV, visible and NIR wavelength ranges, with a small dip around 800 nm. Aluminum does oxidize, but slowly and only for a few nanometers from the surface; the oxide layer affects reflection in the ultraviolet and may slightly increase surface scattering at all wavelengths.

High-quality Al mirrors may be coated with surface layers to prevent oxidation and to improve abrasion resistance. Manufacturers use dielectrics such as disilicon trioxide (Si_2O_3) or, particularly for UV applications, MgF_2 ; these are typically deposited as $\lambda/2$ layers centered around 500 nm. Surface coatings result in more durable technical mirrors, but may also affect the reflection spectrum.

Silver

The highest reflectance values over a broad wavelength range for metallic reflectors are reached using silver; values of $R \geq 0.98$ are typical for visible and NIR wavelengths. We recall from Section 3.1.2 that the plasma resonance in a metal defines its transmission and reflection properties: metals are highly reflecting for wavelengths longer than (frequencies lower than) the plasma frequency and absorb radiation near the plasma wavelength. The plasma frequency of Ag is in the near-UV (310 nm), such that this material is less suitable than Al for short wavelengths. The resulting slight excess absorption at blue wavelengths is what gives Ag its slightly “warmer” color when compared to, for example, aluminum.

The primary problem when using Ag as a technical reflector is that it oxidizes (tarnishes) rapidly, reducing reflectance and strongly increasing scattering. For this reason,

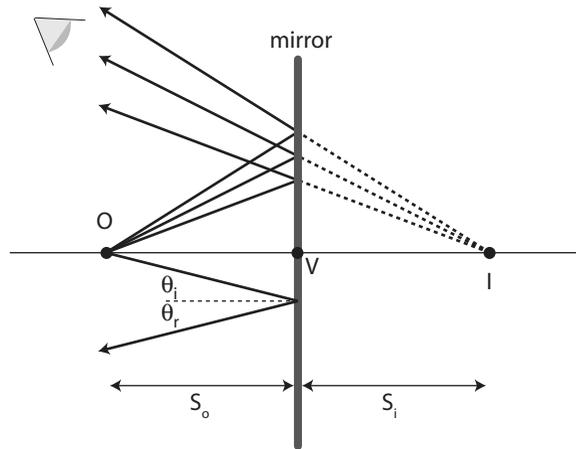


Figure 6.2 Reflection of rays emerging from an object at point O from a planar mirror. All rays reflect at an angle θ_r equal to the incident angle θ_i . The image of the point O , placed at a distance S_o in front of the mirror, is at the point I , found at a distance S_i behind it.

Ag-based mirrors must either be surface-coated with a protective film or, more popularly, used on an internal surface. For the latter case, the back surface of a glass block or prism is coated with Ag, and this layer covered with a protective passivation, such that the silver is completely isolated from the external atmosphere; the mirror surface is then on the backside of the block or prism and transmission is through the glass.

Gold

Gold is unparalleled for reflection at long wavelengths, far into the mid-IR. Although the reddish cast of gold, due to its plasma frequency in the violet (496 nm), and its relatively low reflectance ($R < 0.5$ around 500 nm) make it of limited utility for visible wavelengths, the reflectance of Au averages $R \geq 0.99$ for $0.7 \leq \lambda \leq 20 \mu\text{m}$. For this reason, Au reflecting layers are essential for applications in the IR and is used for many types of thermal reflectors.

Even more so than the previous materials, gold is soft and easily damaged by scratching. Protection layers are thus essential if a robust Au mirror is to be deployed. On the other hand, Au is very inert and does not oxidize, so that environmental exposure to air or humidity is typically not a problem.

6.2 Planar mirrors

Planar mirrors are the simplest reflectors to analyze and the easiest to understand. As summarized in Figure 6.2, the function of the planar mirror (as all mirrors) can be fully described by realizing that all incident rays reflect at an angle θ_r equal to the incident angle θ_i , a process known as specular reflection.

6.2.1 Planar mirror imaging

Let us consider the rays emitted from an object point O located a distance S_o from the mirror position V (for vertex; the reason will become clear when we consider spherical and parabolic mirrors). By tracing a set of three rays upward from O , we see that these diverge when reflected from the mirror, and may be traced back (dotted lines) to a virtual image point I on the other side of the mirror. The position of the image point is at S_i , such that $|S_o| = |S_i|$. The sign convention is such that $S > 0$ if the position is to the *left* of V , so that in our sketch $S_i < 0$.

Since the reflected rays diverge, imaging the point I requires a lens, such as the lens of the eye or the objective in front of a CCD (charge coupled detector) imager. Such an image is termed *virtual*, since it can only be reconstructed using a lens or imaging optical system; in contrast, a *real* image is one which can be projected onto and seen on a screen. In our daily experience, a virtual image is one we see when looking into a mirror, into a microscope or through the viewfinder of a camera; a real image is that which we see on a movie screen or that projected by a beamer onto the wall.

These considerations reflect our daily experience in front of the bathroom mirror. The image we see appears to be placed behind the mirror surface at the same distance from the mirror as we are in front of it; obliquely incident rays are obliquely reflected; and the image can only be seen if you employ (for example) the lens of your eye (or, as in Figure 6.1, a camera objective) to see it. There is no image formed on a piece of paper which we might hold in front of a mirror, as there would be if we held it in front of a projector.

6.2.2 Rotating planar mirrors

The vast majority of micromirror applications require that these rotate, usually to scan an optical beam; we will see examples in Section 6.4. Simple geometry allows us to determine the deflection of the beam when the mirror is rotated about an axis, as shown in Figure 6.3.

We saw above that for a non-tilted mirror ($\phi = 0$), a beam incident at angle θ undergoes a total angle change of 2θ . For mirror tilted at an angle ϕ , we can then easily see that a beam incident at angle θ will be transformed by $2\theta + 2\phi$. This simple relationship will become important in analyzing the beam propagation from scanning micromirror systems.

Example 6.1: Let us analyze a Fabry-Perot interferometer, of the form considered in Section 5.2.5. As shown in Figure 5.15, the interferometer cavity consists of two parallel mirrors and the optical field oscillates in resonance between them. How close to parallel do the mirrors need to be? What deviation from the normal to the mirrors is allowable for the optical beam?

Let us assume that the mirrors are circular and have a radius of $r = 10$ cm; we couple light into the cavity through a tiny hole in the center of one of the mirrors, whose optical

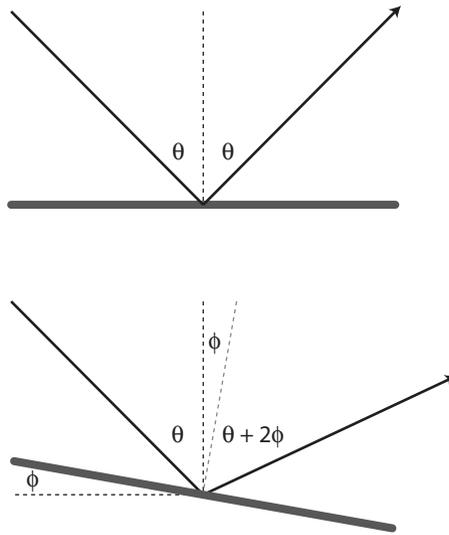


Figure 6.3 A mirror subject to rotation by an angle ϕ ; a beam incident at angle θ is then reflected at $\theta + 2\phi$.

effect we then ignore; and we assume the mirror reflectivity $R = 1$. The cavity length (spacing between the mirrors) is L and we would like at least N_R reflections before the beam wanders out of the cavity.

These are both exercises in geometry, whereby the second question is easier to answer than the first; the student is encouraged to make a sketch to help visualize the ray paths between the mirrors. Considering the second question first, take two perfectly parallel vertical mirrors and propagate a beam at a deviation angle θ_e from the horizontal between them. After propagating across the cavity of length L , the beam has moved upwards (or downwards) by an amount δ given by

$$\delta = L \tan \theta_e.$$

For each reflection, the beam wanders another δ from the center, so that the number of possible reflections N_R for a mirror with radius r is given by

$$N_R = \frac{r}{\delta} = \frac{r}{L \tan \theta_e}$$

or, alternatively, the allowed angular deviation of the beam from the normal for a given minimum number of reflections N_R may be found from

$$\theta_e = \arctan \frac{r}{LN_R}.$$

For the 10 cm mirrors, if we require $N_R \geq 100$, we see that a 20 cm long cavity thus requires a beam aligned to better than 0.3° . For a 1 m cavity, across an optical table, the beam must deviate no more than 0.06° from the normal. It is thus clear that, the longer the cavity, the more demanding the alignment, whereby these results assume perfectly parallel mirrors.

How close to parallel? Let us assume the beam emerges into the cavity with $\theta_e = 0$ (i.e., perfectly normal) but that one of the mirrors is tilted by θ_{tilt} . In this case the geometry is a touch more complicated (the student is *really* urged to go get paper, pencil and a ruler) since the angle of incidence increases with each reflection on the tilted mirror. For N_{RP} reflection pairs (i.e., once right, once left), the total movement of the beam is given by

$$\delta_{tot} \approx \sum_{n=1}^{N_{RP}} 2L \tan [(2n)\theta_{tilt}]$$

where we assume that θ_{tilt} is sufficiently small so that the cavity length is approximately the same for each reflection.

Taking a 20 cm long cavity and requiring at least 100 reflections ($N_{RP} = 50$), we find from this expression that a tilt angle of less than $\theta_{tilt} \leq 0.0056^\circ$ is required to keep the beam inside the 10 cm radius of the mirror ($\delta_{tot} \leq 10$ cm).

These calculations show that the alignment of Fabry-Perot cavities is anything but trivial, and the demands on stability and alignment increase as the cavity becomes larger. It is for this reason, also, that real Fabry-Perot cavities use parabolic, not planar, mirrors.

6.3 Non-planar mirrors

Whereas planar mirrors serve a wide range of useful functions in technical (macroscopic) optics as well as in micromirror structures, mirrors with non-planar surfaces also play important roles in many optical systems. Primarily used for focussing and imaging applications, spherical and paraboloidal mirrors are found in advanced macroscopic optical systems, particularly in telescopes or for use at wavelengths where refractive optics is limited due to the lack of appropriate transparent materials, but also used in such mundane applications as automobile headlights and LED packages.

6.3.1 Parabolic mirrors

The ideal geometric shapes for mirrors used in focussing optics are paraboloidal, ellipsoidal or hyperboloidal; the reader will recall that the parabola, the ellipse and the hyperbola are all conic sections which differ in their eccentricity. Depending on the application, either concave (meaning the mirror surface is curved toward the beam, as in the example of Figure 6.4) or convex (meaning the mirror surface is curved away from the beam) surfaces are employed.

The focussing function of a concave paraboloidal mirror is shown schematically in Figure 6.4. A collimated light beam impinges on the mirror surface parallel to the optical axis, which is the line extending normally from the vertex V of the parabola and intersecting the focal point F . It can be shown, using our simple relationship $\theta_r = \theta_i$,

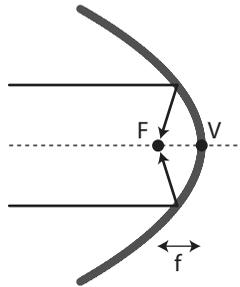


Figure 6.4 Rays impinging on the focus of a concave parabolic mirror. The rays from a collimated beam parallel to the optical axis are focussed at the point F , which is found a distance f , the focal length, from the mirror surface.

that all rays incident on the mirror parallel to the optical axis will meet at the focal point F or, conversely, that light emitted from a point source located at F will emerge from the mirror as a collimated beam, parallel to the optical axis.

Ellipsoidal and hyperboloidal mirrors function similarly; their mathematical form yields two foci and as mirrors they direct the optical rays from one focus to the other. These surfaces, along with paraboloids, are in general very difficult to fabricate, both macroscopically and in micro-optical implementations, and for that reason, spherical mirrors, as discussed in the next section, are often used. For classical optical systems, ellipsoidal and hyperboloidal mirrors are most often found in telescopes and are manufactured using advanced machining technology.

Micro-optically, parabolic micromirrors have been made using microlenses as a mold, yielding 240 μm diameter paraboloid mirrors fabricated in UV-curable photoresist and coated with Au for high reflectance (Merenda et al., 2007). Two-dimensional parabolic micromirrors (a parabola, not a paraboloid of revolution) can relatively easily be defined using high resolution photolithography (Zhang et al., 2007). In addition, microsystems techniques may be employed for generating variable paraboloidal surfaces, using electrostatic deflection of suspended membranes to achieve paraboloids with a deviation of less than $\lambda/10$ from their ideal shape at visible wavelengths, for mirror diameters in the millimeter range (Hokari and Hane, 2008; Mescheder et al., 2005). These micromirrors have the advantage that they are tunable, allowing their use in dynamic focussing applications; we will see related structures when we discuss adaptive optics in Section 6.5 below.

6.3.2 Spherical mirrors

As mentioned in the previous section, non-spherical shapes are in general difficult and expensive to fabricate, for the mirrors we consider here and also for the lenses of the next chapter. Luckily, a parabola, for small deviations from its vertex, can be well-approximated by a circle (Hecht, 2002, Chapter 5.4.3), so that a paraboloidal mirror is well approximated by a spherical mirror for rays close to the optical axis. As shown in

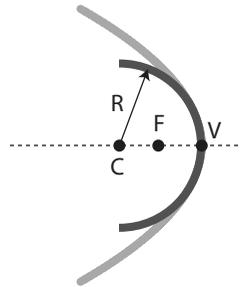


Figure 6.5 A paraboloidal mirror is well approximated by a spherical mirror for rays close to the optical axis. The spherical mirror has a radius R centered at C ; the focal point is F .

Figure 6.5, in which we see the cross section of paraboloidal and spherical mirrors, the parabola is closely matched by a circle of radius R centered at C .

The mirror equation

To allow a simple analysis of imaging using spherical mirrors, it is useful to define several sign conventions; similar rules will apply to the refractive optics we will consider in Chapter 7. In particular, we stipulate:

- Light always comes from the left;
- For concave mirrors, C is to the *left* of V and $R < 0$;
- For convex mirrors, C is to the *right* of V and $R > 0$.

Using these conventions, we can then determine the relationships between objects (or light sources) and their images, using the configuration and notation defined in Figure 6.6. For an object at position O , located a distance S_o from the vertex V of the mirror, its image can be found by ray tracing (no more than $\theta_r = \theta_i$ is required, albeit on a curved surface) to be at position I , a distance S_i from V . From geometry, we derive the mirror equation

$$\frac{1}{S_i} + \frac{1}{S_o} = -\frac{2}{R} \quad (6.1)$$

recalling the sign conventions defined above. This expression for the mirror equation is valid in the paraxial region, close to the optical axis.

If we let the object move to infinity (toward the left, so $S_o \rightarrow \infty$ and thus representing a collimated input beam), the mirror equation simplifies to

$$\frac{1}{S_i} = -\frac{2}{R}. \quad (6.2)$$

Since the rays of a collimated beam parallel to the optical axis impinging on a concave mirror are focussed at the point F , $S_i = f$ and thus

$$f = -\frac{R}{2}. \quad (6.3)$$

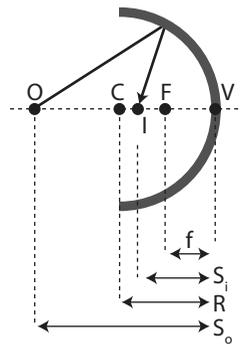


Figure 6.6 A spherical mirror is used to generate an image of the object point O at I ; the spacings between O and I and the vertex V are S_o and S_i , respectively.

Thus the focal point of a spherical lens is the point on the optical axis corresponding to a spacing of half the radius of curvature from the mirror vertex, and is shown by the point F in Figure 6.6. The mirror equation then takes on its most useful form,

$$\frac{1}{S_i} + \frac{1}{S_o} = \frac{1}{f}. \quad (6.4)$$

Focus and collimation

We examine this last expression for two useful special cases. First, if the object is placed at F , then $S_o = f$ and the mirror equation tells us that the image distance $S_i \rightarrow \infty$, implying that the beams are parallel to the optical axis and thus collimated. This situation arises, for example, when an LED (which may be considered to be a point light source if the rest of the optical system is large enough) is placed at the focus of a spherical reflector: the light emerges in a more-or-less collimated beam.

The second special case places the object at infinity to the left ($S_o \rightarrow \infty$), which results in the image at the focus; this case is identical to the previous, of course, only the light moves in the opposite direction. Thus the collimated input light is concentrated at the focus, the configuration used in everything from telescopes to solar concentrators.

Example 6.2: We would like to design micro-reflector to collimate the light from an LED. Let us assume we can etch a hemispherical concave profile into a 300 μm thick Si wafer, and can suspend the LED chip from an etched (or hybridly assembled) cantilever placed at the level of the surface over this reflector. What should the dimensions of the reflector be?

From Equation 6.4, we see that a collimated beam implies that $S_o = f$, such that, using Equation 6.3, we have

$$S_o = -\frac{R}{2},$$

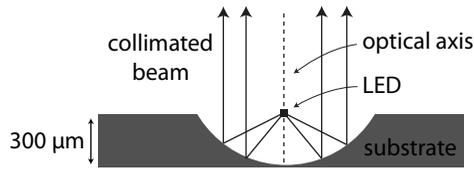


Figure 6.7 An LED positioned at the focus point $300\ \mu\text{m}$ above the bottom (vertex) of a concave mirror with $R = 600\ \mu\text{m}$ etched into a silicon substrate; the LED is held by an etched cantilever, not shown. Light emitted from the bottom of the LED is collimated by the mirror and emerges parallel to the optical axis.

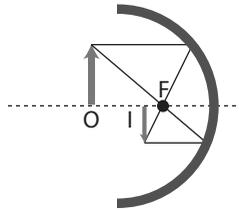


Figure 6.8 Use of a concave spherical mirror for generating an image; the object arrow at O is imaged (inverted and demagnified) at position I .

recalling that $R < 0$ for a concave mirror. Thus the LED must be placed at a position on the optical axis corresponding to half the radius of curvature of the reflector, as shown in Figure 6.7. Since we can etch maximally to a depth of $300\ \mu\text{m}$, and place the LED at the surface, we have $S_o = f = 300\ \mu\text{m}$ and thus $R = 600\ \mu\text{m}$. Therefore we etch a spherical surface with a radius of $600\ \mu\text{m}$ to a depth of $300\ \mu\text{m}$ into the substrate, and suspend the LED at the center, $300\ \mu\text{m}$ over the bottom of the etch profile. The mirror then collimates the light emitted from the LED downwards, toward the mirror, and the collimated beam is emitted upwards, as shown in the figure.

Imaging

If the position of the image of a single point reflected from a spherical mirror surface can be determined using ray tracing, the image of an extended object may likewise be easily assembled. We use the arrow (we opticians always image arrows) in Figure 6.8 as an example and apply two simple rules derived from the considerations of focus and collimation above. We see that: 1) the ray from the arrowhead which impinges on the mirror parallel to the optical axis is reflected through the focus F ; and 2) the ray from the arrowhead which traverses the focus and then strikes the mirror is reflected parallel to the optical axis. The position of the image of the arrowhead is the point where these two rays intersect, as shown in the figure.

The image arrowhead in Figure 6.8 is inverted and smaller when compared to the object arrow. Magnification M [] is thus negative and has an absolute value less than unity; erect images have $M > 0$ and magnified (i.e. larger) images have $|M| > 1$. In

general, the magnification of an image reflected from a concave mirror is given by

$$M = -\frac{S_i}{S_o}. \quad (6.5)$$

Table 6.1 summarizes the types, positions and characteristics of images which can be generated using spherical mirrors and variable placement of the object. We see that both real and virtual images may result when using concave mirrors, depending on the object position; convex mirrors always generate virtual images.

Example 6.3: Let us consider the LED from the previous example again. Say we make a mistake in our mounting process, and the LED source, rather than being at the level of the surface, 300 μm above the bottom of etched reflector, is 50 μm higher, so 350 μm above the reflector vertex. What is the nature of the image created?

Since the LED is no longer at the focal point, we have to apply Equation 6.4, which we can solve for the image position, recalling the relationship between focal length and radius of curvature, to yield

$$S_i = \left(\frac{2}{|R|} - \frac{1}{S_o} \right)^{-1}.$$

We see from Table 6.1 that the image of the LED will be at a finite position, inverted and magnified, where $M = -S_i/S_o$. Plugging in the numerical values for the dimensions and positions, we obtain

$$S_i = 2100 \mu\text{m} \quad \text{and} \quad M = -6.$$

Thus the beam is no longer collimated, but makes an image of the LED source about 2 mm above the mirror, inverted in orientation and magnified by a factor 6. Accurate positioning of sources which are focussed or collimated using spherical mirrors is therefore essential.

6.4 Micromirrors

The physics of reflection from mirrored surfaces does not change as the size of the mirror decreases, so that the considerations presented in the first part of this chapter fully apply to micromirrors. However, the structures, fabrication techniques and functions of micromirrors differ considerably from their macroscopic counterparts. In light of their relevance for a wide variety of applications, we will thus consider some of the aspects of micromirrors in more detail.

At what point does a mirror become a micromirror? As with all the other micro-optical components we consider in this book, the boundary between macro and micro using only size as a criterion is rather vague. An aspect better defined than size is fab-

Table 6.1. Imaging using spherical mirrors: image type, location, orientation and magnification for objects located at various positions with respect to the focus.

concave mirrors				
object location	image type	image location	orientation	magnification
$\infty > S_o > 2f$	real	$f < S_i < 2f$	inverted	$ M < 1$
$S_o = 2f$	real	$S_i = 2f$	inverted	$ M = 1$
$2f > S_o > f$	real	$2f < S_i < \infty$	inverted	$ M > 1$
$S_o = f$		$\pm\infty$		
$S_o < f$	virtual	$ S_i > S_o$	erect	$ M > 1$
convex mirrors				
anywhere	virtual		erect	$ M < 1$



Figure 6.9 Silicon-based two-dimensional scanning micromirrors on a 2-cent coin, with two differing mirror sizes. The chip sizes for both are 3×3 mm. Photo courtesy of Khaled Aljaseem.

rication technology: micromirrors, like most micro-optical components, are generally produced using microfabrication technologies based on semiconductor manufacturing, which we will consider in Chapter 11. Microfabrication, on which the microelectronics and microsystems fields rely, is quite distinct from the classical optical manufacturing techniques used for macroscopic optics, and is the only means by which complex opto-mechanical devices, such as the micromirrors shown in Figure 6.9, may be fabricated on size scales below 1 millimeter. Thus size is perhaps less relevant than fabrication approach, although manufacturing micro-optical components with dimensions larger than a few millimeters using microfabrication techniques is a bit like polishing the Golden Gate bridge with a toothbrush: conceivable but not entirely sensible.

6.4.1 Micromirror structure

Micromirrors are made using a variety of materials, whereby silicon is likely the most popular for non-commercial devices; silicon nitride has also been employed (Lutzenberger et al., 2003). Other semiconductors, metals, ceramics and polymers, as well as

hybrid combinations of these have been used and may have advantages (in cost, fabrication simplicity or compatibility with other processes), but the advanced state of development of traditional silicon micro-machining implies that this material is very well represented in micromirror development. Industrially, metals may be of greater relevance: the DMD micromirror used for displays, which we will discuss in Section 6.6 below, is based on aluminum.

Typical lateral micromirror dimensions lie in the range 10 μm to 1 000 μm , and thicknesses from 1 μm to several 10 μm . Whereas high performance reflection for fixed macroscopic mirrors is usually achieved using the thin-film multilayers discussed in Chapter 5.3, most micromirrors rely on the native reflection of the metal layers, either the bulk structure of the mirror itself or as a deposited thin-film on the mirror surface.

Micromirror movement

The great advantage of micromirrors over their macroscopic counterparts is that they are usually dynamic: they may be tilted, moved, or deformed, and these usually at high speed. Using MEMS-based cantilevers (Petersen, 1977) or tilting micromirrors (Petersen, 1980) for optical scanning was already proposed and demonstrated 30 years ago. Optical scanning or dynamic beam deflection is achieved using a tilting micromirror which is actuated to move through a certain angular range; we will look at the structures employed in more detail below.

Numerous other modes of micromirror movement are of interest as well. Purely vertical motion (piston movement, typically upwards from the substrate) is used to change optical path length; arrays of vertically moving micromirrors, or dynamically deformable reflecting continuous membranes, are used for wavefront correction, as we will see in more detail in Section 6.5. Lateral, in-plane motion of micromirrors is valuable as the basis for optical switches or variable optical attenuators, both important for optical communications applications. Finally, in-plane rotational motion of a mirror or a reflecting diffractive grating structure have been used for fabricating micro-external-cavity tunable lasers (Geerlings et al., 2006). Development of new and clever movable micromirror designs is still one of the most active areas for research at present (Hane and Sasaki, 2007).

6.4.2 Scanning micromirrors

Of all micromirror movement possibilities, the tilting of mirrors used for scanning or dynamically deflecting an optical beam is likely that which is the most developed and with the greatest number of applications (Solgaard, 2009, Chapter 7). The structure of a tilting micromirror is generally that of a thin plate, made from a stiff, micro-structurable material, and mounting mechanics which attach it to a substrate; both the plate and the mechanics are typically fabricated in a single step from the same material using wet or dry etch techniques. The mechanical linkage between mirror and substrate is often a spring; Figure 6.10 shows some simple examples. In Figure 6.10 (a) and (b), one-dimensional scanning mirrors are attached to the substrate using either a torsional or a spring mount; both of these twist or bend as the mirror moves in the y direction. In

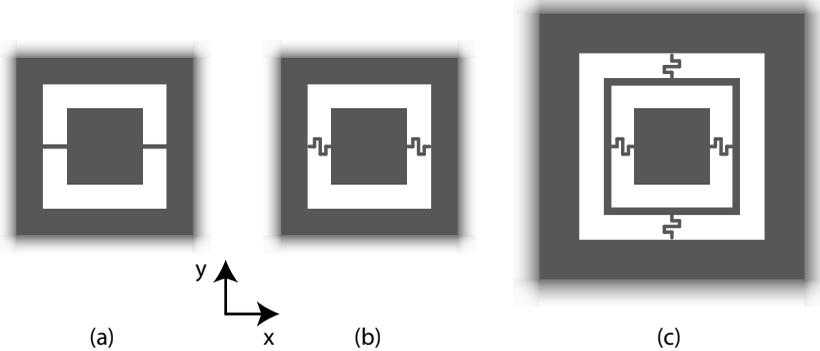


Figure 6.10 Structural features of scanning micromirrors. A one-dimensional mirror with a torsion bar mount for movement in the y direction (*a*); one-dimensional mirror with a spring mount for movement in the y direction (*b*); and a dual-spring gimbal mount with a moving intermediate frame for two-dimensional movement in both the x and y directions (*c*). Not to scale.

Figure 6.10 (*c*), the mirror is mounted in a frame which is turn connected to the substrate by an orthogonal set of springs; this arrangement, known as a gimbal mount, allows two-dimensional movement in both the x and y directions. Independent motion in the x and y (planar) directions then allows a freely positionable mirror usable for a surface scan. Internal cross-bars have also been used to achieve two-dimensional scanning (Kim et al., 2003).

The sketches of Figure 6.10 are not to scale. An SEM (scanning electron microscope) picture of a real one-dimensional scanning micromirror is shown in 6.11, and a light microscope picture of a two-dimensional mirror is presented in Figure 6.12. We see from these structures that typical mirror dimensions range from about $100\ \mu\text{m}$ to $1000\ \mu\text{m}$ on a side, and that the mounting springs have typical widths from 1 to 2 micrometers.

Actuation

Stimulation and control of mirror movement may be by means of one of numerous approaches, whereby electrostatic actuation is likely the most popular. By applying an electric potential between the metal of the mirror and electrodes mounted on the substrate, an attractive force is generated, resulting in a bending of the mirror toward the electrode; the mirrors of Figures 6.11 and 6.12 are actuated in this manner, so that the electrodes are not visible below the mirror. Popular alternatives include comb drives (Jeong and Lee, 2005), electromagnetic (Su et al., 2005; Yalcinkaya et al., 2006) or magnetic actuation (Ji et al., 2007), thermal actuation (Singh et al., 2005), piezo-electric actuation (Kotera et al., 2005) and pneumatic actuation (Werber and Zappe, 2006a,b).

Actuation and movement of micromirrors generally takes place in one of two modes: steady-state beam steering or resonant operation. For beam steering, the actuator positions the mirror in a desired position, at which it is stably held; using various approaches for feedback-control, such positioning can be quite accurate, but is sensitive to shock

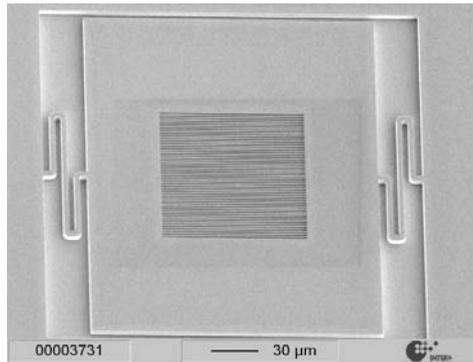


Figure 6.11 SEM photograph of a 180 μm wide silicon micromirror for one-dimensional scanning, attached to the substrate by the springs on the left and right; angular mirror motion is in the vertical (y) direction. The grating structure in the center is a special feature used for position measurement and is not usually present on a standard micromirror; see Section 6.6.2. Photo courtesy of David Kallweit.

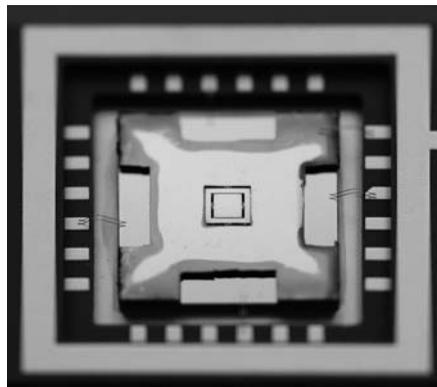


Figure 6.12 A packaged two-dimensionally scanning silicon micromirror; the 500 μm wide mirror in the center is mounted to a movable frame (springs left and right) which is in turn mounted to the substrate using the springs visible at the top and bottom of the frame. Photo courtesy of Khaled Aljaseem and Andreas Fischer.

and vibration, as well as variations in actuation bias. In addition, the maximum achievable angular movement is limited; for most electrostatically actuated mirrors, the out-of-plane motion is several degrees at most (Kim et al., 2002), although electromagnetic or pneumatic actuation can result in mirror tilt of up to 75° (Werber and Zappe, 2006a) or stroke (purely vertical) motion of over $80\ \mu\text{m}$; we will examine these structures in more detail in Chapter 12.

In resonant operation, the micromirror is actuated using a time-dependent stimulus at the mechanical resonance frequency of the mirror/spring system. The maximum angular movement can be much larger in this mode of operation, but the movement is continuous and generally sinusoidal in time. Typical resonant frequencies, a strong function of material, dimensions and structure of the mechanics and the mirror, are in the range

of kHz to tens of kHz. Which operating mode is to be preferred depends on the application; pure scanning applications often use resonant operation to profit from the large achievable scan angle.

6.4.3 Optical characteristics

The optical function of a mirror is relatively simple, but a number of characteristics play a role in determining the optical performance of a micromirror. Its reflectance is typically given by the surface coating, typically a metal, but the details of the surface and the mirror shape and size also play a role in determining mirror merit.

Planarity and roughness

Optical quality is defined by planarity and roughness. The former refers to the deviation of the mirror from an ideal plane, and these deviations should be less than $\lambda/10$ for a given application. A deviation in planarity of less than $1\ \mu\text{m}$ over the surface of a $500\ \mu\text{m}$ mirror is typical for single-crystal silicon structures, but may be worse for other materials, particularly if these are thin, less than $10\ \mu\text{m}$. Parasitic mirror curvature may also be asymmetric, making it difficult to correct, and also degrades under dynamic operation: the mirror deforms when deflected, particularly for resonant mode operation (Conant, 2003, Chapter 1).

Roughness refers to surface deviations on a small size scale, typically in the tens to hundreds of nanometers. Roughness, usually a result of material surface properties or the characteristics of surface films, should also be significantly less than $\lambda/10$, and, for good mirrors, can be less than $10\ \text{nm}$. A typical profile of mirror roughness and deformation is seen in Figure 6.13, which is a white-light interferogram of a single-crystal silicon micromirror with an edge length of $500\ \mu\text{m}$. The overall concave shape of the mirror is clearly seen, as is the roughness, the latter on a size scale of several $10\ \text{nm}$; the overall deviation from a planar surface is about $110\ \text{nm}$, corresponding to about $\lambda/5$ for visible wavelengths.

Aperture

Micromirrors have a limited dimension and the aperture (i.e., the mirror size) also defines the optical far field, meaning the light pattern generated far from the mirror surface. Due to diffraction, as we will see in detail in Chapter 8, the intensity pattern of light reflected from or transmitted through an aperture is given by the spatial Fourier transform of the aperture. For a micromirror placed in an imaging optical system, the far-field image generated will thus be the Fraunhofer diffraction pattern² of the mirror

² Joseph von Fraunhofer (1787–1826), was buried under a collapsed workshop when an apprentice lens-maker in Munich, and the compensation payment he received due to this accident allowed him to lead a life of intellectual independence. This great German physicist was primarily interested in the design of achromatic lenses but his most lasting achievement was the discovery of the spectral absorption lines which today bear his name and his establishment of the spectroscope as a serious scientific instrument.

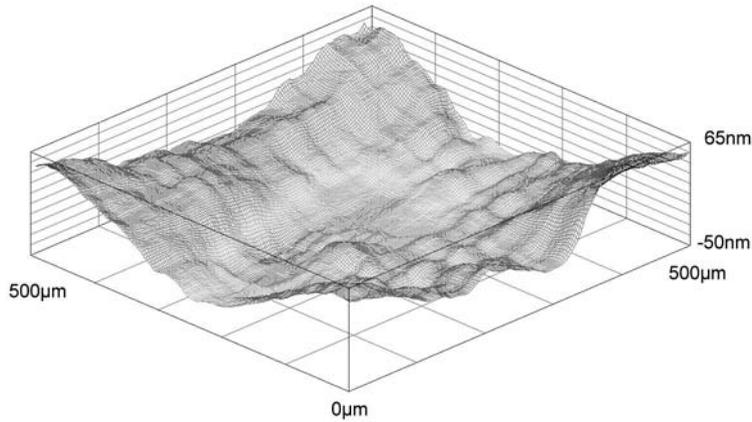


Figure 6.13 Surface profile of a 500 μm single-crystal silicon micromirror, showing roughness and large-scale mirror deformation. The mirror is seen to curve inwards toward the center (a concave shape) and total deviation from planarity is about 110 nm, which for visible wavelengths is about $\lambda/5$. Plot courtesy of David Kallweit.

shape: the image of a focussed spot, when reflected from a micromirror, will thus have a width and structure which is a function of mirror size and wavelength.

Employing some of the concepts of Chapter 8 in advance, let us consider a tilting micromirror used to image a point source, as shown in the cross-section of Figure 6.14. Using basic diffraction theory, we find that the two-dimensional intensity pattern in the xy plane, $I(x, y)$ [W/m^2], generated by a rectangular mirror with dimensions $w_x \times w_y$ [m^2] at a distance d [m] from the mirror surface is given by

$$I(x, y) = I_0 \left[\frac{\sin\left(\frac{\pi w_x}{d\lambda_0} x\right)}{\left(\frac{\pi w_x}{d\lambda_0} x\right)} \right]^2 \left[\frac{\sin\left(\frac{\pi w_y}{d\lambda_0} y\right)}{\left(\frac{\pi w_y}{d\lambda_0} y\right)} \right]^2 \quad (6.6)$$

for illumination wavelength λ_0 and maximum intensity I_0 [W/m^2]. Alternatively, the diffraction pattern of a circular mirror with radius r_0 [m], as a function of the radial variable r [m], is similarly found from

$$I(r) = I_0 \left[\frac{2J_1\left(\frac{2\pi r_0}{d\lambda_0} r\right)}{\left(\frac{2\pi r_0}{d\lambda_0} r\right)} \right]^2 \quad (6.7)$$

where J_1 is a first order Bessel function³ of the first kind and d [m] is again the distance of the image from the mirror surface. These expressions for rectangular and circular diffraction patterns will be derived in Sections 8.2.3 and 8.2.4, respectively.

³ Friedrich Wilhelm Bessel (1784–1846) was a German astronomer and the first to measure the distance to a star using parallax; he also predicted the existence of Neptune, which was only discovered after his death. His eponymous functions were developed to help explain perturbations in planetary motion but now have a wide range of other applications.

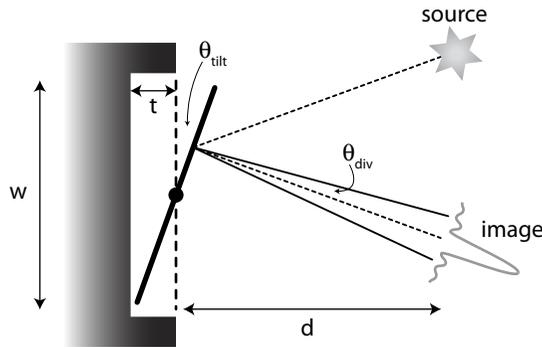


Figure 6.14 Schematic cross-section of a tilting mirror used to image a point source; the dotted line from source to image denotes the path of the ray, but in reality an optical system would generate a (Gaussian) beam which comes to a focus at the image point. The mirror of width w has a maximum tilt angle θ_{tilt} defined by the depth of the cavity below it, t ; the beam reflected from the mirror diverges due to diffraction with an angle of θ_{div} and the image is a distance d from the mirror. Dimensions and angles are not to scale; typical values might be $w = 500 \mu\text{m}$; $d = 1 \text{ mm}$; $t = 20 \mu\text{m}$; $\theta_{\text{tilt}} = 5^\circ$; and $\theta_{\text{div}} = 0.06^\circ$.

Both of the previous expressions predict periodically-varying intensity distributions with a high intensity peak at the center but with numerous satellite maxima which can expand the image of a point considerably. As seen in the calculated diffraction pattern of a circular mirror given in Figure 6.15, the intensity maximum is surrounded by concentric rings of reduced but non-zero intensity for a circularly-symmetric structure; this pattern is typically referred to as an Airy disk⁴. The width of the diffraction spot is typically defined as the spacing between the central maximum and the first dark ring in the Airy disk, which is given by

$$\Delta r = 1.22 \frac{\lambda_0 d}{2r_0} \quad (6.8)$$

for a circular aperture; for a square mirror of width w [m], the spacing to the first minimum is

$$\Delta x = \frac{\lambda_0 d}{w}. \quad (6.9)$$

The factor 1.22 derives from consideration of basic diffraction theory in circular coordinates; see Section 8.2.4.

Example 6.4: Consider a circular micromirror with a diameter of $600 \mu\text{m}$ illuminated with a laboratory standard HeNe laser. What is the spot size on a screen placed 10 cm

⁴ George Airy (1801-1892) was the British Astronomer Royal for 46 years. Active in optics and geophysics, he developed cylindrical lenses to correct astigmatism, developed the concept of the Airy disk due to diffraction, and proposed new explanations for the gravitational anomalies found in mountain ranges.

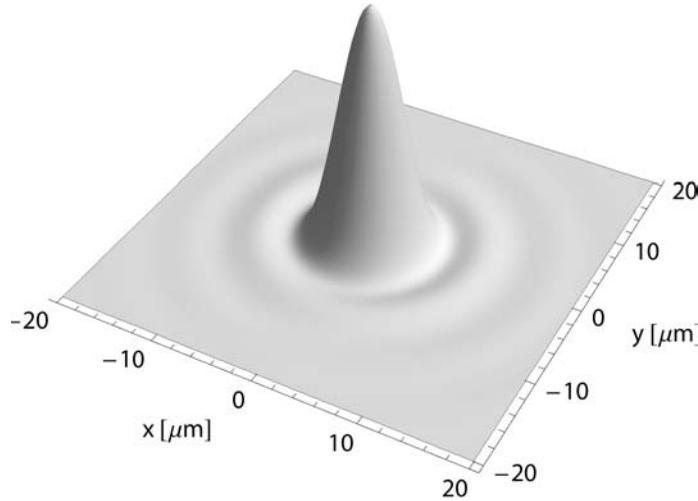


Figure 6.15 Calculated diffraction pattern generated by a circular mirror with radius $r = 100 \mu\text{m}$ illuminated by a HeNe laser, as imaged 1 mm from the mirror surface. The first ring surrounding the peak is the first maximum of the Airy disk.

from the mirror? How about on the laboratory wall, 5 m away? What happens if we use a more macroscopic mirror, with a radius of 5 mm?

We can directly apply the definition of the Airy disk, Equation 6.8, to determine the size of the spot; the distance from the center to the first dark ring is usually employed as a definition of diameter. For the $600 \mu\text{m}$ micromirror, we see that the spot has a diameter of about $64 \mu\text{m}$ for the screen 10 cm away, but around 3.2 mm on the wall, purely due to diffraction from the limited-size aperture of the mirror.

The spot size scales inversely with mirror radius, so for the larger mirror we have considerably smaller spots, a bit less than $8 \mu\text{m}$ for the screen and less than 0.4 mm for the wall. Increasing mirror size is thus always advantageous in reducing diffraction effects.

Divergence angle

The previous relationships may also be expressed as a divergence angle due to diffraction, as seen in Figure 6.15, given by θ_{div} [rad], with a value

$$\theta_{div} \approx \arctan \frac{\Delta x}{d} \approx \frac{\lambda_0}{w} \quad (6.10)$$

for a square mirror of width w [m], and

$$\theta_{div} = 1.22 \frac{\lambda_0}{2r_0} \quad (6.11)$$

for a circular mirror. The values for divergence angle are generally small; for a square mirror with $w = 500 \mu\text{m}$, a typical value is $\theta_{div} \approx 10^{-3}$ rad.

Resolution

Based on these considerations, we can define the resolution of a scanning micromirror; resolution in this case means how closely adjacent spots imaged by the mirror may be spaced such that they may still be distinguished from each other. Two Airy disks (spots) are commonly defined to be distinguishable if the maximum of one is farther than the first dark ring (the first minimum) from the other. Using this definition and the definition of divergence angle from above, we can determine the number of resolvable spots for a square mirror with maximum tilt angle θ_{tilt} (in both directions) as

$$N_{spots} = \frac{2\theta_{tilt}}{\theta_{div}} = \frac{2 \arcsin \frac{2t}{w}}{\frac{\lambda_0}{w}} \approx \frac{2t}{w} \frac{2w}{\lambda_0} = \frac{8t}{\lambda_0} \quad (6.12)$$

for maximum mirror vertical movement t [m], as defined in Figure 6.14; for a circular mirror, the factor 1.22 is included, so that

$$N_{spots} = \frac{8t}{1.22\lambda_0}. \quad (6.13)$$

We note that the previous expressions are only a function of the maximum tilt (given implicitly by the maximum vertical movement t) and the wavelength; mirror size does not appear since, as the mirror becomes larger for a given t , the divergence decreases but the maximum scan angle decreases as well.

As a result, we can summarize the demands on micromirror design to yield optimal optical properties. For small divergence and thus the imaging of as small spots as possible, Equations 6.10 and 6.11 tell us that we need either a small mirror/image spacing, d , which is not always practical; a large mirror, with large w or r_0 ; or a short wavelength, λ_0 . Since d and λ_0 are often fixed for a particular application, small divergence implies that the mirror should be as large as possible. We see thus that decreasing mirror size may be of advantage mechanically (yielding higher scan resonance frequencies and reducing required actuation voltages or powers) but is typically a disadvantage optically.

For scanning applications in which a high resolution is required, it is maximum scan angle, θ_{tilt} , which is desired. For a given mirror size w or r_0 , this requirement implies as large a value for t as possible, and it is for this reason that considerable research effort has been extended to develop out-of-plane mirror designs. For both considerations, low divergence and high resolution, reducing the wavelength is an advantage which, insofar as the application allows it, always improves micromirror optical performance.

6.5 Adaptive micro-optics

Many macro- or micro-optical imaging applications result from the transmission of the optical field through optically inhomogeneous media. Earth-bound astronomers must image through the atmosphere and ophthalmologists can only view the retina through

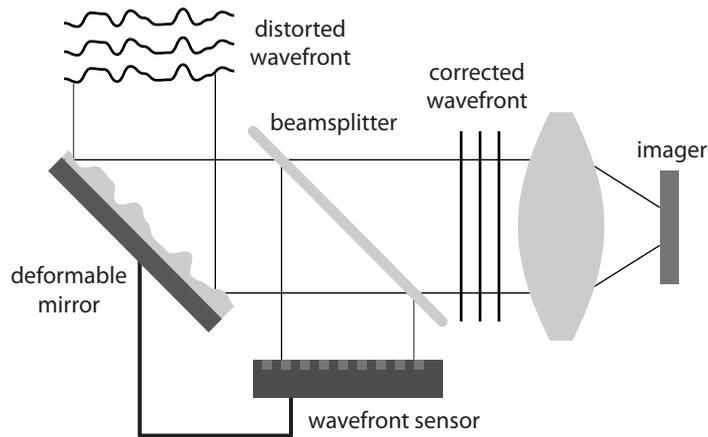


Figure 6.16 Correction of a distorted wavefront using a deformable adaptive optical mirror. A portion of the optical field is directed onto the wavefront sensor, which controls the deformation of the adaptive optical mirror, whose surface adapts to compensate the wavefront distortion and thereby generates an (ideally) planar wavefront.

the vitreous humor inside the eye: both of these media are subject to spatially and time dependent variation of the optical path length through them. As a result, an ideally planar wavefront from which an image is generated becomes distorted, limiting the resolution of the image.

Adaptive optical systems are those whose characteristics are adjusted to compensate for wavefront distortion; they “adapt” to the momentary shape of the wavefront and thus allow dynamic correction of distortion. Key components in adaptive optical systems are deformable mirrors, frequently micromirrors, and we discuss the structures and their implementation here. Useful for this discussion is the characterization and analysis of the wavefront as discussed in Section 2.5.

The use of adaptive optics for wavefront correction has proven to be of value for a number of applications. In ophthalmology, the distortion due to the liquid inside the eye is compensated using adaptive optics to allow high-resolution retinal imaging (Doble et al., 2002); in astronomy, atmospheric distortion is likewise corrected with adaptive optical systems (van Damm et al., 2004), allowing imaging resolution which rivals that of space-based telescopes; and for laser-based free-space optical communications (Tyson, 2002) as well as for beam-shaping of industrial lasers for machining (Vinevich et al., 2004), adaptive optics is playing an increasingly important role.

6.5.1 Wavefront correction

The principle behind wavefront correction is illustrated in Figure 6.16. A distorted wavefront, light emitted from a star or reflected from the retina, is incident on a deformable mirror. A portion of this field is projected onto a wavefront sensor, which calculates the spatial deviation of the wavefront from the planar ideal. This sensor then transmits control signals to the deformable mirror which changes its shape, resulting in

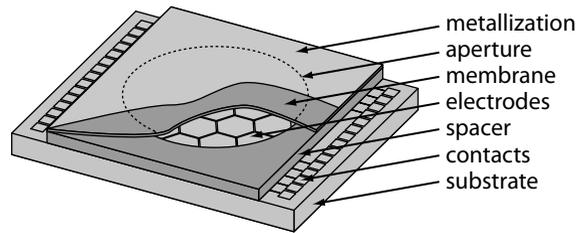


Figure 6.17 Cross-sectional view of a typical deformable membrane micromirror. The electrode array, arranged in a circle under the metallized membrane, is used to deform the mirror by pulling the membrane downwards. The membrane may be a metal, dielectric or polymer. Sketch courtesy of Christoph Friese

localized phase shifts which compensate for the measured distortion. The sensor and deformable mirror are linked in a feedback loop, so that the mirror continuously adapts to changes in the wavefront, allowing a dynamic optical correction.

The wavefront sensor is typically a Shack-Hartmann configuration⁵. The Shack-Hartman sensor uses a two-dimensional array of microlenses arranged over an array of photodetectors, such that any deviation from a planar wavefront gives rise to a spatially-varying intensity distribution from which the wavefront shape can be calculated; the distortion is frequently expressed by the corresponding Zernike polynomials. Other types of wavefront sensors exist, whereby measurement and calculation speed is of essence when employed in an adaptive optical system.

Once the deviation of the wavefront has been determined, the adaptive optical mirror endeavors to correct it. Correction is accomplished by a localized shift in the vertical (normal to the wavefront) position of the mirror, thereby compensating the phase shift which leads to the distortion. Adaptive optical mirrors are thus dynamically movable components, for which the mirror surface shifts to compensate for changes in the measured wavefront.

6.5.2 Adaptive optical mirrors

A variety of adaptive optical mirror concepts has been proposed and demonstrated. Two basic configurations have been employed: a two-dimensional array of individual mirrors which move vertically (piston motion); or a continuous planar membrane mirror whose shape is distorted using a two-dimensional array of actuators, usually underneath it. For both of these approaches, three parameters are of primary interest in defining the performance of the system: resolution (how accurately the vertical position may be defined); stroke (the maximum total vertical motion possible); and speed of response.

⁵ The Shack-Hartmann sensor as it is configured today was invented by Ronald Shack of the University of Arizona in the late 1960s (Platt and Shack, 2001). Johannes Hartmann (1865-1936) was a German astronomer, professor in Potsdam, who invented the “Hartmann test” to characterize refractor telescope optics, the conceptual basis for the Shack-Hartmann sensor as configured today.

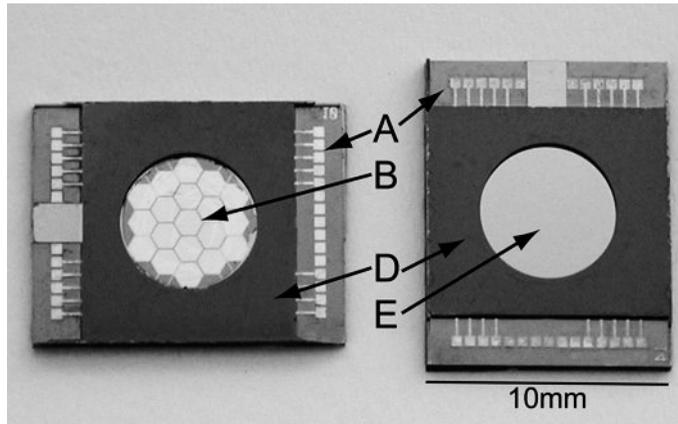


Figure 6.18 Electrode structure *left* and deformable SU-8 mirror membrane *right* of a polymer-based continuously deformable adaptive optical mirror. Shown are the electrode contacts *A*; the hexagonal electrodes *B*; the silicon frame *D*; and the reflecting membrane *E*. Photo courtesy of Christoph Friese

Piston mirror arrays

The classical approach to wavefront correction using adaptive optics was to employ individual mirrors, each independently deformable, arranged in a two-dimensional array (Doble et al., 2002). These mirrors, typically actuated using piezoelectric actuators, tend to be macroscopic components, with diameters in the range of 4 to 8 cm, so that the entire assembly is decidedly not a micro-optical system. Large system size, coupled with the high cost per actuator, has made this approach acceptable for certain high-end applications (notably astronomy, the field which has provided the primary impetus for these developments) but impractical for widespread use in other areas.

Deformable mirrors

Microsystems fabrication techniques have led to a new generation of miniaturized, continuously deformable membrane mirrors for adaptive optics. The basic structure is shown in Figure 6.17; a microfabricated metallized membrane is suspended over a two-dimensional array of electrodes, separated from these by a spacer. Application of a bias between membrane and any given electrode pulls the membrane mirror downward, such that a pattern of variable applied voltages may be applied to deform the mirror to the required shape.

A variety of structures and materials has been employed for micro-optical implementations of deformable membrane micromirrors. Using a SiN membrane metallized with Cr/Au, for example, a 700 μm diameter mirror with a single, central actuator has been demonstrated, with a stroke of 3.5 μm for an applied bias of 200 V (Shao et al., 2004). An array of deformable micromirrors has been fabricated using polysilicon as the membrane material (Bifano et al., 1997); 10×10 arrays of mirrors spaced at 300 μm have been fabricated, with fill-factors on the order of 99%. Alternatively, continuous polysilicon membranes using a two dimensional array of 4 096 actuators with 99.5% fill factor

have been operated with 4 μm stroke at bandwidths of greater than 5 kHz (Cornelissen et al., 2009). Finally, a single $10 \times 10 \text{ mm}^2$ Al-coated SiN membrane has been spanned over a two-dimensional array of 37 electrodes (Vdovin and Sarro, 1994) for continuous deformation; diameters up to 50 mm have been realized using this technology.

Of considerable interest in the fabrication of deformable mirrors is the use of polymers. These allow considerable flexibility in setting membrane and spacer thicknesses and may be subject to easier processing than standard silicon-based materials; the mechanical properties of some polymers are excellent for manufacturing flat, stress-free membranes. SU-8, for example, is a photostructurable polymer (Conradie and Moore, 2002) which is well established in microsystems fabrication. Employing a hybrid assembly approach using a glass substrate and a silicon frame for the membrane, as shown in Figure 6.17, an SU-8 deformable mirror with a Au/Cr high-reflection coating has been demonstrated (Friese and Zappe, 2008). The polymer film, fabricated with thicknesses ranging from 7.5 μm to 9 μm , was suspended over an array of 19 hexagonal electrodes, as seen in the photograph of Figure 6.18. A polymer spacer between mirror and electrodes, also of variable thickness, allowed optimization of the spacing for either high stroke/high voltage, or low stroke/low voltage.

These 5 mm diameter mirrors could be vertically actuated by up to 18 μm , depending on the membrane thickness, the spacer thickness and the applied voltage. For low voltage operation, a 7.5 μm membrane could be deformed by up to 6 μm for 125 V; the high deformation values were reached for thick membranes (9 μm) and high voltages (400 V). Using these mirrors, lower-order aberrations (those defined by lower order Zernike coefficients) could be corrected, and the technology has been shown to be suitable for membrane mirrors up to 15 mm in diameter.

6.6 Micromirrors: case studies

For the various optical components which we introduce and analyze in this and the following chapters, we will consider a few examples of their implementation, if possible as part of a commercial application, as “case studies”. These concluding sections will give students a feel for the relevance of the technologies and components in both academic and industrial environments.

6.6.1 The DMD

The Digital Micromirror Device, or DMD, a trademark of Texas Instruments, is likely the most successful and widely employed micro-optical component, if we exclude optical disc reader heads as not being really “micro-optics.” We have already been introduced to the DMD in the introductory Section 1.3, where we proposed that it is the one micro-optical component that readers are likely to have in their home.

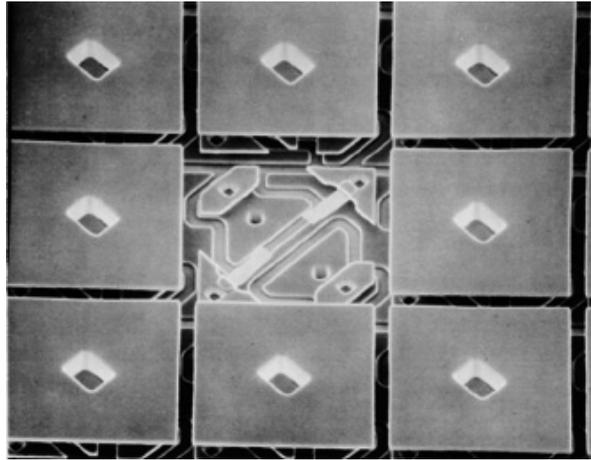


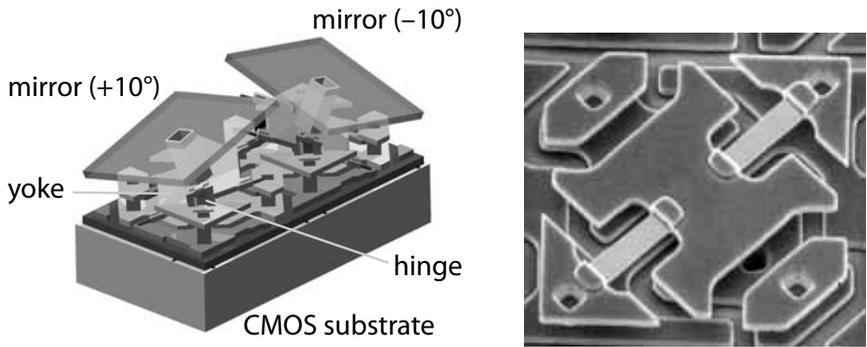
Figure 6.19 A close-up view of nine DMD mirrors; the center one is missing to show the underlying micromechanics. Each mirror has a size $16\ \mu\text{m} \times 16\ \mu\text{m}$ and spacing between mirrors is $17\ \mu\text{m}$. Photo © Texas Instruments; reprinted courtesy of Texas Instruments.

DMD structure

After about ten years of research and industrial development, the DMD premiered in 1985 (Hornbeck, 1993; Van Kessel et al., 1998) and it was first used in digital projectors in 1996. The Digital Micromirror Device is based on an optically simple $16\ \mu\text{m} \times 16\ \mu\text{m}$ aluminum micromirror, shown in Figure 6.19, actuated by aluminum-based micromechanics, shown schematically in Figure 6.20. The mirror and the mechanics are in turn fabricated on a CMOS-based SRAM (static random access memory) substrate. Each SRAM memory cell provides a voltage bias on the electrostatically-deflectable micromirror, moving it to one of two angular positions, at $\pm 10^\circ$. The fact that the mirror only has two stable positions (-10° at rest; $+10^\circ$ when biased) leads to the designation “digital”; the device is not continuously positionable. These mirrors are then fabricated in large arrays, originally spaced every $17\ \mu\text{m}$ but now with a pitch of $13.8\ \mu\text{m}$ and correspondingly smaller mirrors. Since the primary application is for displays, popular display resolutions have defined the array sizes: SXGA, with 1280×1024 pixels, with 1 310 720 individual mirrors, is the largest commercially available.

DMD projectors

The use of the DMD chip for projection is based on a clever optical approach, one that yields a robust system as well as one which generates excellent color saturation. In the one-chip variant, illustrated in Figure 6.21, time-division multiplexing is used to generate a broad color spectrum. The micromirror array is illuminated with a broadband (in the visible) metal halide arc lamp, transmitted through a rotating color wheel, which alternately transmits the primary colors red, green and blue (RGB); the wheel rotates at 60 rpm, and for every rotation, each color is transmitted for 1/3 of the cycle, thus for 5.56 ms. Depending on the position of an individual DMD micromirror, this light will be transmitted through the projection optics or deflected to an absorber: at



(a) Schematic depiction of two DMD mirrors in the $\pm 10^\circ$ positions. (b) SEM photograph of the mirror mechanics, with the mirror removed.

Figure 6.20 Mechanics of the DMD, schematically and in reality. Both © Texas Instruments; reprinted courtesy of Texas Instruments.

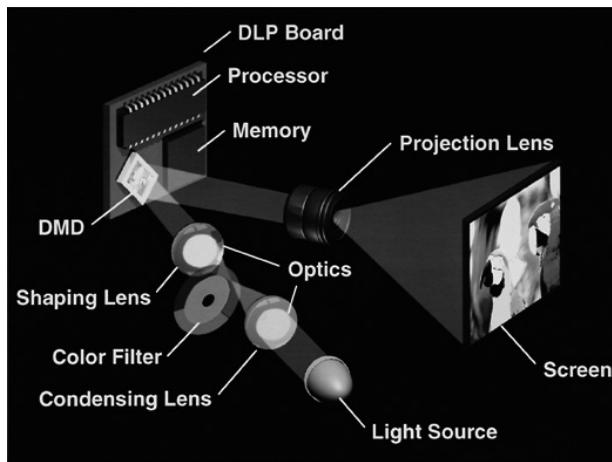


Figure 6.21 Concept of a DMD-based projection system. Diagram © Texas Instruments; reprinted courtesy of Texas Instruments.

the -10° position, no light is transmitted (the pixel is black), and in the $+10^\circ$ position, red, green or blue light (depending in the momentary position of the color wheel) is projected for that pixel. An alternative approach uses three DMDs in a single projector, one for each primary color.

Since the mirrors switch in $20 \mu\text{s}$, the amount of time the mirror spends in the $+10^\circ$ position during the 5.56 ms at which a certain primary color is projected defines the intensity of that color for that pixel. For 24 bit color (8 bits, or 256 gray levels, per color), the mirrors are thus in the “on” state for anywhere from $20 \mu\text{s}$ to $5\,560 \mu\text{s}$ per color, in $20 \mu\text{s}$ intervals; the longer the “on” state, the brighter that color component for that pixel and full on/off contrast ratio is up to 575:1. This process takes place at video

frame rates such that the eye sees continuous RGB color for each pixel, with more than one billion color and gray scale combinations for each of the 1.3 million pixels (Hornbeck, 1998).

The DMD shows how a conceptually simple micro-optical device, a tiltable mirror with two stable positions, can provide high functionality when part of a highly complex microsystem. It is likely that not even the inventors of this component could have envisaged the scope of its technical and commercial success.

6.6.2 Angular control

The DMD mirrors have two stable positions, at $\pm 10^\circ$, defined by landing tips on which the mirror rests when in one of the two actuated states. For freely rotatable micromirrors, however, determination and control of the angular position of the mirror is an important function; key to achieving this is having a means for angular position sensing. With knowledge of the instantaneous angular position, a feedback network may be employed to assure that the mirror remains in that position, or, when used in resonant mode, that the maximum scan angle is stable. Accurate angular positioning or measurement of maximum angular movement are essential for scanning, printing or optical cross-connect applications.

A number of approaches for sensing angular position for micromirrors have been proposed and demonstrated. Electric or electromagnetic concepts usually use changes in resistivity or field distribution as the mirror rotates; optical approaches generally employ changes in reflection or diffraction as a function of angular position.

Electrical rotation measurement

Measurement of rotation angle using electromagnetic means can use the change in electrical properties of the mirror supports as they undergo rotation. The strain induced due to the torsional deformation of the mirror spring gives rise to a piezoelectric effect: the shear stress results in a change in resistivity. Using a current injected through contacts oriented transversely to the micromirror mount, measurement of the induced voltage normal to this direction has been shown to provide a useful electrical signal for sensing rotational position (Sasaki et al., 2006). With a p-type silicon substrate for the mirror structure, a voltage change of about $1 \text{ mV}/^\circ$ of rotation was measured, a sensor signal which may then be used to determine instantaneous angular position.

Alternatively, electromagnetically-actuated micromirrors, which use a combination of integrated coils and external electromagnetic fields to stimulate mirror movement, may use magnetic field sensing of mirror position. Using an integrated coil on the mirror surface and external fixed magnets, movement of the mirror results in a generated current in the sensing coil and thus a signal for determination of mirror movement (Miyajima et al., 2003). Using this sensor in a closed-loop system, it has been shown that, following a warm-up period of about 15 minutes, scan angle fluctuation can be reduced to below 0.02% by monitoring mirror movement in this fashion.

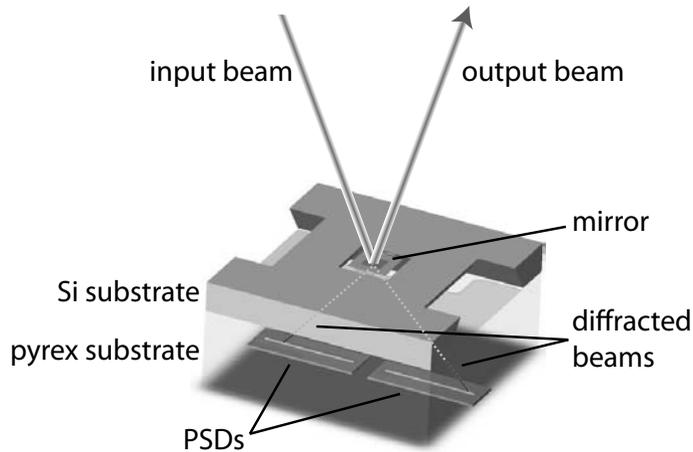


Figure 6.22 Schematic drawing of the optical feedback sensor for determination of mirror angular position. As the mirror rotates, the diffracted orders transmitted through the grating move in opposite directions on the two PSDs, generating a high-resolution angle-dependent signal. Diagram courtesy of David Kallweit.

Optical rotation measurement

Optical techniques for micromirror angular control are attractive due to the high resolution which may be obtained using light-based approaches; most proposals, however, rely on a second optical system, unrelated to the signal being scanned or switched, on which the measurement is based. In contrast, a diffractive angular position measurement which uses a small fraction of the steered light beam has been shown to provide high accuracy while using a relatively simple additional measurement network.

The sensor is based on a transmission grating etched into the mirror surface, as was seen in Figure 6.11 (Kallweit and Zappe, 2006); whereas the example in the figure shows a large grating, the function may be implemented with the grating covering only a small fraction of the mirror surface. As shown in the schematic illustration of Figure 6.22, a portion of the reflected light is transmitted and diffracted by the grating, such that two complementary diffraction orders impinge on the pair of position-sensitive detectors (PSDs) located below the mirror, with a transparent Pyrex substrate as a fixed spacer between them.

As the mirror rotates, the two diffraction orders (the ± 1 or ± 2 orders; we will revisit the concept of diffraction in Chapter 8) generate optical beams which move in opposite directions on the two PSDs, thus generating a sensitive angle-dependent signal. With typical mirror sizes of 200 to 500 μm , thicknesses of 4 to 7 μm and grating periods of 2 μm , an angular resolution of 0.03° could be reached when using a 650 nm laser diode for illumination; typical spacing between mirror and the PSD was 7.9 mm. Due to the high sensitivity of the PSDs, only a small fraction of the incident light is required to obtain a reliable measurement signal: 100 μW is sufficient, such that only 1% of a 10 mW input signal needs to be transmitted through the grating.

This optical angle-sensing approach has been used to stabilize the scanning micromirror in a desired position; immunity from external shock and vibration may thus be achieved. In addition, a desired rotation angle may be set without the need for calibration and software-based lookup tables, usually necessary since, even for mirrors fabricated on the same substrate, process variations during fabrication imply that no two mirrors ever behave identically for a given input signal.

6.6.3 Optical switching

A final important application for micromirrors, one which has also provided considerable impetus for mirror development but in products that consumers are unlikely to purchase for home use, is in the area of optical switching. In optical telecommunications systems, the backbone for most data transmission worldwide, it is frequently necessary to switch the optical signals from an arbitrary input to an arbitrary output; this function arises often in network reconfiguration and data routing.

It is of considerable advantage to switch the optical signals directly, as opposed to converting them to electrical signals, switching these electrically, and subsequently reconverting to an optical signal. An optical switch is thus configured to connect N input signals (typically optical fiber outputs) to N arbitrary output ports (also typically fibers), a function known as optical cross-connect (OXC); the switch is designed to allow a dynamic change of the connected input and output ports, such that the optical data stream may be reconfigured within a few milliseconds. Micromirrors have proven to play a significant role in the development of such high-capacity all-optical switching systems (Marom, 2007).

Optical crossbars

Two approaches for OXC have enjoyed some popularity. Optical crossbars use one-dimensional arrays of input and output fibers, oriented 90° to each other. Between them is a two-dimensional array of pop-up micromirrors, whose surface is oriented at 45° to the input and output arrays. If a given micromirror is in its vertical position, the signal is diverted 90° to the appropriate output fiber; if the mirror remains flat on the substrate, on the other hand, the optical signal continues unreflected until it reaches an actuated (vertical) mirror.

The demands on mirror motion are considerable: although only two stable positions are needed, actuator movement of more than $20\ \mu\text{m}$ and mirror angular movement of 90° are required. An additional problem with optical crossbars is that they scale poorly. For N input and output signals, the system scales as N^2 , meaning N^2 mirrors are needed to switch all possible inputs with all possible outputs. For a 32×32 system, mirrors with a diameter of $640\ \mu\text{m}$ and a lateral chip size of $3.2\ \text{cm}$ are required (Marom, 2007, Chapter 3.07.4.4.3); scaling to larger input port counts is unlikely to be practical.

3D $N \times N$ switches

An alternative approach to optical crossbars is the three-dimensional $N \times N$ switch, which scales as $2N$ with input/output port count, and uses gimbal-mounted, three

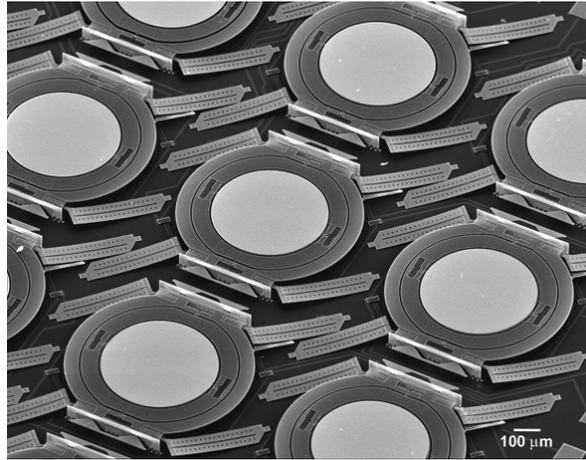


Figure 6.23 An array of three-dimensionally movable micromirrors used for OXC. The mirrors, the light disks at the center with a diameter of $600\ \mu\text{m}$, are mounted on gimbals, allowing three-dimensional positioning. Photo © Alcatel Lucent; reprinted courtesy of Alcatel Lucent.

dimensionally-positionable micromirrors, such as those developed by Bell Laboratories (then a part of Lucent Technologies) (Bishop et al., 2002); an array of these is shown in Figure 6.23. These mirrors, fabricated using silicon MEMS technology and with a metal reflecting surface, couple signals from a two-dimensional fiber array into an arbitrary output fiber in a two-dimensional output fiber array; two two-dimensional mirror arrays, each with N mirrors, are required for a complete switching system, hence the scaling as $2N$ rather than N^2 . The individual mirrors have a diameter of $600\ \mu\text{m}$ and are switchable in less than 10 ms.

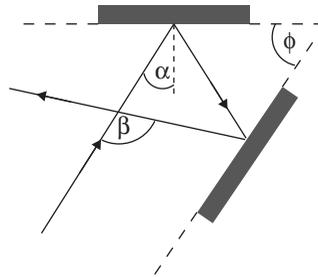
An OXC system based on these micromirrors is relatively compact. For 256 input/output ports, 2 chips (with an area of less than $6.5\ \text{cm}^2$) are employed, spaced by 8.6 cm; mirror angular movement required is less than 10° . However, despite this highly-advanced MEMS micromirror technology which led to optical switching system performance speeds of up to 10 TB/s, the OXC systems based on this technology were no longer marketed after 2002. Alternative technologies, using, for example, electrostatic comb drives for larger angular motion (Yano et al., 2005), have been developed and the performance of entire OXC systems based on MEMS mirrors continues to improve (Zheng et al., 2003). The extremely high demands on component performance for micromirror-based optical switching systems will provide considerable stimulus for continued refinement of micromirror technology.

Problems

1. Concerning Figure 6.1: what is wrong with this picture?
2. You have a thermal evaporator for Au which can be approximated as a hemisphere with a radius of 0.15 m and use this to deposit a $2\ \mu\text{m}$ thick reflecting layer on a wafer with 200 micromirrors, each $100 \times 100\ \mu\text{m}$. Using the current price of gold (density

$= 19.32 \text{ g/cm}^3$), what is the value of the Au layer on each micromirror? What is the ratio of this cost to the total material cost for one evaporation?

3. Use ray tracing to show that the image of your left hand in a planar mirror has the same topology as your right hand.
4. Derive the expression for the total number of possible reflections due to a misaligned beam in a Fabry-Perot resonator given in the first part of Example 6.1, using a simple geometric sketch. Plot the allowed angular error of the optical beam as a function of cavity length (up to 25 cm) for 50, 100 and 200 reflections, assuming circular mirrors with $r = 10 \text{ cm}$.
5. Sketch the geometric configuration of the misaligned mirror in the Fabry-Perot resonator given in the second part of Example 6.1 and derive the expression for the total lateral displacement of the beam given in the example. Assume the tilt angle is sufficiently small so that the mirror spacing is approximately the same for all reflections, even though the mirrors are not parallel.
6. Consider the mirror arrangement shown in the figure below: two mirrors are oriented at an angle ϕ to each other.



What is the angle between the incident and reflected beams, β ? For what ϕ does this arrangement act as a retroreflector, i.e., $\beta = 180^\circ$?

7. Construct a Sagnac interferometer of the form shown in Figure 5.14, replacing the beam splitter with a fourth mirror. Assuming the mirrors to be silver-coated and using a visible wavelength, how many loops around the interferometer can the light make before its intensity is reduced by 99%?
8. Considering the same Sagnac interferometer as the in the previous problem, assume the mirrors are perfect reflectors ($R = 1$), but are misaligned by 0.01° , all in the same direction. Given that the mirrors are square with $w = 2 \text{ cm}$, spaced $L = 10 \text{ cm}$ apart, and that the beam is coupled in at the center of one mirror (never mind how), how many loops through the interferometer can the beam make? Can we make more loops if we use larger mirrors ($w = 10 \text{ cm}$) with larger spacing ($L = 50 \text{ cm}$)?
9. Estimate the scan angle through which a scanning micromirror would have to move if it were used as a supermarket barcode scanner.
10. For what distance from the optical axis is the difference between a spherical and a paraboloidal mirror greater than $\lambda/10$ for visible wavelengths?
11. Using the paraxial approximation, verify that the magnification of a spherical mirror is indeed given by Equation 6.5.

12. For each of the four entries for concave mirrors in Table 6.1, sketch the object and position, orientation and size of the image.
13. Repeat the previous problem, using a spherical mirror with $R = 2$ cm; place objects at 4 cm, 2 cm, 1.5 cm and 0.5 cm and verify that Equations 6.4 and 6.5 hold.
14. Show that the reflection of an object from a convex spherical mirror generates a virtual image and sketch its position.
15. An actuated, stabilized micromirror is used for beam steering an optical communications beam between two satellites. Assuming zero beam divergence and a photodetector diameter of 10 cm on the receiving satellite, what angular stability is required to assure a stable link if the satellites are separated by 5 000 km?
16. In Figure 6.14, the incident and reflected rays cannot be related as shown; that's why the caption says "Not to scale". Given that the mirror is sketched at an angle of 20° from the vertical, and the incident beam propagates at 20° from the horizontal, sketch the incident and reflected rays *to scale*.
17. A two-dimensional scanning micromirror is used in resonant mode; the x axis has a resonant frequency of 1 kHz and the y axis has a resonant frequency of 2.5 kHz. Assuming the resonance is sinusoidal in both cases, sketch the trace of a reflected light beam as it scans a two-dimensional surface. Is much of the scanned area covered? How might you improve the scan coverage?

7 Refractive micro-optics

*Whoever controls the media, the images, controls the culture.*¹

Allen Ginsberg

When most people think of “optics”, the first component that is likely to spring to mind is the lens. Whereas opticians know that the field is much broader than this, and the heft of this book reminds micro-opticians that their field is equally diverse, lenses are nevertheless essential optical elements in most optical systems. Whereas mirrors are even older, lenses have also experienced considerable historical development and indeed much of the story of optics concerns lenses and their combinations. As creators of images, lenses are used to project the world as we perceive it, from the television camera to the human eye.

We consider in this chapter the domain of what is known as geometric or refractive optics, which relies on the refraction of light, to which we were introduced in Chapter 4.1.3, to achieve a desired optical function. In geometric optics, we can consider light propagating as rays; interference effects are (mostly) ignored, as these form the basis of diffractive optics, the subject of Chapter 8.

The physics of macroscopic lenses is identical to that of their microscopic brethren, so that much of the following discussion is not a function of size. Due to their dimensions, however, microlenses are much more susceptible to diffraction phenomena and are fabricated using very different manufacturing techniques than classical optical elements. Since the optical physics needed to understand refractive lenses applies to both regimes, we will initially present lenses independent of their size, and consider micro-optical aspects and special considerations in the case studies at the end of the chapter.

There is a considerable literature available for the student of optics, including many texts which extensively consider the breadth of the refractive optics field; a small sampler might include the classic (Born and Wolf, 1980, Chapters 3, 4 and 5), the didactically excellent (Hecht, 2002, Chapters 5 and 6), as well as (Iizuka, 1983, Chapters 5 and 6), (Smith, 1990) or (Shannon, 1997). For good applications-oriented coverage of geometric optics, the author proposes (Ray, 2002) for analysis of photographic systems or (Smith and Atchison, 1997) for a discussion of natural and man-made optical compo-

¹ Allen Ginsberg (1926-1997), American poet and beatnik.

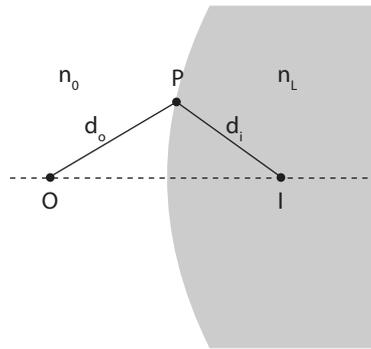


Figure 7.1 Transmission of rays from an object point at O to an image point at I across a curved surface. Fermat's principle tells us at the optical path length, $n_0 d_o + n_L d_i$ must be constant for any transition point P .

nents. The reader is encouraged to consult these for a more in-depth treatment of those topics only insufficiently considered here.

7.1 Lens fundamentals

A lens consists of one or more curved surfaces of a material, optically transparent at the wavelength to be used, whose difference in refractive index with respect to the surrounding media bends impinging light beams in a controlled manner. The shape of the lens surface thus critically defines its optical function: convex lenses, for example, generally lead to a focussing of the incident beam whereas concave lenses are used to generate diverging beams; we define these terms below. We begin our presentation of lenses, then, by analyzing refraction on curved surfaces and derive the mathematical means to design or analyze a refractive optical element.

7.1.1 Refraction at a surface

Since geometric optics is based on refraction, the most important relationship we employ is Snell's law, Equation 4.9, which relates the angles of incidence to those of refraction across a boundary. If we consider a curved surface as that shown in Figure 7.1, we can use Snell's law to determine the change in direction of the rays which are emitted from an object at O in a region of refractive index n_0 [] as they are transmitted into a region with higher refractive index n_L []. We use the notation n_0 to represent the refractive index *outside* the lens, usually air or a liquid, and n_L for the refractive index *inside*, which could be anything from glass to a polymer to another liquid.

If we use a curved surface to focus the emitted rays to a point I , what is the required form of this surface? Helpful in the derivation of the shape is Fermat's principle, to which we were introduced in Section 4.1.1: the optical path length for all rays which converge from a source to a single point must be a constant. Thus, considering Fig-

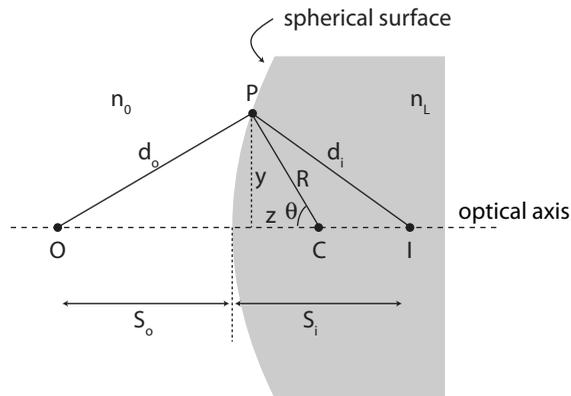


Figure 7.2 Cross-sectional diagram of the optical transmission across a spherical surface of a material with refractive index n_0 ; radius of curvature is R , centered at point C . Rays are emitted from point O and should focus at point I . The optical axis is shown by the dotted line. Not to scale.

ure 7.1, Fermat tells us that

$$n_0 d_o + n_L d_i = \text{constant} \quad (7.1)$$

for any point P along the surface. Using this restriction, we find that as we move the point P while maintaining a constant optical path length, its path describes an ellipse. In general, conic sections (which include ellipses and hyperbolas) represent the ideal shapes for a lens surface. For elliptical or hyperbolic profiles, all rays which are emitted from O (and intersect the interface) will be focussed to the point I ; this focussing function is the essence of a refractive lens.

7.1.2 Refraction at a spherical surface

While an elliptical surface is the optimal one for a lens, this shape is difficult to manufacture, both macro- and micro-optically. Classical lens grinding techniques, as well as a broad range of microfabrication techniques, can easily generate spherical surfaces, such that most lenses are segments of spheres rather than ellipses. Even though the sphere is not the ideal optical shape for a lens, and leads to the numerous aberrations which we will discuss in Section 7.4, it is the shape that the vast majority of refractive lenses employ, such that we will analyze it in detail here.

Consider the cross-section through a spherical lens surface shown in Figure 7.2. Rays emitted from an object point O located in a medium with refractive index n_0 , a distance S_o [m] from the interface, should be refracted to focus at an image point I , in the lens medium with refractive index n_L , a distance S_i [m] from the interface. The lens surface (in this two-dimensional cross-section) is circular, with radius R [m] centered at the point C .

The point P again defines the point on the lens surface at which refraction takes place, except now we restrict the movement of P to be on a circle. For any given P ,

the distance from O is given by d_o [m] and the distance from I by d_i [m]; we will also usefully employ height of P above the optical axis, y [m], which intersects the axis a distance z [m] from the center C .

Path length analysis

Using the geometry of the right triangle with hypotenuse R in the figure, we have

$$z = R \cos \theta \quad (7.2)$$

and

$$y = R \sin \theta; \quad (7.3)$$

for the right triangle with hypotenuse d_o ,

$$d_o^2 = (S_o + R - z)^2 + (R \sin \theta)^2. \quad (7.4)$$

We can re-write this last expression as a function only of the known parameters (S_o , S_i and R), as

$$d_o^2 = (S_o + R - R \cos \theta)^2 + R^2 (1 - \cos^2 \theta) \quad (7.5)$$

$$= [S_o + R(1 - \cos \theta)]^2 + R^2 (1 - \cos^2 \theta) \quad (7.6)$$

$$= S_o^2 + 2S_o R(1 - \cos \theta) + R^2 (1 - \cos^2 \theta) + R^2 (1 - \cos^2 \theta) \quad (7.7)$$

$$= R^2 + (S_o + R)^2 - 2R(S_o + R) \cos \theta. \quad (7.8)$$

Similarly, inside the lens we can analyze the right triangle with hypotenuse d_i to yield

$$d_i^2 = R^2 + (S_i - R)^2 + 2R(S_i - R) \cos \theta. \quad (7.9)$$

Constant path length

The total optical path length L_{opt} for a ray emitted from O , passing through P , to I , is then given by

$$L_{opt} = n_0 d_o + n_L d_i. \quad (7.10)$$

Since we stipulate, by Fermat's principle, that all paths from source to image must have the same optical path length, we take the derivative with respect to the angle θ which defines the point P , or

$$\frac{dL_{opt}}{d\theta} = 0 \quad (7.11)$$

such that

$$\frac{d(n_0 d_o)}{d\theta} + \frac{d(n_L d_i)}{d\theta} = 0. \quad (7.12)$$

Plugging in Equations 7.8 and 7.9 and evaluating, we then obtain

$$n_0 \frac{1}{2d_o} 2R(S_o + R) \sin \theta - n_L \frac{1}{2d_i} 2R(S_i - R) \sin \theta = 0 \quad (7.13)$$

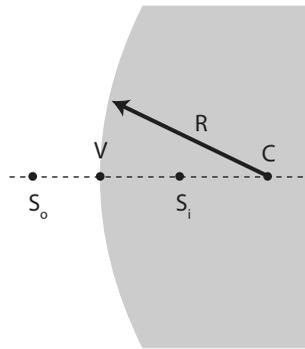


Figure 7.3 Lens parameters for which sign conventions are defined; the sign is a function of their relative position to the lens vertex V .

which yields the relationship

$$\frac{n_0}{d_o} (S_o + R) = \frac{n_L}{d_i} (S_i - R). \quad (7.14)$$

We may re-write this last expression in the useful form

$$\frac{n_0}{d_o} + \frac{n_L}{d_i} = \frac{1}{R} \left(\frac{n_L S_i}{d_i} - \frac{n_0 S_o}{d_o} \right) \quad (7.15)$$

which gives us a relationship between the path lengths, spacings and refractive indices on the object and image sides of the optical interface as well as the radius of curvature.

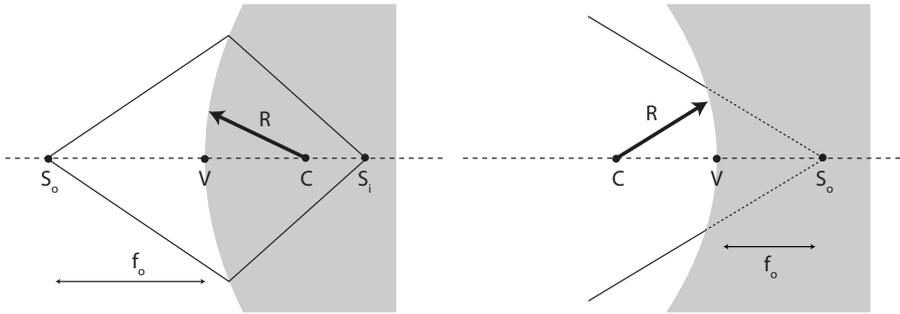
Paraxial approximation

Whereas Equation 7.15 relates the object and image spacings across the interface, it also includes the unknowns d_i and d_o , which we can not determine in general. We can derive a very useful variation of this relationship using the paraxial approximation, which applies if we consider optical rays close to the optical axis. In the paraxial approximation, the angle θ is small, so that $\cos \theta \approx 1$ and $\sin \theta \approx \theta$, and thus $d_o \approx S_o$ and $d_i \approx S_i$.

Using this first order approximation (first order since we restrict ourselves to the first terms of the series expansions of the harmonic functions), we can re-write Equation 7.15 in a simplified form, namely

$$\frac{n_0}{S_o} + \frac{n_L}{S_i} = \frac{1}{R} (n_L - n_0). \quad (7.16)$$

This last expression relates quantities we can determine. If the spacing between lens and object, S_o , as well as the relative refractive indices, are known, we can determine the focus point. We emphasize again that the relationship of Equation 7.16 is based on the paraxial approximation, and is thus valid for rays near the optical axis; this approximation is nevertheless a very useful one.



(a) All lens parameters positive; a convex surface. (b) All lens parameters negative; a concave surface.

Figure 7.4 Sign conventions for convex and concave surfaces.

Sign conventions

To usefully employ Equation 7.16 in realistic systems, it is essential to define the sign conventions of the various parameters. Using the configuration shown in Figure 7.3, and including the focal lengths f_o [m] and f_i [m] which we define below, the generally accepted sign conventions are as follows:

- S_o and f_o are *positive* if found to the *left* of the vertex V ;
- S_i and f_i are *positive* if found to the *right* of the vertex V ;
- R is *positive* if C is found to the *right* of the vertex V .

These definitions, conventionally employed in the optics literature, are important. In the lens equations we shall derive and employ presently, using the proper signs is essential for obtaining the correct results. A favorite error, for example, is to forget that radius can have a negative value.

Example 7.1: Let us plot two examples of configurations for which all the lens parameters are positive and all negative; we recall that the sign is a function of their relative position to the lens vertex V .

In Figure 7.4(a), all parameters are positive. The center of the lens radius of curvature C is to the right of V , the object point O is to the right and the image point I to the left of V . These values are typical for a convex surface, i.e., one for which the vertex is at the maximum extent of the lens material (the lens bends “outward”).

In contrast, Figure 7.4(b) has all parameters negative: the vertex is to the right of C , the “object” is to the left of the vertex (inside the lens) and the image to the left. Such values are typical for a concave surface, one for which the vertex is at the minimum of the lens material (the lens bends “inward”).

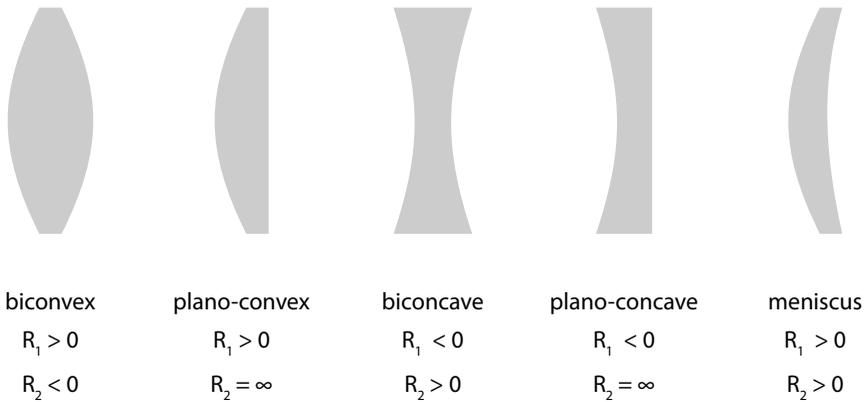


Figure 7.5 Various combinations of convex and concave surfaces lead to different lens types. R_1 and R_2 are the radii of curvature of the left and right interfaces, respectively.

7.1.3 Thin lenses

Up to this point, we have considered a single spherically-curved refractive surface. Most lenses, however, have a second refracting surface from which the optical beams emerge from the lens material, such that the optical function of the lens is the combination of the refractive effects of both of these surfaces. A number of interface combinations are possible, as shown in Figure 7.5: biconvex or biconcave lenses have two convex or concave surfaces, respectively; plano-convex or plano-concave lenses have one planar surface ($R \rightarrow \infty$); and the meniscus lens has one convex and one concave surface.

Thin lens approximation

We can derive a simple relationship between the curvatures of the two lens surfaces and the object and image distances which result. The thin lens approximation allows a simplified analysis, by considering the optical effect of both surfaces yet ignoring the thickness of the lens. Thus we can consider the lens as the sum of refraction at two surfaces, ignoring what happens to the beam as it traverses the bulk of the lens material.

We employ the notation presented in Figure 7.5, where R_1 [m] represents the radius of the left surface and R_2 [m] that of the right, and reemphasize the importance of the sign conventions. Evaluating the function of a complete lens with two interfaces by “adding” the refractive surfaces yields, in the thin lens approximation,

$$\frac{1}{S_o} + \frac{1}{S_i} = \left(\frac{n_L}{n_0} - 1 \right) \left[\frac{1}{R_1} - \frac{1}{R_2} \right] \quad (7.17)$$

where we recall that both object and image are now outside the lens, in a region with refractive index n_0 . This latter expression is known as the Lensmaker’s formula and relates the surface radii of curvature and the refractive indices of the lens and the surrounding medium to the positions of object and image; S_o and S_i are measured from the lens vertex (V , no longer labeled in subsequent figures).

Two special cases are of interest, allowing us to define the focal lengths:

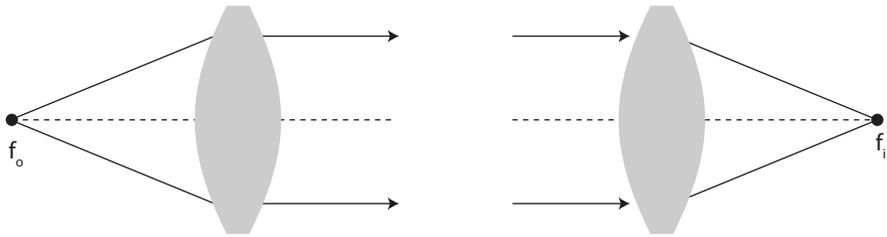


Figure 7.6 Focal points of a convex lens: a point light source at the object focal point f_o results in collimated output *left*; and collimated input is focused at the image focal point f_i *right*

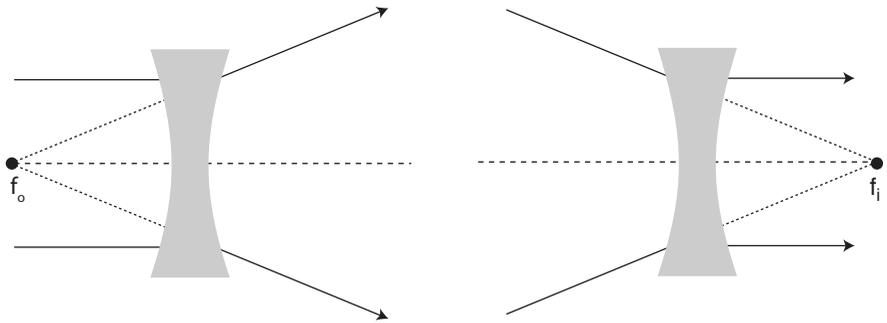


Figure 7.7 Focal points of a concave lens: collimated input diverges and appears to come from a point source at f_o *left*; and light converging to an image focus point f_i emerges collimated *right*.

- $S_o \rightarrow \infty$ yields $S_i \rightarrow f_i$
- $S_i \rightarrow \infty$ yields $S_o \rightarrow f_o$

where f_o [m] and f_i [m] are the focal lengths of the object and image sides of the lens, respectively. Rays from an infinitely distant object ($S_o \rightarrow \infty$) or infinitely distant image ($S_i \rightarrow \infty$) are parallel to the optical axis, and thus collimated. The focal lengths are defined as the points at which collimated input rays meet on the optical axis, as shown in Figure 7.6. We see that thus that collimated rays impinging on a convex lens converge on the focal point or, conversely, emission from a point source at the focus of a convex lens emerges collimated, with the rays parallel to the optical axis².

Concave lenses behave differently, but are also defined by their focal length. As seen in Figure 7.7, collimated input diverges as it traverses a concave lens. The focal point may be determined by tracing the diverging rays backward through the lens, as shown by the dotted lines in the figure; these rays meet at the focal point on the left (object) side of the lens.

The definitions of focus and collimation we have just employed are an idealized simplification. We saw in Section 2.6 that light “rays” propagate as Gaussian beams and the use of lenses is no exception. Thus the focus we consider here is the point at which the

² An alternative definition of focal length, used almost exclusively in optometry, is the diopter, defined as the reciprocal of the focal length in meters.

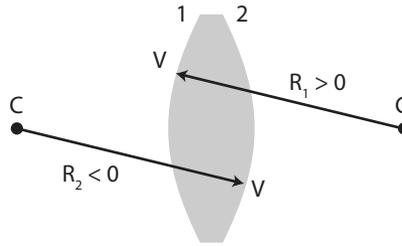


Figure 7.8 Definition of the radii of a biconvex lens.

beam waist is at a minimum and not a mathematical point. The optical effect of a lens on a Gaussian beam is considered in more detail in Section 7.3.3 below.

Gaussian lens formula

For a thin lens, the object and image focal lengths are identical,

$$f_i = f_o \quad (7.18)$$

such that Equation 7.17 may be simplified to the form

$$\frac{1}{f} = \left(\frac{n_L}{n_o} - 1 \right) \left[\frac{1}{R_1} - \frac{1}{R_2} \right]. \quad (7.19)$$

This expression gives a relationship between the radii of curvature of a thin lens and its focal length f , given the refractive indices. Implicit in this equation, and derived from consideration of what happens if object or image move to infinity, is the relationship

$$\frac{1}{S_o} + \frac{1}{S_i} = \frac{1}{f}. \quad (7.20)$$

This last expression is known as the Gaussian lens formula and will prove to be very useful in calculating the imaging properties of thin lenses. Given a lens with focal length f , this simple relationship allows us to determine the image position given any object position.

Example 7.2: Given a bi-convex glass lens ($n_L = 1.5$) in air for which both surfaces have a radius of curvature of 100 mm, where and what is the image of an object at infinity?

Considering the sign conventions for the radius, shown explicitly in Figure 7.8, we see that this lens must have $R_1 = 100$ mm and $R_2 = -100$ mm. Plugging these into Equation 7.19, we obtain

$$\frac{1}{f} = \left[(1.5 - 1) \left(\frac{1}{100} - \frac{1}{-100} \right) \right]$$

such that

$$f = 100 \text{ mm.}$$

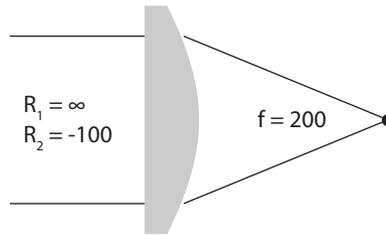


Figure 7.9 Definition of the radii of a plano-convex lens.

Since the object is at infinity, $S_o \rightarrow \infty$ and thus $S_i = f$; thus the rays converge to the focus point at $f = 100$ mm, as in the right side of Figure 7.6.

Example 7.3: Repeat the previous example, but with a plano-convex lens where the right face has a radius of curvature of 100 mm.

We can straightforwardly apply the same relationship, except that now $R_1 \rightarrow \infty$, as seen in Figure 7.9. We thus have

$$\frac{1}{f} = \left[(1.5 - 1) \left(\frac{1}{\infty} - \frac{1}{-100} \right) \right]$$

such that

$$f = 200 \text{ mm.}$$

We thus clearly see, when comparing with the previous example, the role of both surfaces in determining the focal length of a lens.

7.2 Imaging

In micro- and macro-optical systems, we may broadly define two types of functions for refractive lenses: 1) focussing or collimation; and 2) imaging. The former is relatively simple optically: the light from a point source is placed at the focal point of a lens resulting in collimation of the beam, or a collimated beam is focussed onto a surface. Despite the simplicity of the arrangement, collimation is an essential function in, for example, the coupling of light from an LED or laser diode into an optical system; and focussing is required, for example, for bringing optical energy onto a workpiece in laser machining setups. There are many more examples.

The most significant role that macroscopic refractive lenses play in optical systems, however, is for imaging; given an extended object, the function of the lens system is to create a high-quality image of that object. As refractive optical microlenses improve in

quality, imaging will play an increasingly important part in the functionality of micro-optical configurations as well, such that we consider the basics of imaging here.

7.2.1 Real vs. virtual

An optical image can take one of two forms, depending on what is required to view it, as we already saw for the case of reflection from mirrors in Section 6.2.1. A real image is one which can be projected onto a screen whereas a virtual image requires a lens system, such as the eye, to see it. Real and virtual images are also generated by lenses, depending on where the object to be imaged is placed.

Looking ahead to Figure 7.12, we can see the generation of real and virtual images explicitly. If the object position is at $S_o > f$ (Figures 7.12(a) - 7.12(c)), the image is real, as shown by the solid arrow on the right side of the lens. If the object is closer to the lens than the focal length $S_o < f$ (Figure 7.12(d)), the image becomes virtual (the striped arrow) and is on the same side of the lens as the object. This latter situation is the one we encounter when using a magnifying glass to examine insect life: when the ant is placed closer to the lens than the focal length, its image is enlarged and visible using the eye as an imager. Placing the ant *at* the focal point on a sunny day is generally frowned upon.

7.2.2 Ray tracing - by hand

To determine the position, size and orientation of the image of an object placed somewhere near a lens, we employ ray tracing: by following the paths of two or more rays through the lens, using the same approach as we did for spherical mirrors in Section 6.3.2, the corresponding image is the point where these intersect. Ray tracing was traditionally done by hand, with paper, a sharp pencil, and a large eraser, and this approach is of great utility in quickly analyzing simple lens systems, such that we present it here. A more complete analysis, or that of more complex lens systems, is now done with numerical ray tracing, which we discuss in the next section.

Tracing rays

Three rays from any point on an object are particularly simple to follow and for that reason manual ray tracing uses at least two of these, as illustrated in Figure 7.10. Using an arrow as the object³, we can consider the rays which emerge from its tip. The ray which emerges parallel to the optical axis is directed through the image focus f_i after leaving the lens. Likewise, the ray which pass through the object focus f_o emerges from the lens parallel to the optical axis. Finally, it can be shown that the ray which traverses the center of the lens, in the thin lens approximation, is undeviated and is thus a straight line.

³ The use of an arrow is traditional, since the relative size and orientation of the image is immediately clear. Also traditional is placing the object to the left of the lens, such that rays always move from left to right.

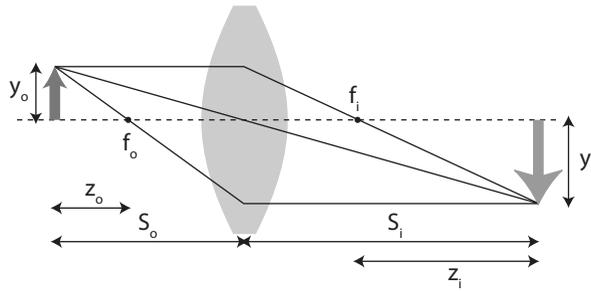


Figure 7.10 The three primary rays used for ray tracing between an object and an image: 1) the ray emitted parallel to the optical axis goes through f_i after passing through the lens; 2) the ray through the lens center is undeviated; and 3) the ray which goes through f_o emerges from the lens parallel to the optical axis.

The intersection of at least two of these three rays uniquely defines the image point. As a result, the image of the object arrow at the left is the image arrow at the right, as seen in the figure. We have only traced the rays from the arrow tip, but any point on the object can be subject to the same treatment, so that even with only this one set, we can place the image as shown: an enlarged, inverted, and real image.

Conjugates

The positions at which light is focussed in an optical system are termed the conjugates. For an arrangement in which rays are traced from an object a finite distance from the lens to form an image, also at a finite distance, we refer to finite conjugates. For the case of collimation, or focussing a collimated beam, in which either the source or image are at infinity, we have an infinite conjugate. Lenses and lens systems are typically designed to be optimal (and have aberrations corrected) for a particular pair of conjugates, and performance degrades if these are used in a different configuration.

Length relationships

The simple geometry of the rays in Figure 7.10 (the various triangles these form) allows us to determine relationships between the various relevant lengths in the system. Given the lens/object and lens/image distances, S_o and S_i , respectively, as well as the object and image heights (y_o and y_i) and finally the spacing between the object and image as the respective focal points ($z_o = S_o - f$ and $z_i = S_i - f$), we can employ basic geometrical relationships to determine that

$$\left| \frac{y_o}{y_i} \right| = \left| \frac{S_o}{S_i} \right| = \frac{f}{S_i - f} \quad (7.21)$$

and

$$\left| \frac{y_i}{y_o} \right| = \frac{f}{S_o - f}. \quad (7.22)$$

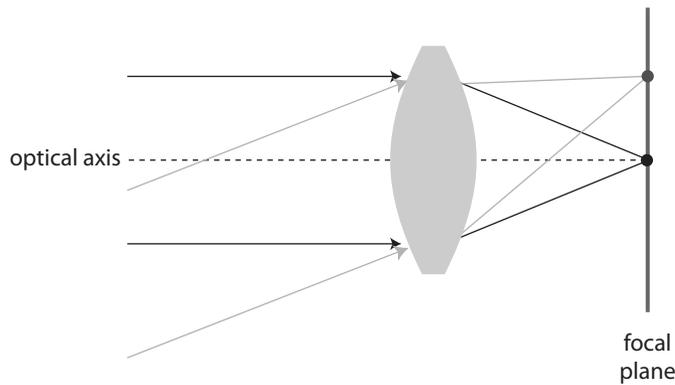


Figure 7.11 The focal plane is the plane in which the focal points for on- and off-axis rays are found; it is only planar in the paraxial approximation.

As a result, we see that

$$(S_i - f)(S_o - f) = f^2 \quad (7.23)$$

$$z_o z_i = f^2, \quad (7.24)$$

which is a useful relationship between the distances of an object and an image from the focal points and the focal length, comparable in utility to the Gaussian lens formula.

Focal plane

As shown in Figure 7.10, the image of a distributed (as opposed to a point) object is of course also distributed. Each point of an image is the focal point of the rays emerging from the corresponding part of the object, and these focal points of the image are distributed in two dimensions on the focal plane. Figure 7.11 shows this explicitly: the second dimension of the plane is normal to the page.

In general, the focal surface is spherical; a planar focal surface is an approximation only valid in the paraxial case, for rays close to the optical axis. We will continue to assume a focal plane here, but will consider aberrations which result from field curvature (the difference between spherical and planar surfaces) in Section 7.4.

Image generation

Using this simple approach for ray tracing and employing the considerations of the previous paragraphs, we can see what happens to the image of an object as we place this at different positions along the optical axis. In Figure 7.12, we trace two rays (the one which emerges parallel to the optical axis and the one which traverses the lens center) for four cases.

Placing the object far from the lens, $S_o > 2f_o$, as in Figure 7.12(a), we find that the image appears between f_i and $2f_i$, is reduced in size, real and inverted. In Figure 7.12(b), the object and image are both at $2f$, and the image is the same size, real, and inverted; this latter case is known as 1:1 imaging, since, apart from the inversion, the image is the same size as the object.

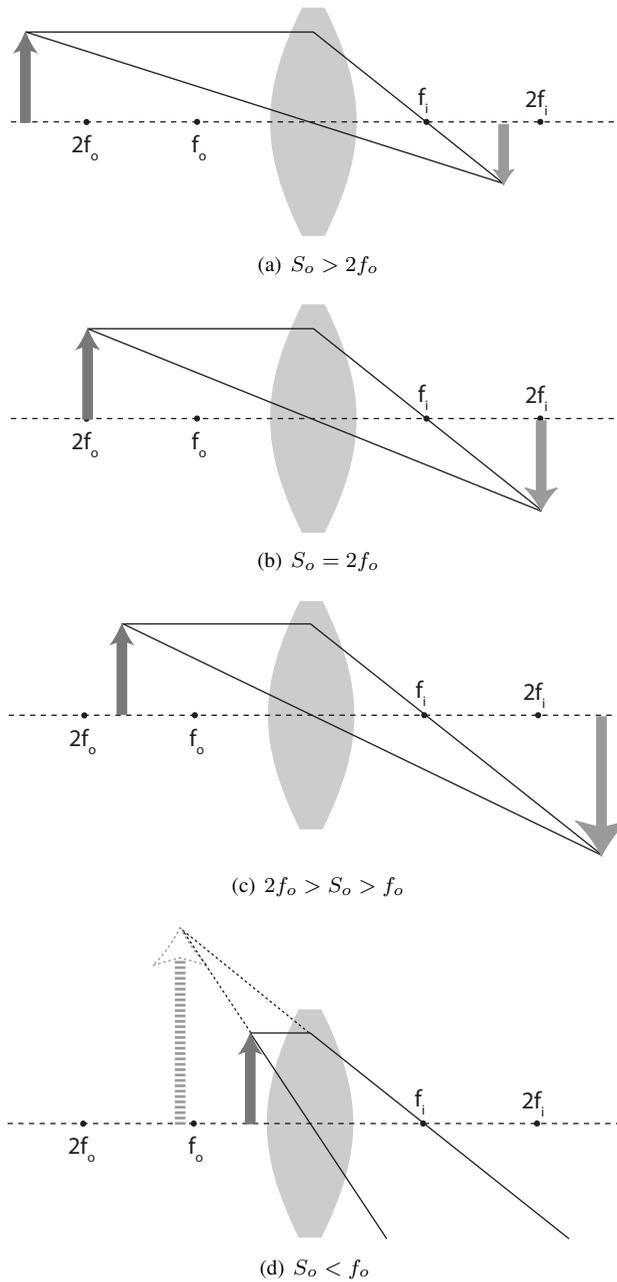


Figure 7.12 Image position and size for an object at four positions along the optical axis on the left side of a biconvex lens.

Table 7.1. Imaging using convex and concave lenses: image type, location, orientation and magnification for real objects located at various positions with respect to the focus.

convex lenses				
object location	image type	image location	orientation	magnification
$\infty > S_o > 2f$	real	$f < S_i < 2f$	inverted	$ M_T < 1$
$S_o = 2f$	real	$S_i = 2f$	inverted	$ M_T = 1$
$2f > S_o > f$	real	$2f < S_i < \infty$	inverted	$ M_T > 1$
$S_o = f$		$\pm\infty$		
$S_o < f$	virtual	$ S_i > S_o$	erect	$ M_T > 1$
concave lenses				
anywhere	virtual		erect	$ M_T < 1$

Moving closer to the lens, for $2f_o < S_o < f_o$, Figure 7.12(c), the image is enlarged, still inverted and real, and found farther from the lens, $S_i > 2f_i$; as we move closer and closer to the lens, the image increases in size and moves farther away. For the special case, not shown in the figures, that the object is placed *at* the focal point f_o , the image is at infinity.

Finally, for the object placed closer to the lens than the focal distance, $S_o < f_o$, the image becomes virtual (represented by the striped, as opposed to solid, arrow, in Figure 7.12(d)), enlarged and erect (non-inverted). As we move the object closer to the lens for this case, the virtual image increases further in size and moves closer to the lens. Table 7.1 summarizes these four cases.

Magnification

In the previous examples, we saw that the image size changes as a function of object position. The transverse magnification, M_T [], of a lens quantifies the size and orientation of the image, and is given by the ratio of image size to object size, namely

$$M_T = \frac{y_i}{y_o} \quad (7.25)$$

for size y (extent in the y direction, normal to the optical axis), as shown for the object height y_o and image height y_i in Figure 7.10. Based on our ray tracing results, we can determine the magnification from the ratio of image to object positions, namely

$$M_T = -\frac{S_i}{S_o}. \quad (7.26)$$

Using the results of the relationships in Equations 7.21 and 7.22, we can combine these with Equation 7.25 to write the magnification in an additional useful form, namely

$$M_T = -\frac{z_i}{f} = -\frac{f}{z_o} \quad (7.27)$$

for spacing between image or object and focus point, z_i or z_o , respectively. These relationships allow a quick determination of the magnification of an image based on the position of object or image with respect to the focus.

The magnitude of M_T gives the ratio of the sizes of image to object; $M_T > 1$ implies magnification and $M_T < 1$ implies de-magnification of the image. The sign of M_T indicates the orientation: $M_T > 0$ implies an erect image whereas $M_T < 0$ an inverted image. It is important not to confuse these two aspects of magnification.

Since, for real images, S_o and S_i are always greater than zero, Equation 7.26 always gives $|M_T| < 0$. This consideration implies that a single lens *always* generates real images which are inverted. A summary of the imaging properties of convex and concave lenses for objects at various positions along the optical axis is given in Table 7.1.

7.2.3 Ray tracing - numerically

Whereas a rudimentary analysis of the imaging properties of a single lens or simple lens combination can be done by hand, a detailed prediction of the optical performance is usually done using numerical ray tracing techniques. Whereas ray tracing software does nothing but apply Snell's law at a succession of interfaces, the ability to rapidly include material information, dispersion, wavelength dependencies, multiple interfaces, off-axis and skew rays and so forth, implies that numerical ray tracing can allow both fast and comprehensive analysis of a lens system. These computer-based optical modeling techniques have been extensively developed and many of the numerical approaches are also of particular value for micro-optical components (Brenner and Singer, 1993; Zeitner et al., 2004).

Basis

A numerical ray-trace proceeds as follows:

- a ray at a certain origin height h is traced to a given incident height y on the lens, thereby defining its angle with respect to the optical axis⁴;
- Snell's law is used to calculate the transmission angle across the interface, given the refractive indices and the wavelength;
- the ray is transmitted to the next interface, defining a new h and y ;
- the process is repeated until the ray emerges from the optical system;
- all of the previous is repeated using rays at various origins and angles, and often at varying wavelengths.

This procedure was done in the past using graph paper, a ruler, a compass, a slide rule and some logarithm tables, which limited practical calculations to a few rays in the tangential plane (that including the primary ray and the optical axis; look ahead to Figure 7.27 in Section 7.4.4) in two dimensions. Computing power now allows a

⁴ The parameters y and h are clearly distinguished if we look ahead to Figure 7.24, for the collimated case where $y = h$, and Figure 7.26, for the general case where $y \neq h$.

completely numerical approach, such that a wide variety of skew rays (those not in the tangential plane) may be rapidly calculated in three dimensions.

Optical path difference

A typical ray trace is performed from a single object point to a single image point, over several paths through the system. One of the most relevant parameters in such a trace is the optical path difference (OPD), defined as the difference between the actual and ideal optical path lengths (OPL). We recall that an ideal point source generates spherical waves; inversely, an ideal point image is formed by spherical waves converging to a point at the center of the sphere. As a result, the OPD is a measure for the difference between the actual path length and the path given by a reference sphere centered at the image point. The OPD is typically calculated and plotted as a function of ray incidence height on the lens, y , and the magnitude of the variation is a measure of lens quality; either the peak-to-peak or root mean square (rms) value of OPD is taken to characterize the system.

If the absolute value of OPD varies less than $\lambda/4$, we generally speak of the Rayleigh limit, or a diffraction-limited lens. A diffraction-limited lens is one in which solely diffraction defines the nature of the generated pattern; as we will see in Section 8.2.4, a “perfectly” focussed spot generates the pattern of an Airy disk, which we already saw in Section 6.4.3 for micromirrors. If the Airy disk generated by a lens has 84% of the total energy in the central peak, then we speak of a diffraction-limited lens.

Spot diagram

If the OPD is larger than several wavelengths, then diffraction may generally be ignored and the deviations of the system from an ideal lens may be determined solely by geometric traces. The most useful means of characterization is then the spot diagram. The input aperture is divided into equal areas, usually squares, and a ray is traced through the center of each square, forming a spot in the image plane. The density of spots is then energy density in the image plan.

The distribution of spots is sometimes used to define a point spread function (PSF), a two dimensional energy distribution function based on the density of traced spots; strictly speaking, however, the PSF is the energy distribution in a spot due solely to diffraction from the aperture, which we consider in the next chapter. A one dimensional cross-section through the PSF yields a line spread function (LSF), a one-dimensional cut through the plane of spots. The PSF and/or LSF provide detailed information about the energy distribution in the image plane.

Strehl ratio

The quality of the focussed spot, and by extension the optical system, is often characterized by the Strehl ratio⁵, defined as the ratio between the calculated peak of the Airy

⁵ Karl Strehl (1864–1940), a Bavarian high school school teacher in the town of Hof, but also a noted mathematician, physicist and astronomer, defined criteria for defining the sharpness of optical systems (primarily telescopes) and was one of the first to address the issue of atmospheric distortion in astronomy.

pattern and that expected for a “perfect” diffraction-limited lens. The Strehl ratio may be approximately related to the OPD by

$$\text{Strehl ratio} = (1 - 2\pi^2 \text{OPD}^2)^2 \quad (7.28)$$

where the OPD term represents the rms value measured across the full lens aperture. A value for Strehl ratio close to unity thus implies a lens close to the diffraction limit.

Modulation transfer function

The resolution of an imaging lens or lens system is defined by the dimension of the smallest feature which can be distinguished. As we saw for micro-mirrors in Chapter 6, two closely spaced spots are considered to be distinguishable if the maximum of the Airy disk of one is further from the other than the first minimum of its Airy disk. The resolution is typically characterized by the modulation transfer function (MTF).

MTF is determined by imaging a pattern of light and dark lines with variable spatial frequency, or period, Λ [m]. The image of this pattern is a convolution of the LSF with the object lines, and the contrast of this image, C [], is found from

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \quad (7.29)$$

where I_{max} [W/m²] and I_{min} [W/m²] represent the maximum and minimum intensities of the image, respectively. The contrast is often also referred to as the “modulation”, hence the designation MTF.

As period Λ decreases, the contrast of the image C also decreases. An MTF characteristic is then a plot of C versus Λ , the latter usually in units of lines/mm; a rapidly decreasing contrast with decreasing period, or a small absolute value for C for a given Λ , implies limited optical resolution. A characteristic MTF plot derived from simulation is shown in the following example and measured MTF curves may be found in Figure 12.10(c) of Section 12.2.

Example 7.4:

A variety of commercial ray tracing programs exists, ranging widely in capability, complexity and cost, and the corresponding literature is extensive: see for example (Kidger, 2002, 2004) or the series (Mouroulis et al., 2008). To illustrate a few of the types of results obtained from numerical optical simulation, we will employ the ZEMAX package here to simulate a simple example optical system. This software enjoys considerable popularity and has been used for design and analysis of microlenses (Stockham and Smith, 2006) as well as DMD-based micromirror systems (Dewald, 2000).

We take the microlens system shown in Figure 7.13; as we will see in Section 7.6.1 below, microlenses tend to be thin, curved structures on a transparent substrate. In the figure, the photoresist microlens (with $n_L = 1.56$) is patterned on the left surface of the 1 mm thick Pyrex substrate and has a radius of curvature of 690 μm and height of 64 μm .

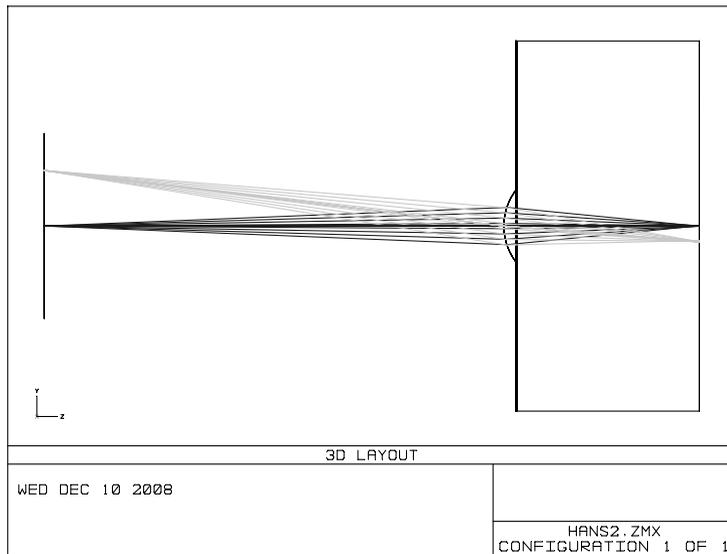


Figure 7.13 An example microlens system for numerical ray tracing using ZEMAX. The photoresist microlens (with $n_L = 1.56$) is patterned on the left surface of the 1 mm thick rectangular Pyrex substrate. The lens has a radius of curvature of 690 μm and height of 64 μm . Two ray bundles are simulated, using $\lambda = 760$ nm.

Two bundles of rays (at $\lambda = 760$ nm) are traced by the program in this example: the rays originate from point sources on the left and, traversing the lens and the substrate, are focussed on the right, exactly on the back side of the Pyrex substrate. A central bundle of on-axis rays and a bundle of oblique rays have been traced through the system. In practice, numerous further ray bundles can easily be traced to reflect the possible source positions. The quality of the focus can be visualized by examining the spot diagram of the on-axes rays, as shown in Figure 7.14. We see that the focus spot is circularly symmetric and takes the form of the Airy disk, as we expect, with a diameter of about 3 μm .

Numerical simulation can also directly yield characteristics such as the modulation transfer function. Figure 7.15 shows the diffraction-limited MTF and the actual MTF of this microlens, plotted as a function of spatial frequency in lines per millimeter, for an illumination wavelength of 780 nm. We see that the microlens is very close to the diffraction limit, implying that the primary aberrations play a relatively small role in determining lens performance in this example.

In Example 7.11 below, we will again use ZEMAX to analyze the aberrations in this lens by looking at the form of the focus spots generated by both ray bundles.

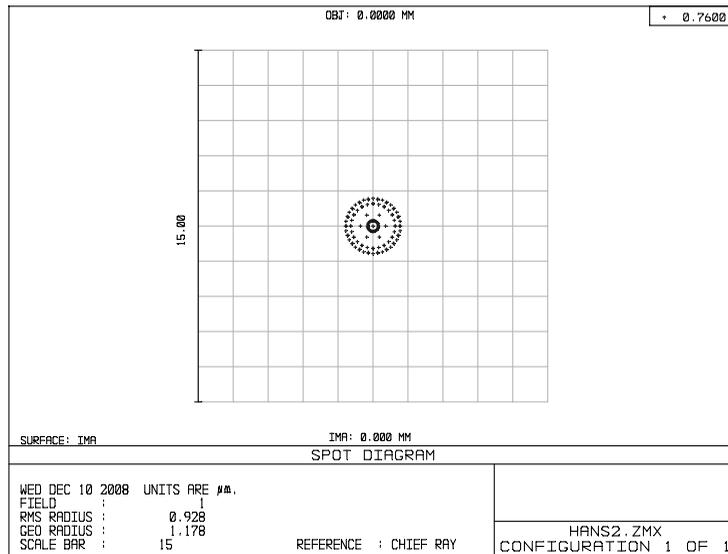


Figure 7.14 The spot diagram at the focal point of the on-axis rays for the microlens of Figure 7.13. Grid spacing is 15 μm .

7.2.4 Aperture

Lenses are not infinitely wide but have a finite diameter: the transparent region of a lens is defined as the aperture. As seen in Figure 7.16, where we have drawn the lens mount explicitly, a lens has a maximum aperture diameter D [], which then limits the amount of light that can enter the component. Two parameters, numerical aperture and f-number, derive from the finite aperture and represent important lens and optical system characteristics.

Numerical aperture

The numerical aperture, NA [], is a lens or lens system parameter which can be interpreted as the light-gathering power of the component. Consider the dual-lens arrangement in Figure 7.17: for emission from a point source, there is a maximum angle at which light may still be coupled into the lens or, as in this case, the lens system. For rays incident at $\theta > \theta_{max}$, either the ray will strike outside the diameter of the lens, or the ray successfully refracted by the first biconvex lens will miss the second, the plano-concave, lens.

Based on this maximum angle of incidence, the numerical aperture is defined as

$$NA = n_0 \sin \theta_{max} \quad (7.30)$$

for refractive index n_0 outside the lens. For a point source at the focus of the lens, the maximum angle is given by

$$\theta_{max} = \arctan \left(\frac{D}{2f} \right) \quad (7.31)$$

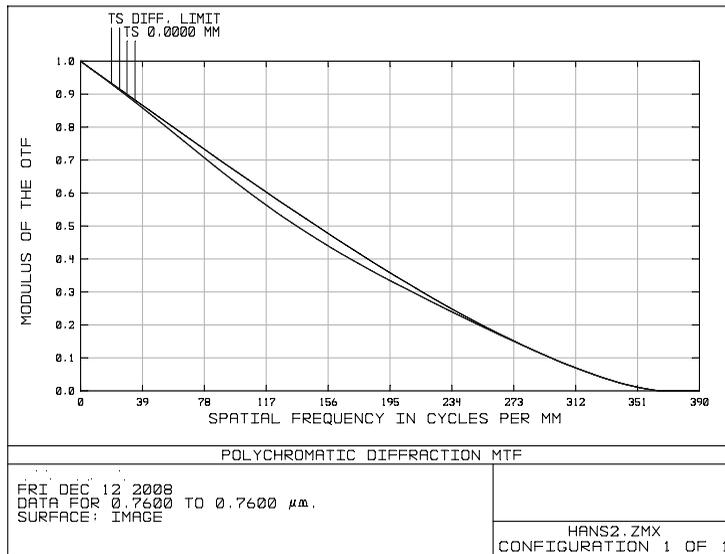


Figure 7.15 The calculated MTF for the microlens in Figure 7.4. Two curves, closely overlapping, are shown, both calculated for $\lambda = 760 \text{ nm}$; the slightly higher one is the diffraction-limited MTF for this lens, and the lower is the actual MTF. The abscissa gives the spatial frequency (in lines per millimeter) and the ordinate the value for contrast.

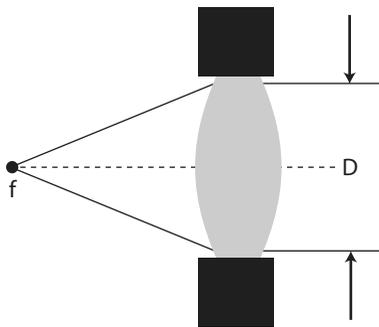


Figure 7.16 Every lens has a finite aperture D , typically defined by the lens mount as shown by the black frame. The aperture limits the amount of light that it can collect and leads to the definition of NA and $f/\#$.

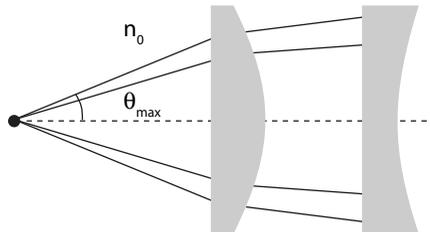


Figure 7.17 Definition of the numerical aperture. θ_{max} is the maximum incidence angle for a light ray to enter the lens.

for aperture diameter D , such that we can approximate

$$NA = n_0 \sin \left[\arctan \left(\frac{D}{2f} \right) \right] \approx n_0 \frac{D}{2f} \quad (7.32)$$

for lenses with large aperture, $D \gg f$. A large aperture thus implies a high value for numerical aperture.

For operation in air, with $n_0 = 1$, we see thus that numerical aperture is limited to $NA < 1$; typical values are $0.7 \leq NA \leq 0.95$ for high-quality refractive optics. To overcome this limitation, the lens can be operated in an immersion liquid with $n_0 > 1$; from Equation 7.30, we see then that $NA > 1$ may be achieved. The use of immersion was already proposed by Robert Hooke in the 17th century and thus has a long history; it is used today in high-end microscopy (Mansfield and Kino, 1990; Koyama et al., 1999; Ippolito et al., 2001; de Lange et al., 2001) and is key to achieving the extreme resolution of state-of-the-art lithography systems (Switkes et al., 2003; Mulkens et al., 2004; Owa and Nagasaka, 2004). Elastomer-based rubber micro-ball-lenses with $NA \geq 1.25$ have been developed for compact, high-resolution microscopes using the principle of solid immersion (Gambin et al., 2006).

A large value for NA is typically desired. High numerical aperture implies a high light-gathering power, such that more light can be coupled into the lens. More significantly for many applications, high NA leads to high optical resolution, meaning the minimum size of the spot to which a collimated beam can be focussed or the minimum spacing between two structures resolvable in a magnifying optical system, as we saw in our definition of MTF. As we will see in Chapter 8 below, diffraction effects limit the resolution of finite-aperture systems, and the width of a focus spot is given by

$$d = 1.22 \frac{\lambda}{NA} \quad (7.33)$$

for operating wavelength λ . Thus a high NA implies a small spot size or conversely a high resolution for the lens or lens system. It is for this reason that high-end optics often uses lenses with large diameters.

f-number

Related to the numerical aperture as a measure of the light-gathering power of a lens is the so-called “f-number”, written as $f/\#$ []. Again considering the lens aperture defined in Figure 7.16, this parameter is found simply from

$$f/\# = \frac{f}{D} \quad (7.34)$$

for focal length f . As is easily seen from this last expression, $f/\#$ decreases as aperture D increases, such that a small f-number implies higher light-gathering power for a given focal length.

$f/\#$ is most popularly employed for photographic optics, where the “speed” of an objective lens is given by this parameter. High-speed lenses are those which allow short exposures since they have higher light-gathering power, and thus a small $f/\#$ due to the large aperture. The multi-kilogram (and multi-kilo-€) telephoto lenses favored by

sports photographers and paparazzi are so large because $f/\#$ scales with focal length: a 300 mm $f/2$ lens requires an aperture of at least 150 mm.

Example 7.5: A 50 mm standard camera objective is specified as $f/1.4$. This value for $f/\#$ is reached when the aperture ring, the metal diaphragm which opens and closes the aperture, is fully open and is thus the highest “speed” for that lens. What is the aperture and why do the aperture stops on the ring make steps of 1.4, 2, 2.8, 4, 5.6, . . . ?

The aperture of the lens is easy enough to calculate from Equation 7.34 as

$$D = \frac{f}{f/\#} = \frac{50 \text{ mm}}{1.4} \approx 36 \text{ mm}.$$

The seemingly odd sequence of apertures is due to the fact that each is a factor $\sqrt{2}$ larger than the previous. Increase in $f/\#$ by $\sqrt{2}$ implies decrease in D by $\sqrt{2}$, which implies a decrease in the illuminated *area* of the lens by a factor 2. Since, for a given input optical power density, the open area of the lens defines the power transmitted through the optical system, each step of the aperture ring thus reduces the power (on the CCD or film) by a factor of two.

7.3 Advanced lenses

We have discussed the optics of refractive lenses using a number of approximations, which allow calculation of numerous configurations by hand. We consider in this section some additional aspects of lenses important for characterization of the individual components or systems. In some cases, considerably more calculation is required for a comprehensive analysis than presented above, but the student should be aware of the concepts in any case.

7.3.1 Thick lenses

The simple calculations and imaging constructions we considered above used the thin lens approximation: a “thin” lens is one whose thickness is considerably smaller than its diameter. For a thin lens, we traced up to three rays through the lens but ignored what happens in the bulk of the lens. Whereas this approximation works well for many simple systems, many refractive lenses are far from thin and we need to consider more exactly what happens as the ray traverses the glass.

Figure 7.18 shows a more detailed depiction of the ray moving through the material of a lens of thickness t . Refraction at both interfaces is clearly seen, defining the ray direction in the material. If we trace the incoming and emerging rays to the point where they intersect inside the lens, this intersection defines one of the principle planes P_2 . For a ray moving toward the right, the principle plane defines the position at which we

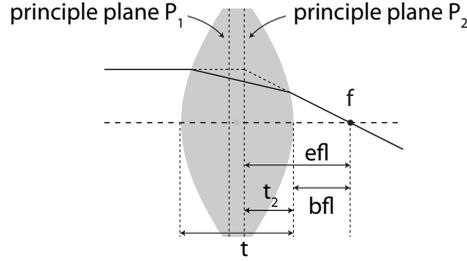


Figure 7.18 Traversal of a ray through a thick lens of thickness t . The effective focal length efl is the distance between the principle plane P_2 and the focus; the back focal length bfl is the spacing between the lens surface and the focus.

would place an equivalent (infinitely) thin lens with the same refractive power; a second principle plane P_1 can be defined toward the left side of the lens for rays moving to the left.

Focal length

For a thick lens of the form shown in Figure 7.18, with radii of curvature R_1 and R_2 for the left and right interfaces, respectively, thickness t as well as lens and surrounding media refractive indices n_L and n_0 , respectively, geometric ray tracing allows us to calculate the focal length f as

$$\frac{1}{f} = \left(\frac{n_L}{n_0} - 1 \right) \left[\frac{1}{R_1} - \frac{1}{R_2} + \left(\frac{n_L - n_0}{n_L} \right) \frac{t}{R_1 R_2} \right]. \quad (7.35)$$

This expression for focal length is, for the case of a thick lens, known as the effective focal length (efl), as it is measured from the position of the principle plane P_2 .

In contrast, we can also usefully define the back focal length (bfl) as the spacing between the surface of the lens and the focal point as

$$bfl = f - \left(\frac{n_L - n_0}{n_L} \right) \frac{ft}{R_2} \quad (7.36)$$

where the difference between effective and back focal lengths is given by the spacing between the principle planes and the lens surface, given for the left and right surfaces by

$$t_1 = -f \left(\frac{n_L - n_0}{n_L} \right) \frac{t}{R_2} \quad (7.37)$$

and

$$t_2 = -f \left(\frac{n_L - n_0}{n_L} \right) \frac{t}{R_1}. \quad (7.38)$$

We can verify from the previous expressions that for a thin lens, for which $t \rightarrow 0$, we then have $(t_1, t_2) \rightarrow 0$ and the expression for focal length approaches that for the thin lens approximation, Equation 7.19, such that both $(efl, bfl) \rightarrow f$. For thick lenses with short focal lengths, however, the difference between efl and bfl can be significant.

Imaging

Imaging with thick lenses may employ the same simple relationships between object and image position which we employed for thin lenses. We recall that the Gaussian lens formula, Equation 7.20, related S_i and S_o with f , and this relationship also holds for thick lenses if we use the principle planes to define the image and object positions. Accurate determination of P_1 and P_2 is thus an essential first step in the calculation of the imaging properties of a realistic thick lens system.

In practice, the image through a thick lens is traced in two parts. First the image is calculated behind the first surface, assuming the lens is infinitely thick, such that the image is inside the lens. This image then becomes the “object” for the second step, which calculates the image emerging from the lens through the second surface. We thus assemble the final image based on the effect of both surfaces independently.

Example 7.6: Consider a PMMA microlens with an aperture of 500 μm fabricated on a bulk substrate. What are the differences between effective focal length and back focal length for radii of curvature 4 mm and 400 μm ?

Such a micro lens, of which we will see numerous versions in Section 7.6, is plano-convex; the radius of curvature is thus R_2 , with $R_1 \rightarrow \infty$. Determining the refractive index of PMMA to be $n_L = 1.491$ (Table 3.5), we can then apply the Lensmaker’s formula, Equation 7.19, to determine the efl for both cases,

$$\frac{1}{f} = \left(\frac{n_L}{n_o} - 1 \right) \left[\frac{1}{R_1} \right]$$

which yields 8.15 mm for $R_2 = 4$ mm and 815 μm for $R_2 = 400$ μm .

Determination of the bfl requires that we know the thickness of the lens. Some geometry allows us to relate the lens thickness t to the radius of curvature R_2 and its aperture D as

$$t = 2 \left(R_2 - \sqrt{R_2^2 - \left(\frac{D}{2} \right)^2} \right)$$

giving 15 μm for $R_2 = 4$ mm and 176 μm for $R_2 = 400$ μm . The microlens with the smaller radius is thus much thicker than its partner, as we might expect. Comparing the bfl values, we can plug t into Equation 7.36 to yield a bfl of 8.14 mm for $R_2 = 4$ mm and 697 μm for $R_2 = 400$ μm .

If we consider the percentage difference between efl and bfl for the two cases, we find about 0.13% for the large radius, but 14% for $R_2 = 400$ μm . Thus the thick lens, which has dimensions typical for microlenses, has considerably different values for effective- and back-focal lengths, and these need to be considered in the design of an optical system.

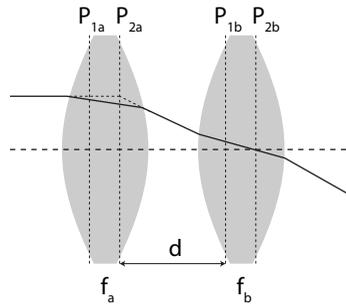


Figure 7.19 A system of two convex lenses with focal lengths f_a and f_b , whose principle planes are separated by a spacing d .

7.3.2 Multiple-lens systems

Most refractive optical systems which perform more than a rudimentary function employ multiple lenses; as we will also see in Section 7.4 below, assemblies of lenses are generally required to reduce the aberrations of an optical system. Design of multi-lens systems rapidly becomes very complex, and is thus the bread and butter of optical design teams. We just take a glance at what happens if we combine two convex lenses here and refer the interested reader to the excellent literature (Smith, 1990).

Consider the pair of convex lenses shown in Figure 7.19; the spacing between the lenses is given by the separation of the principle planes P_{2a} and P_{1b} , as indicated. For focal lengths f_a and f_b of the left and right lenses, respectively, it can be shown that the focal length of the two-lens combination, f_{ab} is given by

$$f_{ab} = \frac{f_a f_b}{f_a + f_b - d} = \left[\frac{1}{f_a} + \frac{1}{f_b} - \frac{d}{f_a f_b} \right]^{-1}. \quad (7.39)$$

For the thin lens approximation, the principle planes coalesce and the spacing d becomes the separation between the centers of the lenses.

Just as a single focal length may be defined for this lens pair, a single pair of principle planes may also be defined for the lens system. The spacing between the lens centers (O_A and O_B , for the left and right lenses, respectively) and the left and right principle planes of the dual-lens system, P_1 and P_2 , respectively, are given by (Smith, 1990, Chapter 2)

$$\overline{O_A P_1} = \frac{f_{ab} d}{f_b} \quad (7.40)$$

$$\overline{O_B P_2} = -\frac{f_{ab} d}{f_a} \quad (7.41)$$

where positive values imply a position to the right of the lens element.

Recalling that transverse magnification is given by $|M_T| = S_i/S_o$ (Equation 7.26), and that object and image spacing are related by the Gaussian lens formula, $1/f = 1/S_o + 1/S_i$ (Equation 7.20), we see that moving the lenses with respect to each other, thus changing d , gives rise to a change in f_{ab} and thus a change in M_T . The pair of

lenses thus acts as a primitive zoom; although zooms are typically configured using three or more lenses (usually a movable concave lens positioned between two convex lenses), this simple example gives a hint of the functionality available when using lens combinations.

Combining the powers of lenses, both positive and negative for convex and concave, respectively, can be done in this manner for an arbitrary number of elements; matrix methods (Hecht, 2002, Chapter 6.2.1) or numerical ray tracing (Section 7.2.3) are most often used for an efficient analysis of multi-lens systems. Although most micro-optical components are designed for use as single elements, advances in fabrication and assembly technology have led to both tunable micro-optical lenses (considered in Chapter 12) as well as reconfigurable lens systems of the type considered here.

Example 7.7: Consider the two dual-lens systems shown in Figure 7.20; both of these function as telescopes, using an objective lens with f_1 to the left and an eyepiece with f_2 on the right. The Keplerian⁶ configuration uses two convex lenses spaced $f_1 + f_2$ apart, whereas the Galilean⁷ setup uses a concave eyepiece, with f_2 positioned at f_1 as shown. Such lens arrangements are termed afocal, meaning that the system does not form an image, i.e. both object and image are at infinity.

How does the telescope work? By considering the trace of the rays through the focal point and the fact that NA is the same for both lenses, we can relate the diameters of the collimated beams, as indicated in the figures, as

$$\frac{d_1}{d_2} = \frac{f_1}{f_2}.$$

Thus a collimated beam is reduced in diameter (and we will take advantage of this feature when we design a beam expander in Example 7.9), but, then, why does a telescope magnify? It is not the width of the beam which is relevant, but the angle with which a ray emerges from the eyepiece. Consider the ray incident at θ_1 in the upper figure, which crosses the intersection of the focal points, the image plane for both lenses, at a height y .

The simple geometry of the system allows us to write

$$\frac{\theta_2}{\theta_1} = \frac{y/f_2}{y/f_1} = \frac{f_1}{f_2} = M$$

⁶ Johannes Kepler (1571-1630), German astronomer and mathematician, is best remembered for his three laws of planetary motion, which did much to cement the heliocentric model of the solar system into accepted knowledge; he was also the first to correctly explain the optics of the human eye. His version of the telescope with a convex eyepiece was an improvement over Galileo's version and is the standard configuration used today.

⁷ Galileo Galilei (1564-1642), Italian astronomer, fabricated his own telescope upon hearing of its invention in 1609 and promptly discovered four of Jupiter's moons as well as the mountains on earth's moon. As is well known, these discoveries, coupled with his support for the Copernican view of the universe, did little to ingratiate himself with the authorities of his time.

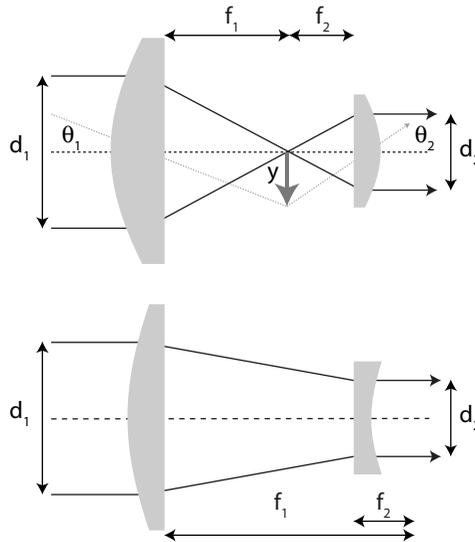


Figure 7.20 Two 2-lens systems which form the the basis for a telescope: the Keplerian *top* uses two convex lenses whereas the Galilean *bottom* employs a convex objective and a concave eyepiece.

for *angular* magnification M . Thus for an objective lens with a longer focal length f_1 than that of the eyepiece f_2 , $M > 1$ and the angular separation between rays in the generated virtual image increases, the essential function of this two-lens system.

7.3.3 Gaussian optics

We saw in Section 2.6 that, considered exactly, light does not propagate as rays but as Gaussian beams. Whereas the ray model is extremely useful for a wide variety of lens analyses, some situations require that we consider the Gaussian shape of the beam; these considerations are particularly relevant when looking at the details of focussing or collimation.

Transmission through lenses

A nice property of the Gaussian beam is that when it is transmitted through a circularly symmetric aperture, such as (typically) a lens, the beam remains Gaussian; the beam waist width, $2W_0$, the depth of focus, $2z_0$ and the radius of curvature, $R(z)$, may change, but the functional form remains Gaussian. A glance at Table 2.5 is useful to remind ourselves of the relevant parameters defining a Gaussian beam.

Consider the transmission of a Gaussian beam through a convex lens as shown in Figure 7.21: the beam moves from left to right and at the input (at the left) has a radius of curvature R_L and a waist width W_{0L} at position z_L . The radius of the transmitted beam (at the right), R_R , is then related to the input radius and focal length f of the lens

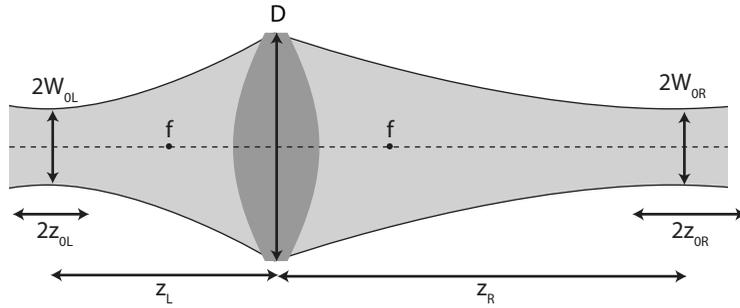


Figure 7.21 The change in dimensions of a Gaussian beam when transmitted through a lens with focal length f : the beam waist width of $2W_{0L}$, with depth of focus $2z_{0L}$, located at Z_L , is transformed into a beam with waist width of $2W_{0R}$, with depth of focus $2z_{0R}$, located at Z_R .

as

$$\frac{1}{R_L} - \frac{1}{R_R} = \frac{1}{f}. \quad (7.42)$$

We see that $R_L > 0$ since the input beam is diverging, whereas $R_R < 0$ since the output beam is converging.

Based on this simple relationship, the beam waist width and position can be determined at the output, and are given by

$$2W_{0R} = \frac{2W_{0L}}{\sqrt{1 + \left(\frac{\pi W_{0L}^2}{\lambda R_R}\right)^2}} \quad (7.43)$$

and

$$z_R = \frac{R_R}{1 + \left(\frac{\lambda R_R}{\pi W_{0L}^2}\right)^2} \quad (7.44)$$

for illumination wavelength λ .

These expressions can be simplified somewhat by defining the magnification, M , as

$$M = \left| \frac{f}{z_L - f} \right| \left[1 + \left(\frac{z_{0L}}{z_L - f} \right)^2 \right]^{-1/2} \quad (7.45)$$

such that we can easily relate the beam widths, positions and depths of focus on the left and right sides of the lens as

$$2W_{0R} = 2MW_{0L} \quad (7.46)$$

$$2z_{0R} = 2M^2 z_{0L} \quad (7.47)$$

$$z_R - f = M^2 (z_L - f). \quad (7.48)$$

Using these expressions, after calculating the magnification using Equation 7.45, it is straightforward to determine the characteristics of the Gaussian beam after it traverses the lens.

Example 7.8: Let us see what happens if we move the Gaussian beam waist far from the lens; since the Gaussian beam then becomes a spherical wave, we expect the results to approach those of normal ray optics.

For large z_L , $z_L - f \gg z_{0L}$, so that

$$\frac{z_{0L}}{z_L - f} \ll 1$$

implying that the magnification becomes simply

$$M \approx \left| \frac{f}{z_L - f} \right|.$$

If we recall Equation 7.27, and that $z_L - f = z_o - f$ (where z_o is the spacing between focal length and object position), we see that the previous expression is that for transverse magnification in the ray optic approximation.

Furthermore, using the relationship $2z_{0R} = 2M^2 z_{0L}$, we can show (see Problems), that the relation

$$\frac{1}{f} = \frac{1}{z_L} + \frac{1}{z_R}$$

holds. This last relationship is of course the Gaussian lens formula, which we also derived from consideration of the ray-optical characteristics of a lens. Thus, in the limit, the Gaussian beam characteristics of lenses approach those of ray optics.

Special case: focussing

One important function for single lenses is focussing a collimated beam. Since we know that “collimated” implies, strictly seen, a Gaussian beam with a large depth of focus, we can apply the principles just outlined and realize that focussing a Gaussian beam is best accomplished if the lens is placed at the waist of a collimated beam (one with large $2W_0$ and $2z_0$) as shown in Figure 7.22. In that case, $z_L = 0$ so that

$$M = \frac{1}{\sqrt{1 + \left(\frac{z_{0L}}{f}\right)^2}} \quad (7.49)$$

and $2W_{0R} = 2MW_{0L}$ from above. From Equation 7.48, we furthermore see that

$$z_R = M^2(-f) + f = f \left[\frac{1}{\left(\frac{f}{z_{0L}}\right)^2 + 1} \right]. \quad (7.50)$$

For approximately collimated input, $z_{0L} \gg f$ and thus, from Equation 7.49, $M \approx f/z_{0L}$ whereby $M \ll 1$. We then have, for the focussed side,

$$2W_{0R} \approx 2 \frac{f}{z_{0L}} W_{0L} \quad (7.51)$$

$$z_R \approx f. \quad (7.52)$$

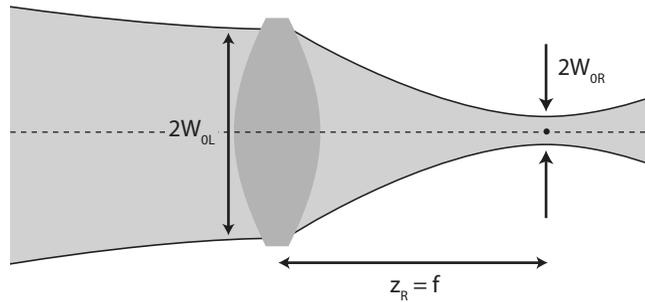


Figure 7.22 Focussing a Gaussian beam: the lens is placed at the beam waist of a wide approximately collimated beam, with large width $2W_{0L}$; on the right (focus) side, the beam waist minimum ($2W_{0R}$) is close to the focal point of the lens.

The focussed spot is found on the focal plane and has a width $2W_{0R} \propto 2f/z_{0L}$, which is small since we have $z_{0L} \gg f$. For the smallest focussed spot, then, a short focal length and a large depth of focus are desirable. Since the latter parameter, from Table 2.5, is given by $2z_0 = kW_0^2$, the focussed spot size may be expressed as

$$2W_{0R} \approx 2 \frac{f\lambda}{\pi} \frac{1}{W_{0L}} \quad (7.53)$$

so that spot size decreases with decreasing focal length and wavelength, and increasing input beam width. Thus a wide input beam with large depth of focus, meaning well-collimated, is optimal for achieving a fine focus spot.

If we fully illuminate the aperture of the lens, such that $2W_{0L} = D$, we can then re-write the previous expression as

$$2W_{0R} \approx 2 \frac{f\lambda}{\pi} \frac{2}{D}; \quad (7.54)$$

recalling that f-number was defined as $f/\# = f/D$, the focus point width may be expressed as

$$2W_{0R} = \frac{4\lambda}{\pi} f/\# \quad (7.55)$$

showing again that resolution is improved for “faster” lenses, i.e., those with a lower f-number.

Special case: collimation

For collimation of a Gaussian beam, we can simply use the system in reverse. “Point” sources may frequently be represented by Gaussian beams with a small beam waist width and small depth of focus; a laser diode, for example, may be approximated in this manner, albeit with a strong astigmatism.

Let’s consider Figure 7.22 from right to left: we place the source (the beam waist minimum) at the focal point of the convex lens (labeled $2W_{0R}$). The collimated beam

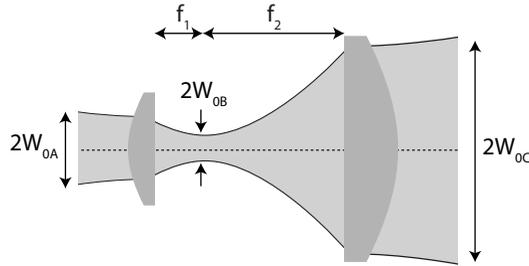


Figure 7.23 A Gaussian beam may be expanded from $2W_{0A}$ to $2W_{0C}$ using an inverted telescope consisting of two plano-convex lenses in an afocal arrangement. The first plano-convex lens with f_1 reduces the beam to diameter $2W_{0B}$ and the second, with f_2 , expands it to $2W_{0C}$.

emerges to the left, with width $2W_{0L}$ and depth of focus

$$z_{0L} = \frac{f}{M} \quad (7.56)$$

where we recall that M is *small*. By considering reciprocity (that (most) optical systems are invariant if we trace them left-to-right or right-to-left), this conclusion may seem trivial, but we nevertheless see clearly here how collimation is realized with a Gaussian beam.

Example 7.9: How can we expand a Gaussian beam using what we have learned about lenses? Let us assume we have a beam with narrow waist $2W_{0L}$, but large depth of focus z_{0L} . Using two lenses, we can expand the beam to maintain a large depth of focus but with a larger beam waist width.

Expansion is best performed using two convex lenses in an “inverse telescope” arrangement: basically, we take the Keplerian telescope of Figure 7.20 and turn it around so the narrow beam enters the eyepiece from the right, the arrangement given in Figure 7.23. From the results of Example 7.7, we can then use ray optic analysis to write

$$\frac{d'_2}{d'_1} = \frac{f'_2}{f'_1}$$

for (narrow) input beam width d'_1 and (wide) output beam width d'_2 . For $f'_2 > f'_1$, the beam is thus expanded, as desired.

How does this look for a Gaussian beam analysis? We see from the afocal arrangement of Figure 7.23 that the first lens, with a small f_1 , placed on the side of the input beam, generates a focus spot, with small $2W_{0B}$ and reduced z_{0B} , at the focal point. We see from Equation 7.49 that, for $z_{0A} \gg f_1$, we have, for the left lens

$$M_1 = \frac{1}{\sqrt{1 + \left(\frac{z_{0A}}{f_1}\right)^2}} \approx \frac{f_1}{z_{0A}}$$

and the focussed spot is at the focal point, f_1 . Using Equations 7.46 and 7.47, we can then write

$$2W_{0B} = 2M_1 W_{0A} = \frac{2f_1}{z_{0A}} W_{0A}$$

and

$$2z_{0B} = 2M_1^2 z_{0A} = 2 \frac{f_1^2}{z_{0A}^2} z_{0A} = \frac{2f_1^2}{z_{0A}}.$$

Since the focal point f_1 is also the focal point f_2 for the right-hand lens, the same expressions may be used in reverse to define the width of the outgoing beam. We thus have

$$2W_{0B} = 2M_2 W_{0C} = \frac{2f_2}{z_{0C}} W_{0C}$$

and

$$2z_{0B} = \frac{2f_2^2}{z_{0C}}.$$

Equating the expressions for $2W_{0B}$ and $2z_{0B}$, we then obtain

$$\frac{z_{0C}}{z_{0A}} = \frac{f_2^2}{f_1^2}$$

and finally

$$\frac{W_{0C}}{W_{0A}} = \frac{z_{0C}}{z_{0A}} \frac{f_1}{f_2} = \frac{f_2}{f_1}.$$

We thus see that the Gaussian optics and ray optics results agree. The expanded beam, with a width $2W_{0C}$, is expanded by a factor f_2/f_1 , for $f_2 > f_1$; the depth of focus is likewise increased, by a factor f_2^2/f_1^2 , so that the output is more “collimated” than the input beam.

7.4 Primary aberrations

For the discussion to this point, we have assumed ideally spherical lenses and analyzed these in the paraxial approximation, i.e. for rays close to the optical axis. We have also implicitly ignored any effects of the material and assumed simply that it is a region with a given constant refractive index.

A more realistic examination of lens characteristics quickly shows the limitations of this idealized view. Real images are not paraxial, so that most applications require that we consider the details of off-axis illumination. In addition, as we saw extensively in Section 3.2, material dispersion leads to a variation in the refractive index with wavelength, such that n is not constant unless we employ a single wavelength.

These two considerations lead to lens aberrations, non-idealities in the imaging properties of lenses which can lead to a significant departure of real performance values from those determined by the idealized models considered above. Aberrations may be divided into two families: achromatic, or Seidel, aberrations, derive solely from the shape of the lens and are present even for imaging with a single wavelength; these are the topic of this section. Chromatic aberrations, due to the dispersion which results from the wavelength-dependence of refractive index, are the topic of the next section. We note that the aberrations we consider here are not due imperfections in the shape or material of the optics, but are intrinsic to spherical lenses, even those manufactured perfectly and without defects.

The consideration of and compensation for optical aberrations is probably the task which most preoccupies the classical lens designer. Whereas putting together a rudimentary imaging system is not particularly difficult, the talent for designing a high-quality optical system with minimal aberrations is that which distinguishes an exceptional optical designer. It is for the minimization of aberrations that most macroscopic optical components consist of complex combinations of lens elements.

7.4.1 Theoretical basis

Recall that the paraxial approximation was based on the assumption that rays close to the optical axis result in an angle θ in Figure 7.2 sufficiently small such that $\sin \theta \approx \theta$, leading to the simplified and useful form of Equation 7.16. This first order approximation is derived from taking only the first term of the series expansion for the sine,

$$\sin \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \dots \quad (7.57)$$

If we include the second term of the expansion in our approximation, we then obtain a more comprehensive form of Equation 7.16, namely

$$\frac{n_o}{S_o} + \frac{n_L}{S_i} = \frac{1}{R} (n_L - n_o) + y^2 \left[\frac{n_o}{2S_o} \left(\frac{1}{R} + \frac{1}{S_o} \right)^2 + \frac{n_L}{2S_i} \left(\frac{1}{R} - \frac{1}{S_i} \right)^2 \right]. \quad (7.58)$$

This expression provides a third-order approximation for the imaging properties of a lens. Recalling that y in Figure 7.2 is the height above the optical axis that an optical ray strikes the lens and that h is the height of the origin of this optical ray⁸, the deviations from the first-order model we considered previously scale with the product of h and y , leading to five so-called primary aberrations: spherical aberration ($\propto h^3$); coma ($\propto hy^2$); astigmatism and field curvature ($\propto h^2y$); and distortion ($\propto y^3$). Taking still higher order terms in the series expansion (leading to aberrations $\propto h^5$, $\propto h^7$ and so forth) leads to 5th, 7th or higher order aberrations. Fifth order are known as secondary aberrations, and the relative relevance of these, and higher order aberrations, decreases rapidly.

⁸ Again, look ahead to Figures 7.24 and 7.26 for a depiction of the parameters y and h .

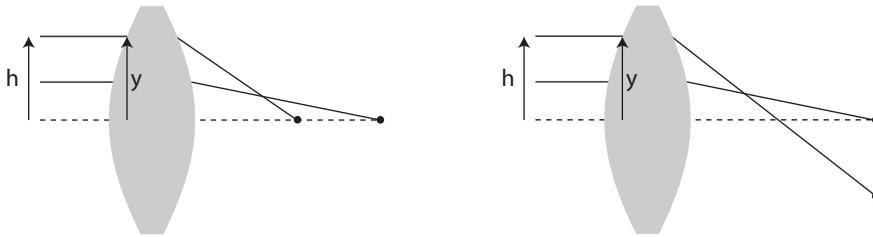


Figure 7.24 Spherical aberration, due to increasing distance h of the beam from the optical axis; longitudinal SA *left* and lateral or transverse SA *right*.

The third order, or primary, aberrations are frequently referred to collectively as Seidel aberrations⁹. The five Seidel aberrations are present in spherical lenses, even for illumination with a single wavelength and we consider these separately in the following sections; for a more detailed discussion, particularly with regard to correction, we refer the reader to (Smith, 1990, Chapter 3) or (Hecht, 2002, Chapter 6.3).

7.4.2 Spherical aberration

Consideration of Equation 7.58 shows that imaging is a function of y , the distance from the optical axis that the beam strikes the lens. The equation shows that, for a given fixed lens radius R and an infinite object distance, $S_o \rightarrow \infty$, the image distance S_i , or the focal point, varies as y increases, shown in Figure 7.24 for a collimated input beam, for which $y = h$. This effect gives rise to spherical aberration, as shown in the sketch.

The effect of spherical aberration, or SA, is thus that the focal length f changes as a function of ray height from the optical axis, y . For *positive* spherical aberration, the focus moves to the left (i.e., toward the lens) as ray height increases, giving rise to longitudinal SA, as seen in the left of the figure. Alternatively, if we follow the rays to the focal plane, we see that the position of the focus is increasingly far from the optical axis as y increases, giving rise to lateral or transverse SA, seen on the right. Longitudinal and lateral SA are just two different manifestations of the same aberration.

Correcting spherical aberration

Spherical aberration is a fundamental problem with spherical lenses, and occurs even for on-axis images¹⁰. There are a number of approaches for correction or minimization of SA. If complexity (or the budget to pay for it) is not an issue, aspheric lenses, whose profile is machined to more closely approximate the elliptical ideal, may be employed, and SA completely eliminated. In addition, aplanatic lenses are meniscus

⁹ Philipp Ludwig von Seidel (1821-1896), German astronomer, optician and mathematician, spent most of his career in Munich and was not only the first to make a systematic study of geometric aberrations in lenses but also the first to apply probability theory to optics as well as to the distribution of diseases and changes in climate.

¹⁰ The expression “on-axis” refers to rays parallel to the optical axis, and should not be confused with paraxial. “Off-axis” rays are oblique or skew rays.

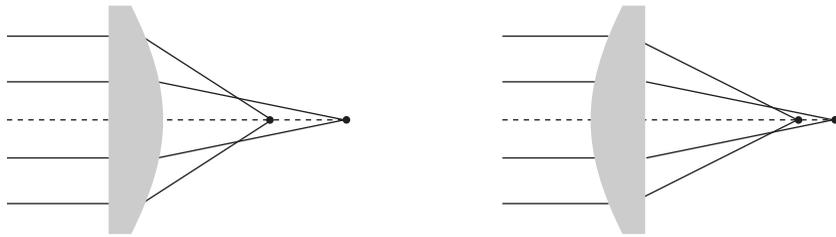


Figure 7.25 How orientation of a plano-convex lens is important in minimizing SA when focussing a collimated beam; with the planar side toward the collimated side *left* SA is worse than for the curved side toward the collimated beam *right*.

lenses which do not introduce spherical aberration (or coma) into the optical system if properly deployed. Alternatively, since concave lenses have *negative* SA, combinations of convex and concave lenses may be employed, allowing for compensation of spherical aberration due to the two polarities.

Using the correct lenses and properly placing these in the optical system are also useful tips for reducing the extent of spherical aberration. For 1:1 imaging applications, or if $S_i = S_o$, a biconvex lens with equal curvature on both sides optimizes SA. For microlenses, multiple layers of varying refractive index in the lens body have been used to minimize spherical aberration (Das et al., 1998). In the following example, we see how proper orientation of a lens is also essential for reducing spherical aberration.

Example 7.10: A plano-convex lens has less spherical aberration than a biconvex lens, such that the former is preferred for focussing the collimated emission from a laser diode. However, how should be orient it to minimize spherical aberration and thus minimize the achievable focus spot?

Figure 7.25 shows the two possibilities. The form of the rays is of course different on the two sides of the lens: on the collimated side, the rays run parallel to the optical axis; on the focus side, they are angled toward the focus. By ray tracing, as shown in the figure, we find that orienting the curved side of the lens toward the collimated beam results in less SA than by orienting the planar side in that direction. As a rule of thumb, the lens should be oriented so that the angle of incidence, with respect to the incident lens surface, of the incident rays is as close as possible to the angle of the refracted rays, with respect to the second surface. As a result, the orientation on the right is to be preferred.

By extension, if we use a plano-convex lens to collimate the emission from a laser diode (which may be considered to be a point source), the planar side should be directed toward the laser to provide optimal collimation. Most commercial collimators are of the plano-convex configuration, and mounting it backwards is a common mistake.

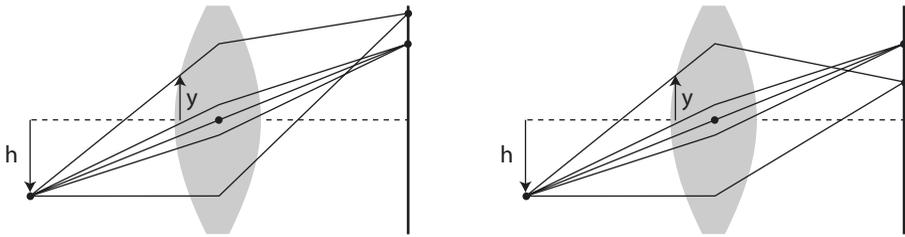


Figure 7.26 Coma is due to variable transverse magnification as a function of ray incidence height y . Positive coma *left* has the principle rays focused above the marginal rays; the opposite holds for negative coma *right*.

7.4.3 Coma

The transverse magnification of a lens, M_T , is a function of the intersection height of the ray with the lens surface, y . As the diagram of Figure 7.26 shows, the principle rays (those which traverse the lens through its center¹¹) and the marginal rays (those which traverse closer to the edge) are focused to different positions on the focal plane. This height-dependence of focus position leads to coma: the image of off-axis rays often has the distinct comatic shape, whereas on-axis objects are not subject to this aberration.

Coma leads to asymmetric, strongly distorted focus points, whose shape (high intensity near the center, with decreasing intensity in a “tail” which extends away from the center of the focal plane, as shown in the simulation result in Example 7.11) is reminiscent of that of a comet, hence the name. Two types of coma may be distinguished, as shown in the figure: positive coma results in the principle rays coming to a focus above the marginal rays; for negative coma, the opposite holds. Due to its asymmetric structure, coma is a disturbing aberration in optical systems but can easily be seen using a simple magnifying glass on a sunny day: rotate the lens so that it is no longer normal to the optical axis defined by the sun, and the focal point will take on the distinct comet-like shape.

Correcting coma

Coma is strongly dependent on lens shape as well as the positions of stops and apertures. An aperture, or stop, is a defined region through which light can be transmitted through a lens, as we saw for a single mounted in lens in Figure 7.16. Whereas the aperture of a single lens is given by its diameter, as shown in the figure, stops (typically opaque diaphragms with a defined opening or, in micro-optical systems, often deposited opaque layers with etched perforations), can be placed at different positions around a lens or lens system, limiting the transmission of off-axis and marginal rays.

As a result, whereas there is no coma for on-axis objects, for off-axis imaging correct placement of field stops can reduce the effects of this aberration. Such an approach is particularly fruitful if a controlled amount of spherical aberration is retained in the

¹¹ Strictly speaking, the principle ray traverses the center of the aperture stop of the lens system; for a single lens, this corresponds to the lens center.

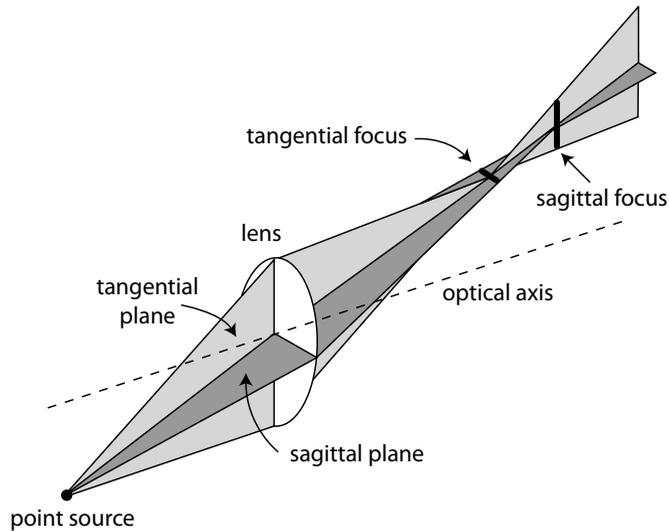


Figure 7.27 Astigmatism results from a difference in the horizontal and vertical magnification for off-axis objects and images. The rays in the tangential and sagittal planes focus at two different positions.

system. This latter consideration is due to the fact that the five primary aberrations are not independent of each other: by considering the aberrations in their standardized order (SA, coma, astigmatism and field curvature, and distortion), it is often the case that correction of one of these is facilitated by incorporating a controlled amount of the previous one on the list into the optics. Hence reduction of coma can benefit from a small amount of SA in the system.

7.4.4 Astigmatism

For off-axis rays, the focal length may differ for the horizontal and vertical directions, even for spherically-symmetric lenses, giving rise to astigmatism. Consider the image of an off-axis point by tracing a principle ray through the lens center, as illustrated in Figure 7.27. We may define two orthogonal planes of rays: the tangential, or meridional, plane is the one which contains the principle ray and the optical axis. The sagittal plane is the one which contains the optical axis and is orthogonal to the tangential plane. The rays found in the tangential plane are known as meridional rays and can be defined using only two coordinates (x and y); all other rays, those not found in the tangential plane, are known as skew rays and require three coordinates (x , y and z) to be characterized.

In Figure 7.27, we see that the rays in these two planes come to a focus at two different positions on the image side of the lens, at the tangential and sagittal foci, respectively. As a result, there is a region between these two foci (known as the “circle of least confusion”¹²), at which the focus diameter is minimized, but is best characterized by a

¹² The “circle of *most* confusion” is typically found on the desk of 10-year-olds doing their homework.

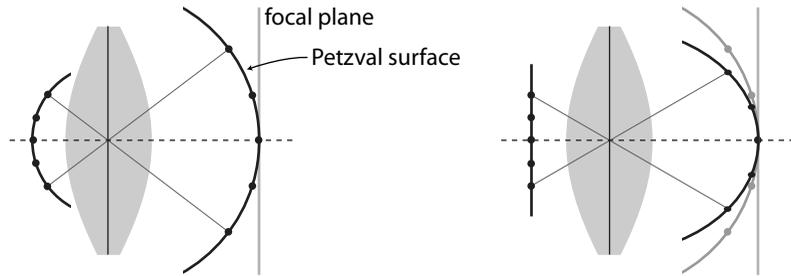


Figure 7.28 An ideal optical system would have object and image describing a segment of a circle *left*; the image is on the Petzval surface. The image of a planar object describes a different Petzval surface, which is displaced from the focal plane *right*.

more-or-less fuzzy blur. As with coma, astigmatism is worse the farther the object point is located from the optical axis and is zero for on-axis objects.

Astigmatism is very difficult to correct and is typically done using a pair of lenses (a doublet), spaced apart, with a properly positioned field stop in between. Astigmatism as a primary aberration, incidentally, is not the same as astigmatism of the eye. For the latter case, the lens has different curvatures in the two orthogonal directions, and correction may be achieved using a corresponding correction lens with variable curvatures. We stress, however, that astigmatism, as we consider it here, is also present for off-axis imaging using circularly symmetric lenses.

7.4.5 Field curvature

The ideal form for objects and images in a lens-based imaging system would be as portions of circular contours, as shown on the left of Figure 7.28. Since objects are arbitrary in shape but often planar, and since imaging is generally on a focal plane, the difference between the planar approximation (which holds only in the paraxial case) and the curved ideal leads to aberrations due to field curvature, also known as Petzval curvature¹³.

Field curvature may be positive or negative, depending on whether the image curvature is increased or decreased with respect to the ideal Petzval surface, respectively; convex lenses have positive field curvature, concave lenses negative. Astigmatism may be considered to be special case of field curvature: the meridional and sagittal planes define two curved image fields, with different radii of curvature, and a reduction in astigmatism is equivalent to equating these two radii.

The fact that concave and convex lenses have different signs for field curvature suggests a method for reducing this aberration. Indeed, using a convex/concave lens pair

¹³ József Miksa Petzval (1807–1891), was born in what was at the time Hungary (now Slovakia) and became professor of mathematics in Vienna. Despite the fact that he dedicated himself primarily to the development of mathematical tools akin to the Laplace transform, he is best known for his study of aberrations in lenses and his manufacture of high-quality achromatic photographic optics.

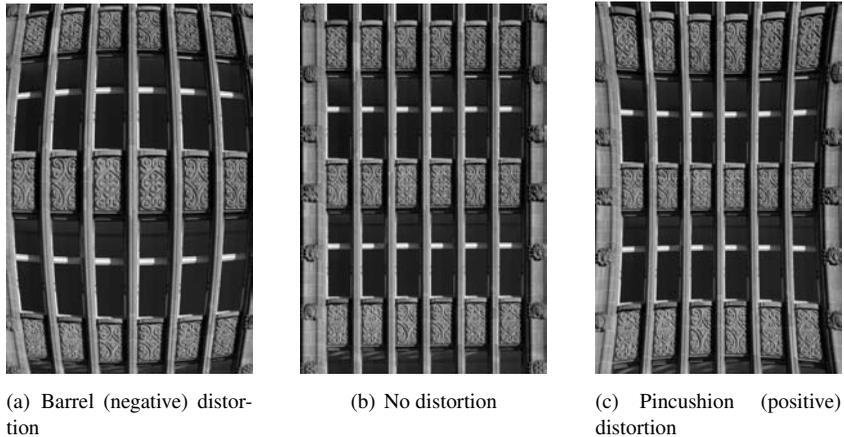


Figure 7.29 Windows at the University of Zurich subject to varying amounts of distortion; barrel distortion *left* typically results when photographing with a wide-angle objective whereas telephoto lenses sometimes have slight pincushion distortion *right*.

while maintaining the condition

$$n_1 f_1 + n_2 f_2 = 0 \quad (7.59)$$

or

$$f_1 = -f_2 \quad (7.60)$$

for lenses of identical material for which $n_1 = n_2$, field curvature may be corrected.

7.4.6 Distortion

The last of the primary aberrations is distortion, which results from transverse magnification M_T which varies with ray height h . As the object or image distance from the optical axis increases, M_T either increases to yield positive, or pincushion, distortion or decreases to yield negative, or barrel, distortion. The two cases are clearly seen in Figure 7.29, where the undistorted windows of Figure 7.29(b) deform in the presence of barrel negative (Figure 7.29(a)) or positive (Figure 7.29(c)) distortion.

Correction of distortion is also best achieved using spaced doublets with an intermediate field stop. Distortion is markedly worse for thick lenses and can be negligible for truly thin lenses. As we realize from the discussion above, aberrations are closely intertwined and correction is usually only possible by employing multiple lenses and cleverly positioned field and aperture stops. Most of the effort which goes into the design of a macroscopic multi-lens system is in the reduction of aberrations, and the topic is considerably more extensive than the brief view we have given here.

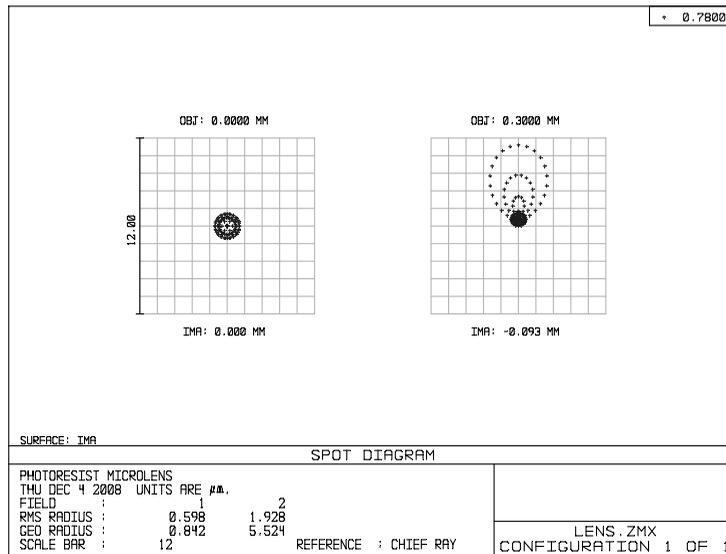


Figure 7.30 The spot diagrams for the two ray bundles traced in Figure 7.13 of Example 7.4. The on-axis focal spot *left* shows the expected circular symmetry (as seen previously in Figure 7.13, but with a different scale) whereas the oblique rays *right* generate a distinctly comatic image.

Example 7.11: Using ZEMAX and the same microlens arrangement we considered in Example 7.4, let us use numerical simulation to visualize some of the aberrations.

Considering the ray bundles traced in Figure 7.13, Figure 7.30 shows the calculated spot diagrams on the focal plane. As we saw in Example 7.4, the focus spot of the on-axis ray bundle (on the left) appears circular, as we expect for a paraxial beam. The off-axis bundle (on the right), on the other hand, generates a distinctly comatic focal point. Coma, which becomes worse the farther the ray origin is from the optical axis, generates the distinct “comet-shaped” image apparent from the spot diagram.

If the same microlens is subject to illumination by an oblique ray bundle, astigmatism is also clearly seen in the ray tracing results. Figure 7.31 shows the sagittal and tangential foci, which are found in different planes. The circle of least confusion, at which the smallest spot size can be found, lies between these focal points, as we saw in Figure 7.27.

7.5 Chromatic aberrations

The primary aberrations we considered in the previous section are inherent to the shape of the lens, independent of material and illumination wavelength. We recall from our

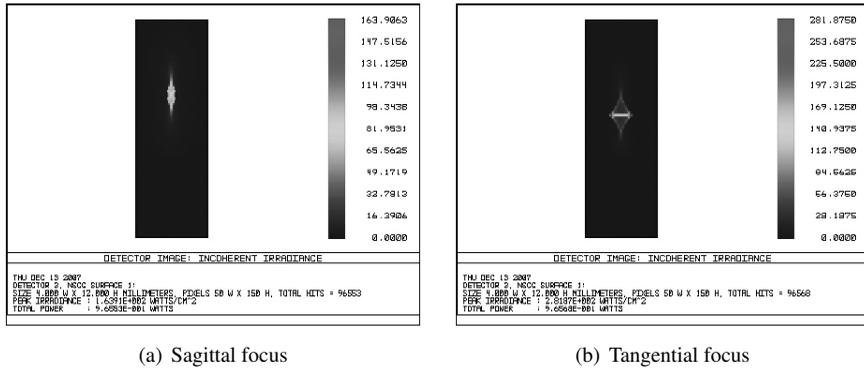


Figure 7.31 ZEMAX simulations of the photoresist microlens in Figure 7.13 with strongly oblique illumination at $\lambda = 760$ nm showing the distinct astigmatism; the two foci are due to rays in the sagittal and tangential planes.

discussion in Section 3.2 that materials are subject to dispersion, meaning that the refractive index varies with wavelength. As a result, a lens can only be optimally designed for operation at a single wavelength and the resultant deviation from ideal performance at other wavelengths is termed chromatic aberration.

7.5.1 Effect of chromatic aberration

For most optical materials, dispersion implies that the refractive index decreases with increasing wavelength; Figure 3.5, for example, shows the variation of real refractive index, n_R , of SiO_2 at visible and near-infrared wavelengths. As a result, the angles of refraction vary with wavelength, implying that ultimately the focal length is also wavelength-dependent. This effect is illustrated in Figure 7.32, where we see that the focal length for short (blue) wavelengths is closer to the lens than that for long (red) wavelengths; this difference gives rise to axial chromatic aberration. As a result, the focussed spot of a collimated beam with a wide spectrum (for example, of white light) is more a diffuse blur, whose color varies as the focal plane is moved. Analogously to the case for astigmatism, the position with the smallest spot size is referred to as the circle of least confusion.

Change in focal length with wavelength leads to a wavelength-dependent magnification for the image, resulting in lateral chromatic aberration. Whereas lateral and axial chromatic aberrations share the same origin, the manifestation of the effect is different for the two cases. Lateral chromatic aberration leads to two-dimensional images whose focal plane varies with wavelength, as seen in Figure 7.33, leading to composite images with colored fringes or halos when viewed in a single plane.

Chromatic aberration (CA) is quantitatively characterized by taking the difference between focal lengths or magnifications for two standard wavelengths; the Fraunhofer C (red) and F (blue) lines, presented in Table 3.1 for definition of the Abbe number (ν),

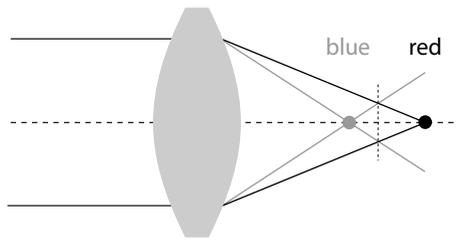


Figure 7.32 Axial chromatic aberration: the focus of longer wavelengths (red) is longer than for shorter wavelengths (blue). The “circle of least confusion” is the area of optimal focus between these two extremes.

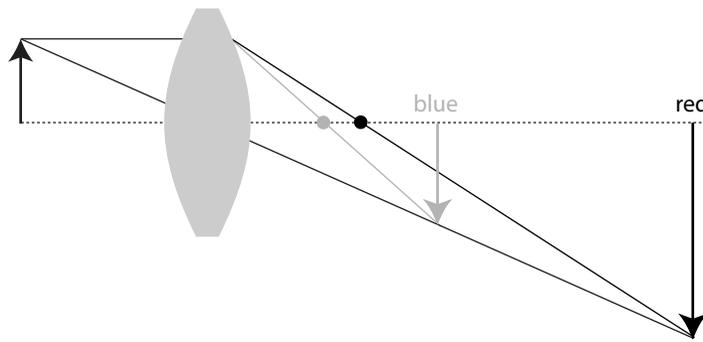


Figure 7.33 Lateral chromatic aberration results from wavelength-dependent magnification, so that the position of the focal plane and the size of the image varies with wavelength.

are commonly used. As a result, we may quantify axial chromatic aberration as

$$CA_{axial} = f_C - f_F \quad (7.61)$$

and lateral chromatic aberration using

$$CA_{lateral} = M_{T(C)} - M_{T(F)} \quad (7.62)$$

for transverse magnification M_T . Smaller numerical values for CA imply less chromatic aberration.

7.5.2 Correcting chromatic aberration

In the illustrations above, we saw that the focal length for blue wavelengths is shorter than that for red wavelengths. A useful rule of thumb for determining the wavelength-dependent position of f is that light is refracted more toward the thicker part of the lens and blue is refracted more than red. As a result, f_F (blue) is found to the left of f_C (red), for rays coming from the left.

By applying the same argumentation to a concave lens, we find the opposite condition, namely that f_F is to the right of f_C . This result suggests a means for correcting chromatic aberration: by combining convex and concave lenses with the appropriate

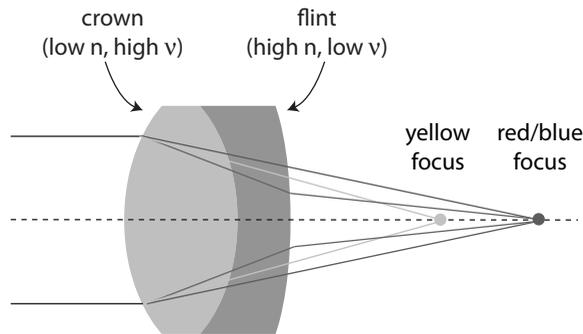


Figure 7.34 A chromatically-corrected Fraunhofer doublet consists of a convex lens, of crown glass with low refractive index and dispersion, cemented to a concave lens, of flint glass with high refractive index and dispersion. Red and blue rays have the same focus, at the right, and yellow, comes to a focus further to the left.

refractive indices and dispersion (Abbe numbers), CA can be minimized for a certain wavelength range, and reduced to zero for two wavelengths.

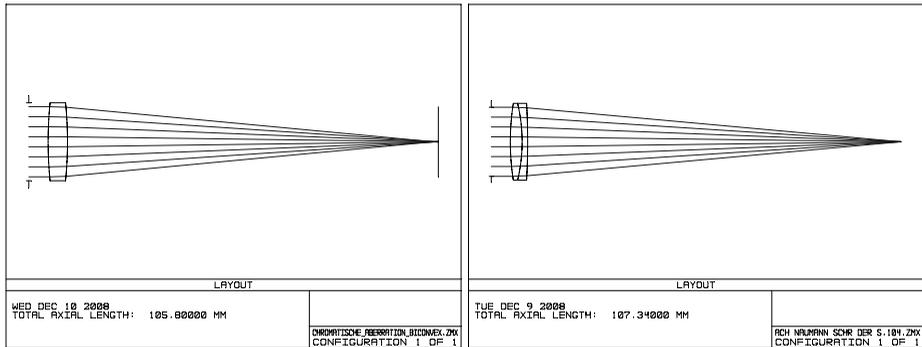
An example of such a corrected lens is the Fraunhofer doublet shown in Figure 7.34. The doublet consists of a convex lens, of crown glass with low refractive index and dispersion (high Abbe number, ν), followed by a concave lens, of flint glass with high refractive index and dispersion (low ν). The two lenses may either be cemented (the convex surface is concomitant with the concave surface) as shown, but may also contact only at the center or only at the edges, if the absolute values of the radii of the two lenses are different. Other types of achromatic doublets are also in use, for example the Steinheil configuration which uses two meniscus lenses. Apochromatic lenses are those designed so that three wavelengths come to a common focus.

This combination of two lenses allows correction of chromatic aberration for two wavelengths, and these are usually chosen to be at the extremes of the spectrum, thus typically red and blue. Thus $f_C = f_F$ for a properly designed doublet. Other wavelengths focus closer to the lens, however, and the strongest deviation is typically for yellow (the Fraunhofer d-line, for example), since this wavelength is furthest from the extremes. Thus, whereas the doublet greatly improves CA performance, some chromatic aberration remains.

Example 7.12:

One of the truly powerful aspects of numerical optical simulation is that complex material characteristics, particularly dispersion, can directly be considered. We will thus use ZEMAX again to demonstrate chromatic aberration; to illustrate the effect more clearly, we will use macroscopic lenses, with a diameter of 20 mm and an aperture of 18 mm, rather than microlenses for the simulation.

In Figure 7.35(a), we see the simulation of a standard BK7 biconvex lens, focussing a collimated beam. Figure 7.36 shows the width of the optical beam (by means of the spot diagram) transmitted through this lens as a function of defocus, the distance from



(a) A BK7 biconvex lens with a diameter of 20 mm and an aperture of 18 mm. (b) A SK-11/SF-9 achromatic lens with a diameter of 20 mm and an aperture of 18 mm.

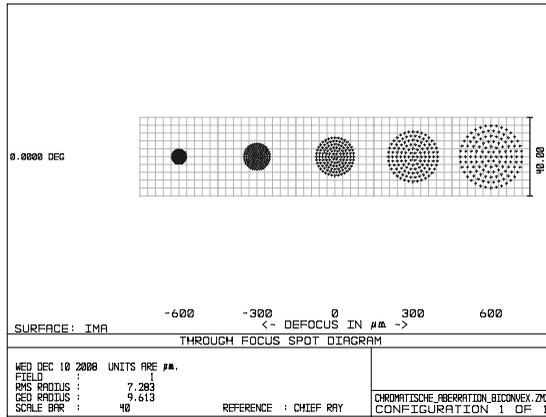
Figure 7.35 ZEMAX simulations of standard BK7 biconvex lens and a SK-11/SF-9 achromatic doublet lens, both bringing a collimated beam to a focus for different wavelengths.

the focus point; defocus zero is thus on the focal plane, defined for $\lambda = 587$ nm for this example. Looking thus at the position of the focus for three standard wavelengths (486, 587, and 656 nm, the Fraunhofer F, d and C lines, or blue, yellow and red, respectively), we see that the position of the focus moves from left to right by about 1 200 μm as the wavelength increases, clearly demonstrating chromatic aberration.

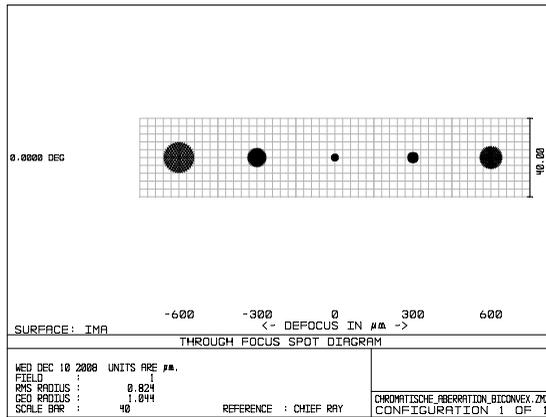
We compare these results with those attained using an achromatic doublet made from SK-11 crown glass and SF-19 flint glass, as shown in Figure 7.35(b). In Figure 7.37, the beam width as a function of defocus is calculated for the achromatic doublet: the positions of the focus for the shortest (7.37(a)) and longest (7.37(c)) wavelengths overlap at zero defocus, whereas the yellow (7.37(b)) has about a 200 μm longer focal length. Whereas all three foci at the three wavelengths are not perfectly coincident, they overlap much more closely than for the case of the simple biconvex lens, showing that chromatic aberration has been improved. We also see that the size of the focus spots for the achromatic lens is smaller than for the simple biconvex lens, an additional aspect to the improved performance of the achromat.

7.6 Microlenses: case studies

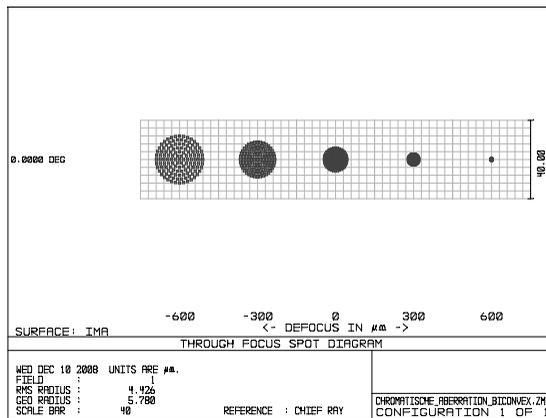
As classical refractive lenses shrink in size to become microlenses, the optical physics does not change: our conclusions above, from imaging to aberrations, hold equally well for microlenses as they do for macroscopic lenses. Diffraction effects do become more important as size decreases, and diffractive optical structures represent a significant, if not the largest, portion of the micro-optics menagerie. We will, however, consider



(a) 486 nm

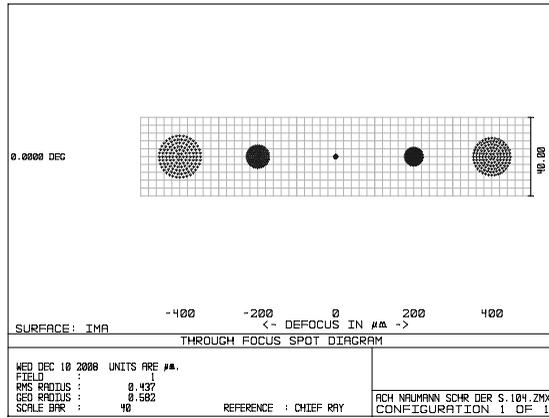


(b) 587 nm

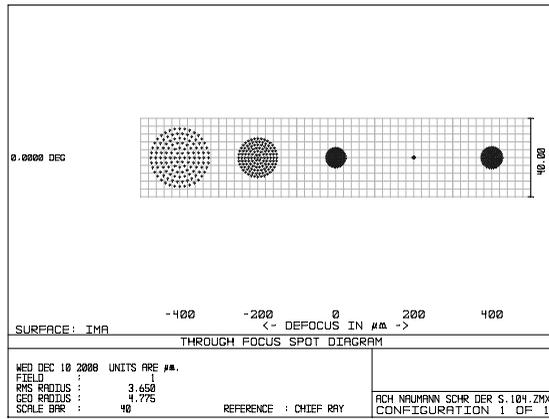


(c) 656 nm

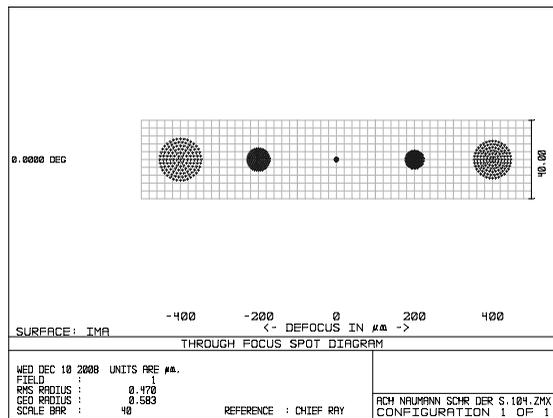
Figure 7.36 The width of the optical beam as a function of defocus position for the biconvex lens of Figure 7.35(a) for three wavelengths; a defocus of zero is defined for $\lambda = 587$ nm. The position of the focus clearly moves from left to right as the wavelength increases.



(a) 486 nm



(b) 587 nm



(c) 656 nm

Figure 7.37 The width of the optical beam as a function of defocus position for the achromatic lens of Figure 7.35(b) for three wavelengths. The positions of the focus for the shortest (7.37(a)) and longest (7.37(c)) wavelengths overlap at zero defocus, whereas the yellow (7.37(b)) has a larger focal length.

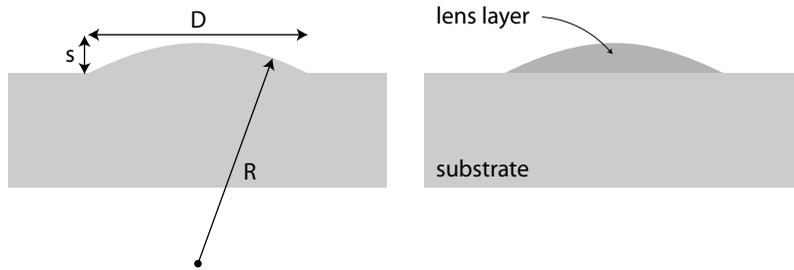


Figure 7.38 Typical microlens structures, fabricated on a planar substrate; either the substrate material itself is patterned *left* or the lens is fabricated in a surface layer *right*. The substrate is typically much thicker than the lens. The sag height is given by s , the aperture diameter by D and the radius of curvature by R .

purely refractive effects here and defer the extensive discussion of diffraction effects to Chapter 8.

Nevertheless, the structures and fabrication techniques do change as refractive lens diameters decrease. Perhaps most significantly, different approaches for manufacturing lenses are required for microlenses than for macroscopic lenses, resulting in subtle changes in the overall lens structure. In addition, as with many aspects of microsystems, small sizes also give rise to effects and configurations which are inapplicable in macroscopic systems. As a result, we consider some of the specialized forms of micro-optical lenses in the following sections, including structures which have no macroscopic pendant.

7.6.1 Refractive microlenses

Classical lens grinding techniques become increasingly less useful as lens diameters shrink below about 1 mm, so that microlenses are fabricated differently, and thus have distinct aspects to their structure which distinguish them from their macroscopic counterparts (Hutley, 1997; Daly, 2001). As shown in Figure 7.38, refractive microlenses are usually not free-standing structures such as those we have considered to this point, but usually formed either by structuring a surface layer on a transparent substrate, or by structuring the surface of the substrate itself.

Whereas some microsystems-based free-standing out-of-plane microlenses have been proposed and demonstrated (King et al., 1996), these remain relatively exotic. The bulk of microlens structures is roughly planar and the optical effect of the substrate (which is typically much thicker than the lens) needs to be considered in ray tracing and in the design of the optical system.

Scaling

As the dimensions of purely refractive microlenses decrease, the resultant change in a number of optical parameters must be considered carefully (Lohmann, 1989; Ozaktas et al., 1994). If we scale all physical features uniformly by a scale factor $S < 1$ (Sinzinger

and Jahns, 2003, Chapter 2), we obtain

$$D' = SD \quad (7.63)$$

$$(x, y, z)' = S(x, y, z) \quad (7.64)$$

$$f' = S^2 f \quad (7.65)$$

for aperture diameter D , physical feature size (x, y, z) and focal length f . Thus aperture and feature sizes scale linearly, and focal length as the square of the scale factor.

As microlens size decreases, the relevance of primary aberrations decreases; smaller lenses generally have shorter focal lengths such that the relative effect of wavefront aberration is smaller. This fact implies that microlenses are much more easily diffraction-limited than aberration-limited, and the diffraction pattern generated by the (small) aperture quickly dominates the other aberrations. The relative effect of diffraction on the performance of a lens is frequently given by the Fresnel Number, N_F [] (Sinzinger, 2007, Chapter 3.02.1.1.2); a value for $N_F \gg 1$ implies purely geometric optics (i.e., diffraction is negligible).

The effect of diffraction becomes more prominent the farther one moves from the lens, such that the maximum achievable focal length is also limited by diffraction. On the other hand, since the chromatic variations resulting from diffraction and refraction have opposite sign, diffraction effects can also be used to make novel types of achromatic microlenses (Ruffieux et al., 2006).

Sizes

Microlenses are generally spherical sections, with a spherical refractive surface, as shown in Figure 7.39. Numerous types of aspheric microlenses have also been fabricated (Grunwald et al., 1997; Chen and Tseng, 2005; Totsu and Esashi, 2005; Shi et al., 2007), but the resultant profiles are generally difficult to control. The relationship between the radius of curvature R , the aperture diameter D and the lens thickness, also known as the sag height s [m], for spherical lenses is given by basic geometry as

$$R = \frac{\left(\frac{D}{2}\right)^2 + s^2}{2s}. \quad (7.66)$$

Typical microlens diameters are from about 8 μm to 1 200 μm . Lens thickness, or sag height, is usually limited by viable etch depths as well as available surface film thicknesses, and have typical values of 10 to 20 μm ; sag heights greater than 80 to 100 μm are difficult to achieve. As a result, the realizable radii of curvature are generally on the order of several hundred micrometers to a few millimeters. These dimensional relationships imply that microlenses tend to be “flat”, with little curvature for their diameter. We note that the dimensions just considered are, however, considerably larger than the wavelength in the visible regime.

Structures

As we will see in Chapter 11, when we consider fabrication in detail, typical microlens configurations use a transparent substrate, such as glass, polymer or, if transparent at

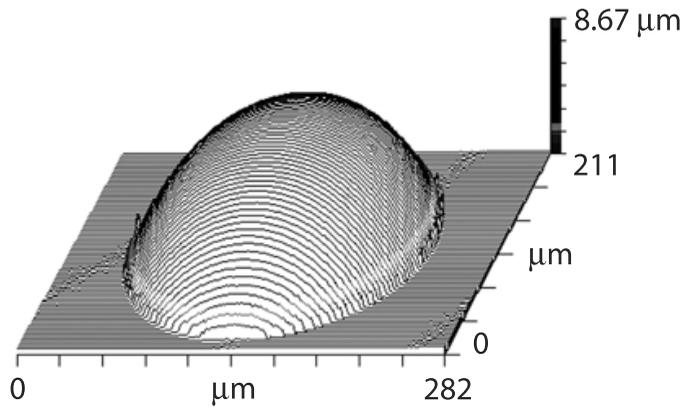


Figure 7.39 The profile of a polymer microlens fabricated on a silicon substrate, measured using a white light interferometer. The sag height of the lens is about 8 μm . Measurement courtesy of Fabian Zimmer.

the wavelengths in use, a semiconductor. These substrates generally have thicknesses ranging from several 100 μm to 1 mm, and need to be of optical quality. The refractive lens surface is either etched into the substrate itself (as seen on the left of Figure 7.38) or patterned into a surface layer deposited on the substrate (as on the right of Figure 7.38). Typical surface layers include photoresist (transparent for $\lambda \gtrsim 650 \text{ nm}$, with refractive index about $n \approx 1.6$), a wide variety of polymers and, particularly for tunable and fluidic optics, liquid droplets. These are patterned primarily using photolithography, etching and molding techniques; machining using single-point diamond turning has also been used, but generally does not achieve the optical quality surfaces required for refractive lenses.

Characteristics

Due to their small size, microlenses tend to be slightly more difficult to characterize than their macroscopic counterparts, but the tools are by and large the same (Daly, 2001, Chapters 5 & 6) (Sinzigler and Jahns, 2003, Chapter 4). To determine the mechanical shape of the lens, surface profilometers are quite useful: contact approaches, which use a stylus dragged over the profile, or, for small structures, an atomic force microscope (AFM), may be used for hard materials, whereas non-contact approaches such as white light interferometry are applicable if the sidewalls are not too steep. A typical measurement result, of a PDMS-based¹⁴ microlens, is that of Figure 7.39.

The optical characteristics are measured in a similar manner to standard lenses, save that a microscope is often required to accurately determine, for example, the focal length. Due to the relatively flat profile (large R with respect to D) of microlenses, the focal lengths generally range from several 100 μm to, at most, 10 mm; typical val-

¹⁴ Polydimethylsiloxane, a silicon-based polymer used, among other things, for contact lenses.

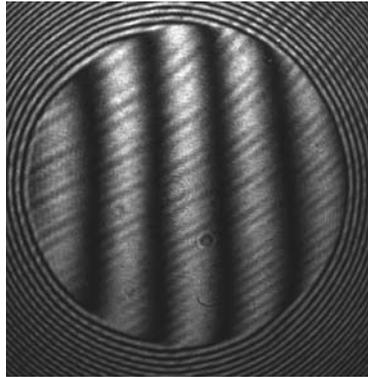


Figure 7.40 An interferogram resulting from the measurement of a 300 μm diameter PDMS microlens using a Mach-Zehnder interferometer. The lens has been tilted slightly to generate the vertical interference fringes; the trained eye would see a slight curvature toward the left. The surface quality of the lens is about $\lambda/7$. Measurement courtesy of Bernd Aatz.

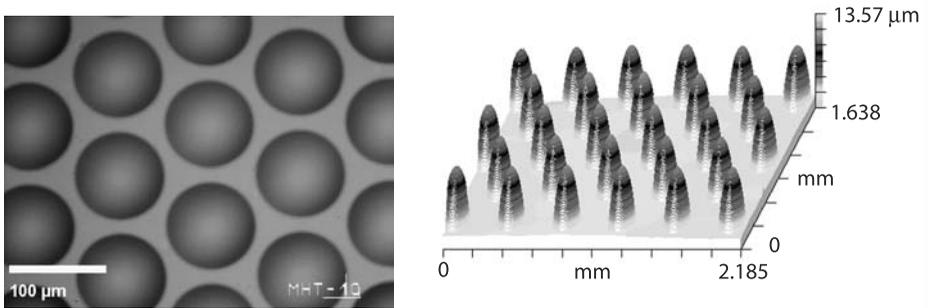
ues are in the millimeter range. As a result, numerical apertures tend to be in the range $0.15 \leq NA \leq 0.45$; typical f-numbers are $1 \leq f\# \leq 3$, but values as low as 0.8 have been demonstrated.

Interferometers also play an important role in microlens testing (Miyashita et al., 2004; Lindlein et al., 2004), as these allow detailed examination of the wavefront; Mach-Zehnder or Twyman-Green arrangements are most practical (Reichelt and Zappe, 2005). The lens profile may be measured directly, or the deformation of the wavefront may be characterized. A typical microlens profile as measured by a Mach-Zehnder interferometer is shown in Figure 7.40. As with macro-lenses, characterization of aberrations is important, and high spherical aberration is a particular problem (Testorf and Sinzinger, 1995); it has been shown, for example, that multi-layer microlenses may be designed to minimize SA (Das et al., 1998).

Arrays

The planar fabrication techniques derived from electronics and MEMS technologies are generally those used for microlenses. As a result, fabrication of microlenses in arrays is as easy as fabrication of individual elements, and this feature allows the design of entirely new optical systems.

Figure 7.41 presents views of a polymer-based microlens array, fabricated by micro-contact printing (see Section 11.2.2); a uniform array of 80 μm diameter microlenses with a sag height of 13.5 μm is generated using this highly-parallel stamping technique, which yields hundreds of very uniform, precisely positioned microlenses. The example shown has a relatively large spacing between the individual elements, but microlens arrays may also be written with a higher packing density, or fill-factor; for rectangular arrays, these lenses are then no longer circular at the edges. The optical function of a microlens array becomes that of a two-dimensional periodic structure (where the period is the lens spacing) convoluted with the function of the individual lens (Sinzinger,



(a) Microscope photograph of a polymer microlens array. (b) Profiles of the same microlens array as measured by a white-light interferometer.

Figure 7.41 Microlens arrays, fabricated using micro-contact printing to define circular hydrophilic areas on the substrate and then using surface tension to define the liquid polymer lenses, which are then cured to solidify them. Lenses and measurements courtesy of Wolfgang Mönch.

2007, Chapter 3.02.2.5.2); as the lens spacing decreases, diffraction effects become more prominent.

Applications

As individual elements, refractive microlenses are relevant for optical microsystems, in which space is at a premium and the optical paths (and thus focal lengths) are short. Coupling between optical fibers, or between fibers and light sources, is one area in which microlenses are heavily represented.

Microlens arrays, however, play an even more important role than individual lenses. Arrays are used for beam homogenization and shaping, as well as concentration, for example of an optical field onto an array of detectors such as a CCD. In addition, since most imaging systems require a large field of view (object and image area), if a single macroscopic lens is employed, it needs a correspondingly large aperture, which in turn implies a long focal length. Large f thus limits the possible reduction in the lateral size of optical systems, which explains why macroscopic optical systems have large spacings between the elements. We can, however, use microlens arrays to perform highly parallel imaging: each microlens represents a sub-aperture of the system, and the image is generated by the overlap of the image segments generated by each microlens (Völkel et al., 1996). Since microlenses have a short focal length, spacings between object and image can be very small, and the dimensions of the overall optical system compressed. Image quality, however, is a critical function of fill-factor and microlens uniformity.

7.6.2 GRIN lenses

Gradient index, or GRIN, lenses¹⁵ use a fundamentally different principle for realizing the refractive power of a lens. The optical power of a GRIN lens is based on a continuous variation of the refractive index profile in a solid material rather than on refraction at a surface.

Principle

Consider the application of Snell's law to a continuous refractive index distribution which varies only in the y direction, $n(y)$, normal to the propagation direction z . The paraxial wave equation, in which $y(z)$ represents the path of an optical ray, becomes (Saleh and Teich, 1991, Chapter 1.3)

$$\frac{d^2y}{dz^2} = \frac{1}{n(y)} \frac{dn(y)}{dy}. \quad (7.67)$$

A particularly useful form of $n(y)$ is

$$n^2(y) = n_{max}^2 (1 - \alpha^2 y^2) \quad (7.68)$$

with a maximum value n_{max} at $y = 0$ and a grading constant α . If we take $\alpha^2 y^2 \ll 1$, we obtain

$$n(y) \approx n_{max} \left(1 - \frac{1}{2} \alpha^2 y^2 \right), \quad (7.69)$$

corresponding to a parabolic refractive index distribution in y . In this case, Equation 7.67 simplifies to

$$\frac{d^2y}{dz^2} = -\alpha^2 y \quad (7.70)$$

which has harmonic solutions of the form

$$y = y_0 \cos \alpha z + \frac{1}{\alpha} \frac{dy_0}{dz} \sin \alpha z \quad (7.71)$$

where y_0 is defined as $y(z = 0)$.

This last expression is a solution to the paraxial wave equation and defines a ray which has an oscillatory trajectory about the center of the index distribution, $y = 0$. The period of oscillation is $2\pi/\alpha$ and is known as the pitch of a GRIN lens.

Rod GRIN lenses

In a GRIN lens, the above-described principle for bending the path of an optical ray is applied by using a circularly symmetric, parabolically varying index profile in a solid cylindrical glass rod; z is the axis of the cylinder and y is the radial direction. A rod GRIN lens then uses this harmonically periodic optical path for focussing or collimating a beam, as shown in Figure 7.42. The sketches, which represent a cross-section of the

¹⁵ GRIN lenses are occasionally referred to as Selfoc lenses whereby the latter is a trademark of Nippon Sheet Glass; "graded" rather than "gradient" index is also sometimes encountered.

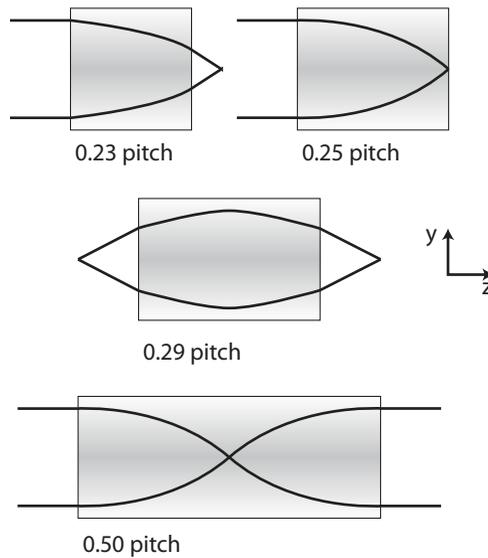


Figure 7.42 Schematic representations of the operating principle behind GRIN lenses; shown is a cross-section through a cylindrical lens. The lines denote the path of the optical field, and the shading the variation of refractive index, decreasing from the center. The pitch refers to the optical length of the lens: a 1 pitch GRIN lens has the same field at the output as was at the input.

cylindrical lens, show the effect of lens length, normalized to the pitch. A so-called 1 pitch GRIN lens (not shown), would result in the same field distribution at the output as was projected into the input. The 0.5 pitch lens inverts the optical field, and a 0.25 pitch lens brings a collimated field to a focus exactly on the end surface of the lens ($f = 0$). Since having a focus on the material surface is usually impractical, a 0.23 pitch lens can be used to place the focal point just beyond the lens surface.

We thus see that the optical function of GRIN lens is determined by its length, and indeed one purchases lenses of specified length for a particular application. The gradient index leads to small modal dispersion, such that we will revisit this refractive index distribution when we consider GRIN fibers for long-distance optical data transmission in Section 9.6. The parabolic index profile is difficult to generate and fabricate, but commercial GRIN lenses are of high quality.

Rod GRIN lenses are popular in micro- or macro-optical systems which have strong demands on lateral spacing; since the focal length of a GRIN lens can be made arbitrarily short (down to zero), spacing between components can be reduced. Individual GRIN lenses are useful in optical communications systems (Tomlinson, 1980); cemented together into one- or two-dimensional arrays, GRIN lenses are most popularly employed in compact planar imaging systems such as copiers, scanners and fax machines.

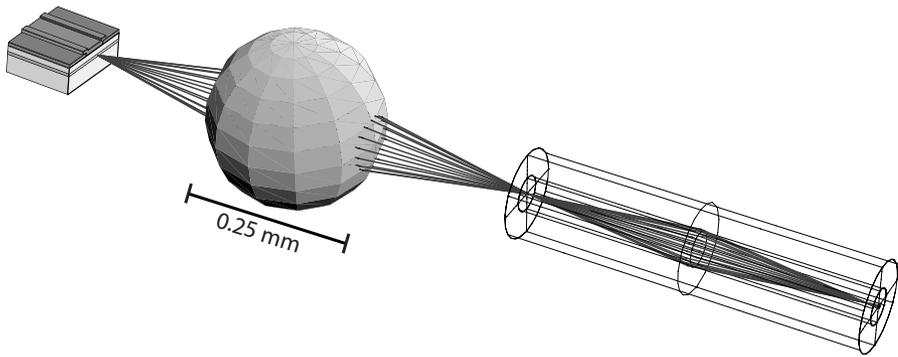


Figure 7.43 A ball lens used for coupling the light from a laser diode chip *left* into an optical fiber *right*. In reality, the emission from the laser is highly astigmatic, such that the angle in the vertical (relative to the chip surface) direction is larger than that in the lateral direction.

Planar GRIN lenses

A continuous refractive index distribution may also be generated below the surface of a slab of an optical material, such as glass or a semiconductor (Oikawa, 2004). Using diffusion of dopants to change the index, the distribution of n becomes that of the diffusion profile, which may also be used for focussing normal to the surface of a planar substrate. Such an arrangement is less useful as a discrete micro-optical component, but can be employed for vertical integration of highly-compact optical microsystems or for the definition of two-dimensional arrays.

7.6.3 Ball lenses

It is often thought that the first lenses were formed by filling spherically-shaped bottles with water and using the resultant sphere for focussing light. The ball microlenses we consider here rely on the same structural form, but are a few orders of magnitude smaller. As the name suggests, the ball lens is a sphere of glass, and the spherical profile is used for refracting an optical beam.

As a complete sphere, the ball lens is most certainly not a “thin lens” and its imaging qualities are limited. However, it is easily positioned, since orientation is irrelevant, and only the position of the center must be assured. Due to their small size and that fact that, unlike microlenses, they are not mounted on a larger substrate, they do require a micromanipulator (or at least tweezers held in a steady hand) for placement and have a propensity to roll around the optical table and fall into the little holes, lost forever.

Structure

Ball lenses are spheres, manufactured by batch polishing of glass, and are occasionally AR-coated. Typical materials include BK7, sapphire, silica, CaF_2 and LaSFN9. Refractive index values are generally in the range 1.5 to 1.8; LaSFN9, for example, is a high-index glass with $n = 1.845$ at $\lambda = 633 \text{ nm}$.

Typical diameters for ball lenses range from 110 μm to 5 mm, with typically available standard sizes of 200, 250, 300, 500 and 800 μm . Focal lengths are very short, generally 100 to 250 μm . Due to their squat spherical shape, spherical aberration is very strong.

Optical characteristics

Ball lenses are rarely used for imaging, so that the relevant optical characterization revolves about determination of focal length. Given a ball lens with diameter D made from a glass with index n_L , operated in air, the effective focal length, measured from the center of the lens, is easily found from

$$\text{efl} = \frac{n_L}{n_L - 1} \frac{D}{4} \quad (7.72)$$

and the back focal length, which is ultimately the spacing from the ball surface to the focal point,

$$\text{bfl} = \text{efl} - \frac{D}{2}. \quad (7.73)$$

Since ball lenses are (very) “thick” lenses, the difference between efl and bfl is significant; typical values for bfl are less than 100 μm .

The numerical aperture depends on the aperture employed; for an aperture diameter D_A , where $D_A \leq D$, we can write NA as

$$NA = \left(\frac{n_L - 1}{n_L} \right) \frac{2D_A}{D}. \quad (7.74)$$

The numerical aperture is thus primarily a function of the fraction of the lens diameter used as the aperture, and for $D_A/D = 2/3$, $NA \approx 0.45$ for a BK7 ball lens.

Applications

Due to their ease of alignment and positioning in an optical system (they fit well into a wet-etched cavity on a silicon substrate, for example), ball lenses are often used for coupling light from a source to sink, such as from a laser diode into an optical fiber, as shown in Figure 7.43, or from fiber-to-fiber. Their small size makes them well suited to incorporation into hybridly-assembled optical microsystems (Lohmann and Syms, 2003). Ball lenses are also often attached to the end facets of fibers, to generate collimated or focussed emission; for fiber coupling, care should be taken to match the NAs of lens and fiber, to assure maximum coupling efficiency.

7.6.4 Micro-Fresnel lenses

In the early nineteenth century, the need to improve the efficiency of lighthouses required ever larger lenses, which, even if possible to manufacture, mount and rotate, would become so heavy that the stability of the building would be called into question. While not the first to propose segmented lenses, it was Jean Augustin Fresnel who deconstructed an aspheric lens into segments and had it installed in the Phare de Cordouan, a wonderfully ornate lighthouse still standing in the Gironde estuary north

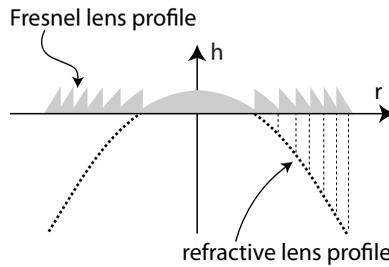


Figure 7.44 How a Fresnel lens is assembled: the aspheric profile of the lens, shown in exaggerated form by the dotted line of the parabola, is folded upwards in discrete segments. Each segment retains the curvature of that part of the lens from which it was taken.

of Bordeaux, in 1822. Whereas other optical luminaries, such as David Brewster, also made what they referred to as “stepped surface” lenses, Fresnel’s was the first which decomposed an aspheric profile and was thus free from spherical aberration. Segmented lenses are today thus generally referred to as Fresnel lenses.

Principle

Micro-Fresnel lenses, a few orders of magnitude smaller than the original, use the same principle. These are purely refractive optical elements formed by essentially removing the bulk of an aspheric lens, since it is only the surface which has a refractive effect. The principle is shown in Figure 7.44: segments of the continuous surface are folded upwards while retaining the original profile. The result is a thin segmented lens, in which the aspheric (or arbitrary) profile is found in the concentric segments. For N segments, the lens thickness is thus reduced by $1/N$.

We stress that the classic Fresnel lens is purely refractive: the segments are orders of magnitude too large for diffraction to play a role. For micro-Fresnel lenses, the relevance of diffraction increases as the segments decrease in size, but the basic functionality we consider here is still that of geometric optics. Fresnel zone plates and other diffractive structures, which superficially look quite similar to refractive Fresnel lenses, will be considered in Chapter 8.

Structure and performance

Micro-Fresnel lenses may be made by machining the segmented structure into glass, by grinding or polishing techniques, or, more popularly, by injection molding or embossing of plastics. For the latter case, the mold master is typically machined into brass by mechanical means; alternatively, photoresist-based structures may also be defined using continuous-relief exposure, required for exactly defining the profiles of the segments. We will explore some of these technologies in Section 11.4. The thin nature of the Fresnel lens implies that it is well suited to fabrication using the thin film technologies of microsystems engineering.

An example micro-Fresnel structure is shown in Figure 7.45; this 2.7 mm diameter lens is fabricated by hot embossing in PMMA, and has a maximum structure height of 88 μm .

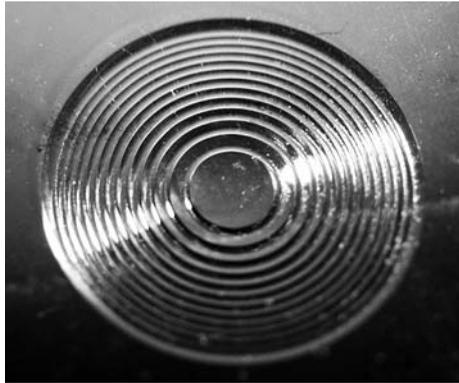


Figure 7.45 A Fresnel microlens fabricated by embossing into PMMA; the lens has a diameter of 2.7 mm and a maximum sag height of 88 μm . Photo courtesy of Andreas Mohr.

Applications

For a given optical power, Fresnel lenses are much thinner than their bulk refractive counterparts, implying that they have reduced optical absorption when compared to normal lenses. As a result, Fresnel lenses are often used for focussing and collimation, particularly for high-power applications, since lower absorption implies less thermal stress on the lens and a higher damage threshold. This feature is also relevant for micro-Fresnel lenses which are frequently used for collimation of emission from laser diodes, which also have a high optical intensity.

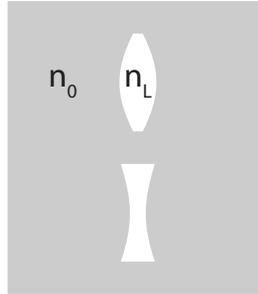
When used for focussing, the convex (“grooved”) side of the Fresnel lens should face the infinite conjugate, to reduce the effects of spherical aberration; this effect is the same as that for standard plano-convex lenses, as we saw in Example 7.10. For the Fresnel lens, in addition, this orientation avoids total internal reflection inside some of the segments of the lens, particularly toward the edge, which would lead to vignetting. An additional consideration for Fresnel lenses is that any off-axis rays may be subject to shadowing (due to the segments) and generate significant coma in an image.

Due to the optical effects of the transitions between the segments, micro-Fresnel lenses are rarely used for imaging. The steps modulate the wavefront and give rise to scattering and stray light, thereby reducing the contrast in any imaging system. Most applications revolve about focussing and collimation, where the large apertures available with low overall sag height, facilitate their use in a variety of integrated microsystems configurations. Fresnel lenses are perhaps a special case of general refractive / reflective microstructures used for manipulating an optical field; there are numerous creative examples of specialized optical surfaces (Rooman et al., 2005). These are used for input / output coupling, beam shaping and generating customized fields of view.

Problems

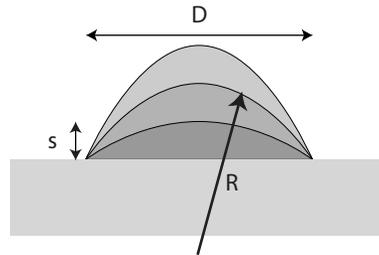
1. Using Fermat’s principle, show that the ideal surface shape for a refractive lens is an ellipse.

2. Given a polycarbonate biconvex lens with $f = 100$ mm in air, determine the image distances S_i for $S_o = \infty, 1000, 200, 150, 100,$ and 50 mm.
3. Repeat the previous problem for the same lens used in water ($n = 1.33$).
4. Imagine the situation where we employ a lens with refractive index $n_L < n_0$; in two dimensions, this might be a lens “cavity” filled with air etched into a plexiglass substrate, as shown in the figure below. How does Equation 7.19 change? Describe the effect of a convex and a concave lens made under these conditions.



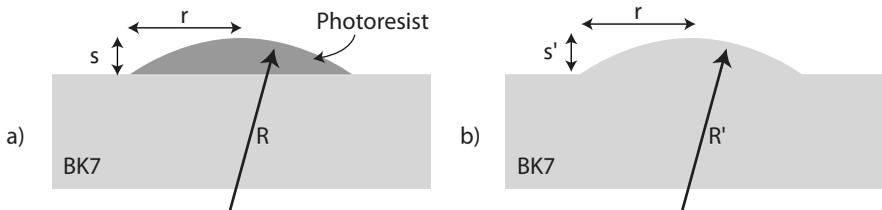
5. For a 5 cm thick biconcave lens ($n_L = 1.52$) with radii of curvature $R_1 = -20$ cm and $R_2 = 10$ cm, describe the image generated by a 2 cm tall object placed 8 cm to the left of the lens. Use the paraxial approximation.
6. Consider a two-lens system, the first (on the left) with $f_1 = 100$ mm and the second with $f_2 = -50$ mm, separated by 70 mm. What is the effective focal length of the system and what optical function does it perform? Calculate the bfl and determine the position of the image plane for an object placed 10 m to the left of the system?
7. A biconvex lens with radii of curvature 5 cm is placed 2 cm to the right of a biconcave lens with radii 6 cm. Determine the position, size and orientation of the image of an object placed 12 cm to the left of the biconcave lens. Assume the thin lenses have $n_L = 1.5$ and the setup is operated in air.
8. Given a glass ($n = 1.52283$) biconvex lens with $f = 10$ cm, measured in air. You use it to image a fish in water ($n = 1.33$), 100 cm from the lens. What is the focal length in this measurement configuration, what kind of image is generated and where, and what is the magnification? The fish is interested in optics and swims closer, to within 20 cm of the lens. Answer the questions about the image again.
9. Show that the Gaussian lens formula results when a Gaussian beam with beam waist much further from the lens than the focal length is transmitted through a lens.
10. Show how Equation 7.58 predicts spherical aberration (change in S_i).
11. Sketch the positions of the red and blue foci that you expect due to chromatic aberration in a concave lens.
12. Derive an expression for the axial chromatic aberration of a plano-convex lens inserted backwards into a collimated beam (i.e., planar side toward the collimated side), as a function of $n_{L(red)}$ and $n_{L(blue)}$ and R_2 . For the resourceful student: calculate and compare numerical values for axial CA using BK7 and LaSFN9 glasses for a lens with $R_2 = -50$ mm. (Yes, you have to find the dispersion data yourself.)

13. Given a microlens technology which uses a $20\ \mu\text{m}$ thick surface layer with $n = 1.64$, calculate the radius of curvature for lenses fabricated with aperture diameters of 200, 500 and $1\ 000\ \mu\text{m}$. Estimate the focal length.
14. The photoresist reflow technique uses the surface tension-defined sphere (or portion thereof) to form a microlens on a substrate (we will discuss this fabrication process in detail in Section 11.2.1). For a given diameter (aperture) on the surface, D , increasing the volume of the droplet gives rise to a larger sphere, with larger sag height s , but with a smaller radius of curvature, R , as shown in the figure.



Given a surface layer with $n_L = 1.64$ and a fixed diameter $D = 100\ \mu\text{m}$, calculate the possible range of radii of curvature R achievable, determine the minimum possible R and the maximum possible NA. Plot R , f and NA as a function of sag height.

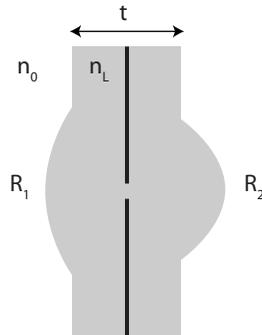
15. You would like to fabricate a microlens with $f = 500\ \mu\text{m}$ and a diameter of $2r = 100\ \mu\text{m}$ optimized for red light ($\lambda = 565\ \text{nm}$). You have two technologies available, as shown in the figure. Both are fabricated on a BK7 glass substrate, whereby lens (a) is made from photoresist ($n = 1.6$) and lens (b) from BK7 itself.



For each approach, calculate the required lens radius of curvature and the sag height.

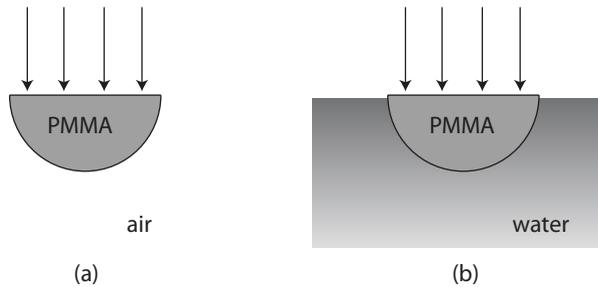
16. For the BK7 microlens given in the previous problem, repeat the design for green rather than red illumination. You know that, for BK7, $n = 1.52283$ for $\lambda = 486\ \text{nm}$ and $n = 1.51432$ for $\lambda = 656\ \text{nm}$. Determine the chromatic aberration.
17. You have an embossed glass microlens array where the lenses of $1\ \text{mm}$ diameter have a radius of curvature of $750\ \mu\text{m}$ and the planar substrate is $300\ \mu\text{m}$ thick. From the data sheet, you know that $n_C = 1.796$ and $n_F = 1.830$. Illuminating the array from the *planar* side with collimated light, calculate the bfl for red light. Now turn the ray around so the *convex* side faces the collimated beam and calculate the bfl again. Repeat the calculations for blue light and determine the CA.
18. Consider the dual-sided microlens structure shown in the figure below. The lenses, with radii of curvature $R_1 = 1\ \text{mm}$ and $R_2 = 0.5\ \text{mm}$, are made from the same

material as the substrate, of thickness $d = 3$ mm, with $n_L = 1.5$. The black structure at the center is a stop.



What is the position of the image created due to an object positioned 10 mm to the left of the lens system? What is the function of the stop?

19. A PMMA ball lens is cut in half and used as a biconvex lens, as shown in the figure. You know that $n_{PMMA} = 1.49$ at 800 nm and also recall that $n_{air} = 1.0$ and $n_{water} = 1.33$.



Determine the focal length for the PMMA lens for use in air at $\lambda = 800$ nm, as in (a). Calculate $f/\#$. Then repeat the calculations for using the lens in water, as in (b). Estimate the numerical aperture for both situations. Why is it difficult to specify NA precisely?

20. For the PMMA lenses of the previous example, you find the following dispersion behavior from the data sheet:

λ [nm]	n
300	1.55
400	1.52
500	1.50
800	1.49

Calculate the lateral chromatic aberration between the infrared and ultraviolet regimes for the use of this lens in air.

21. Given a microlens on a glass substrate, suggest a means to measure its profile *without* using an SEM, an AFM, a white light interferometer, or a stylus profiler.

8 Diffractive micro-optics

*The youth is never free from the danger of its grating interference.*¹

William Godwin, 1797

When a wave strikes an obstacle, the edge of this obstruction affects the further propagation of the field. For obstructions very much larger than the wavelength, the result is a shadow, well known to us from our everyday experience. If we look in detail at the edges of a shadow, or consider the transmission of light around obstructions which have size scales on the order of the wavelength, we see variations in the intensity, due to interference effects at the edge of the opaque region. This phenomenon is known as diffraction and plays a very important role in micro-optics.

For transmission of light through a small aperture, diffraction effects can strongly affect the intensity distribution of the transmitted optical field, and these effects increase in relevance the smaller the apertures become. The definition of a diffraction-limited lens, which we already saw in Section 7.2.3, results from these considerations, and we will examine the physical origins and applications of diffraction in this chapter. As we saw in Section 6.4.3 for micromirrors, and will see again for microlenses below, diffraction can limit performance in micro-optical systems, but is more often a source of optical functionality of great utility for a wide variety of novel components (Turunen and Wyrowski, 1997; Kress and Meyrueis, 2000).

We will begin with an overview of diffraction and its manifestation in a number of simple configurations, including rectangular and circular apertures. The diffraction grating, probably the most established and widely employed diffractive component, is then the subject of a detailed examination, followed by consideration of diffractive lenses; both of these components have no macroscopic equivalent and are thus classical micro-optical elements. We will then consider Fresnel-zone plates, diffractive/refractive hybrid lenses, diffractive optical elements and, as case studies, spectrometers and wavelength-selective mirrors, demonstrating the utility of this branch of optics.

¹ William Godwin (1756–1836), English philosopher and anarchist, in (Godwin, 1797), p. 68.

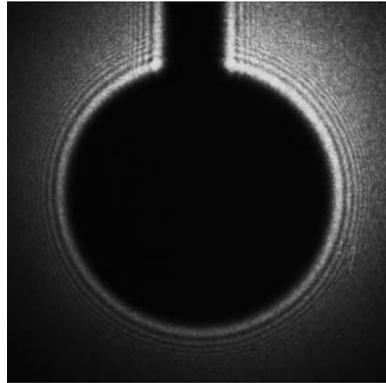


Figure 8.1 Photograph of diffraction around the head of a pin; illumination is by a green HeNe laser emitting at 543 nm. The interference patterns are clearly seen at the edge of the pinhead. Image courtesy of Bernd Aatz.

8.1 Diffraction

Diffraction results from interference between waves which encounter an obstruction in the optical path, and is of particular interest for apertures and obstacles with size ranges on the order of several to several hundred wavelengths. The optical intensity distribution generated by the interaction of an incident optical wave with an aperture is termed the diffraction pattern, an example of which is seen in the interference fringes generated at the edge of the shadow of a pinhead in Figure 8.1. It is the analysis and/or generation of diffraction patterns which forms the focus for the field of diffractive optics (Goodman, 2004). We will thus begin by examining diffraction for a number of simple cases which will prove to be useful building blocks for more complex diffractive systems.

8.1.1 Huygens' principle

As we saw in our historical introduction of Chapter 1, Christiaan Huygens² developed a useful picture to describe wave propagation: every point on a propagating wave at a given time may be considered to be a point source emitting a “wavelet”, and propagation is defined by the sum of these wavelets which form a new wave at the next instant of time. This model becomes very useful if we consider what happens at an aperture, as shown in Figure 8.2. An unobstructed plane wave is the sum of an infinite number of point sources on an infinitely extended wavefront and these sum to generate a following planar wavefront, as seen on the left of the figure.

If we place an obstruction in the field, as shown in the right of Figure 8.2, only a limited number of point sources are transmitted, so that the aperture in this example acts

² Christiaan Huygens (1629–1695), Dutch polymath, active in Holland and France, discovered the moons of Saturn using his own lenses, invented pendulum clocks, proposed a wave theory of light, and developed a 31-tone musical system.

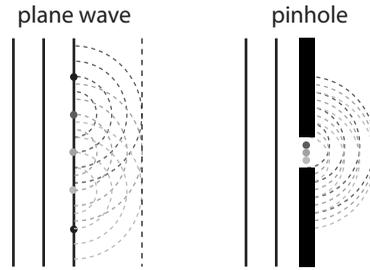


Figure 8.2 Huygens' principle defines propagation of a wave as the sum of “wavelets” emitted from point sources on a wavefront. A plane wave *left* is the sum of an infinite number of these point sources; a pinhole aperture *right* has only a limited number of point sources, such that it acts more like a slightly extended point source.

more like a slightly-extended point source, or pinhole. We see from the schematically sketched wavefronts that the result is approximately a spherical wave, as we would expect for a point source. This approach may be generalized to arbitrary apertures (or obstructions) placed into an optical field: the transmitted field may be calculated as the sum of point sources distributed across the transparent part of the aperture. We will use Huygens' principle to describe diffraction from a number of apertures.

8.1.2 Near- and far-field

As we will see presently, diffraction is strongly wavelength-dependent: the relative magnitudes of aperture size and wavelength define the nature and strength of the effect. For aperture dimension a [m], where $a \gg \lambda$, the aperture essentially creates a shadow and diffraction effects may be negligible. If, on the other hand, $a \ll \lambda$, quantum mechanical or other effects, such as plasmons, play a dominate role in defining transmission; we address some of these topics in Chapter 14.

In between, for $a \approx \lambda$, diffraction dominates and the nature of the diffracted field depends on the wavelength, aperture size and distance to the point of observation. The near-field, or Fresnel, regime results for the image close to the aperture, whereas the far-field, or Fraunhofer, region is obtained for large spacings between aperture and imaging plane. The far-field (Fraunhofer) regime is obtained when the observation distance R [m] satisfies

$$R > \frac{a^2}{\lambda}. \quad (8.1)$$

In the near-field, the diffraction pattern varies strongly with distance, changing form as the observation plane is moved from the aperture. In contrast, the far-field diffraction pattern has a stable form as the image plane moves, and only changes in size.

Example 8.1: For illumination with a red HeNe laser ($\lambda = 633$ nm), at what distance from the aperture does the transition from the Fresnel to the Fraunhofer regime take

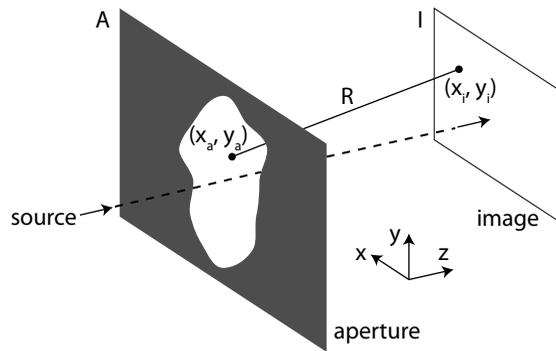


Figure 8.3 Diffraction due to illumination of an aperture A generates a pattern on the image plane I . Huygens' principle implies that each point (x_a, y_a) in the aperture acts as a point source and the intensity distribution at each point on the image plane, (x_i, y_i) , is found from the sum of all the sources in the aperture. The distance R is that between any pair of points, (x_a, y_a) and (x_i, y_i) .

place for a microlens with a diameter of 1 mm? How does it compare with a macroscopic lens with 2.5 cm aperture?

We can simply apply Equation 8.1 to find that the transition distance is about 1.6 mm for the microlens, but almost 1 m for the large lens. Stable far-field diffraction effects thus require considerable spacings for macroscopic optical systems.

In general, most optical applications take advantage of the diffraction far-field, or Fraunhofer, regime. Luckily, much of the mathematical description is simplified for this case, such that calculations of far-field diffraction patterns can often be done by hand for simple apertures.

8.1.3 Fresnel-Kirchhoff formula

To determine the diffraction pattern generated by an arbitrary structure, we can apply Huygens' principle to the open aperture: each point on the aperture area is considered to be a point source. Considering the arrangement in Figure 8.3, we have an arbitrarily shaped transparent aperture A which generates a diffraction pattern on an image plane I (which may in general also be a curved surface). The position of any point on the aperture is given by the coordinates (x_a, y_a) and on the image by (x_i, y_i) , where these are separated by a distance R ; the z axis is in the propagation direction, roughly parallel to R .

Recalling that the field strength of a spherical wave decreases as $1/r$ with distance (Kraus and Carver, 1992, Chapter 14.3), the differential electric field strength dE_i [V/m] at (x_i, y_i) due to a differential point source dE_a [V/m] at (x_a, y_a) is given by

$$dE_i(x_i, y_i) \propto \frac{1}{R} e^{jkR} dE_a(x_a, y_a) \quad (8.2)$$

where as usual $k = 2\pi/\lambda_0$, assuming propagation in free space. The spacing R between (x_a, y_a) and (x_i, y_i) may be expressed in Cartesian coordinates as

$$R = [(x_a - x_i)^2 + (y_a - y_i)^2 + z_i^2]^{1/2} \quad (8.3)$$

for the aperture at $z = 0$.

The total field at the image plane is then the integral of these point sources over the aperture A , yielding

$$E_i(x_i, y_i) = \frac{1}{j\lambda} \iint_A E_a(x_a, y_a) \frac{e^{jkR}}{R} dx_a dy_a. \quad (8.4)$$

This previous expression is known as the Fresnel-Kirchhoff formula³ for diffraction and will prove to be very useful for determining diffraction patterns from arbitrary apertures. Since an arbitrary field has an arbitrary amplitude E_0 [V/m] and phase ϕ [rad], the electric field of the point source at any position (x_a, y_a) takes the general form

$$E_a(x_a, y_a) = E_0(x_a, y_a) e^{j\phi(x_a, y_a)} \quad (8.5)$$

such that the Fresnel-Kirchhoff formula may be written as

$$E_i(x_i, y_i) = \frac{1}{j\lambda} \iint_A E_0(x_a, y_a) e^{j\phi(x_a, y_a)} \frac{e^{jkR}}{R} dx_a dy_a. \quad (8.6)$$

From this relationship, we see that the total field at any position in the image plane may be determined from the integral of all the effective point emitters in the aperture; the intensity and phase of the emission at any point in the aperture is explicitly included, such that the previous expression is a general formulation of the Fresnel-Kirchhoff integral.

Fraunhofer regime

We can simplify the form of the Fresnel-Kirchhoff integral and at the same time explicitly include the Cartesian coordinates by considering the Fraunhofer, or far-field, regime (i.e., when Equation 8.1 is satisfied). In this case

$$z_i^2 \gg (x_a - x_i)^2 + (y_a - y_i)^2 \quad (8.7)$$

so that the binomial expansion yields

$$R \approx z_i + \frac{x_i^2 + y_i^2}{2z_i} - \frac{x_i x_a + y_i y_a}{z_i}. \quad (8.8)$$

If we assume a constant intensity and phase over the aperture, the Fresnel-Kirchhoff integral then becomes

$$E_i(x_i, y_i) = E'_0 \iint_{-\infty}^{\infty} E_0(x_a, y_a) e^{-j(k_x x_a + k_y y_a)} dx_a dy_a \quad (8.9)$$

³ Gustav Kirchhoff (1824–1887), a Prussian physicist, is an icon for electrical engineers on account of his laws describing current flow, but he was also the first to define black-body radiation as well as explain the origin of the dark absorption lines in the sun's spectrum.

where we have defined the pre-factor E'_0 [V/m²] as

$$E'_0 = \frac{E_0}{j\lambda} e^{jk \left(z_i + \frac{x_i^2 + y_i^2}{2z_i} \right)}, \quad (8.10)$$

and the propagation constants (with units [m⁻¹]) as

$$k_x = \frac{kx_i}{z_i} \quad (8.11)$$

$$k_y = \frac{ky_i}{z_i}. \quad (8.12)$$

The integrals in Equation 8.9 can be taken from $-\infty$ to $+\infty$ since only the open region of the aperture plane contributes to the total electric field; the remaining areas are opaque so that $E_0 = 0$.

Remembering that $k = 2\pi/\lambda_0$, we realize that Equation 8.9 represents the Fourier transform of the aperture from (x_a, y_a) to (x_i, y_i) . We can thus write quite simply, using the notation FT to represent the Fourier transform,

$$E_i(x_i, y_i) = E'_0 \text{FT} [E_a(x_a, y_a)]. \quad (8.13)$$

The preceding implies that the far-field diffraction intensity pattern due to a given aperture can be calculated by taking the Fourier transform of that aperture. This conclusion will prove to be of considerable utility in the sections which follow.

8.2 Practical apertures

Using the Fresnel-Kirchhoff integral, and the realization that a diffraction pattern corresponds to the Fourier transform of the aperture, we can analyze a few useful cases which frequently correspond to structures encountered in micro-optical systems.

8.2.1 Single slit

Let us consider the simple aperture shown in Figure 8.4: a slit of width a in the x direction, and very much longer in the y direction. The slit is illuminated at a wavelength λ_0 (assuming operation in air) and we assume a constant optical intensity within the aperture. The image plane is shown as a curved surface, somewhat exaggerated in curvature, a distance R from the aperture.

The diffraction pattern in the Fraunhofer (far-field) regime, $E_S(x_i)$, is then found by applying the Fresnel-Kirchhoff integral in the x direction, so that

$$E_S(x_i) = E'_0 \int_{-\infty}^{+\infty} e^{-jk_x x_a} dx_a \quad (8.14)$$

$$= E'_0 \int_{-\frac{a}{2}}^{+\frac{a}{2}} e^{-jk_x x_a} dx_a \quad (8.15)$$

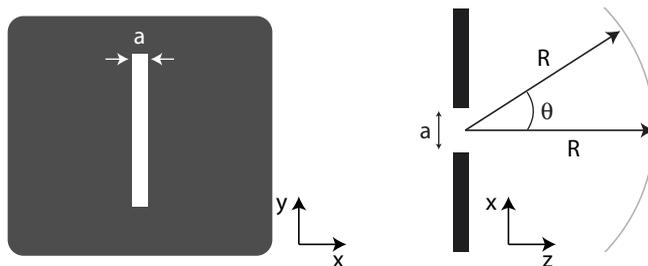


Figure 8.4 A single slit of width a seen from the front *left* and top *right*; the axes correspond to those in Figure 8.3. The image plane is shown as an exaggerated curved surface a distance R from the slit.

where the limits may be taken as $\pm a/2$ since there is no contribution to the integral outside the aperture. This integral evaluates to

$$E_S(x_i) = E'_0 a \frac{\sin \left[\frac{ak_x}{2} \right]}{\left[\frac{ak_x}{2} \right]} \quad (8.16)$$

$$= E'_0 a \operatorname{sinc} \left[\frac{ak_x}{2} \right], \quad (8.17)$$

where we employ the notation $\operatorname{sinc}(\zeta) = \sin(\zeta)/\zeta$.

We may write the argument of the previous expression explicitly as a function of x_i and z_i , since

$$\frac{ak_x}{2} = \frac{\pi a}{\lambda} \frac{x_i}{z_i}. \quad (8.18)$$

Alternatively, since for $z_i \approx R$, θ , the angle in the (\mathbf{x}, \mathbf{z}) plane as shown in Figure 8.4, is given by

$$\sin \theta = \frac{x_i}{z_i}, \quad (8.19)$$

we may also use the practical form

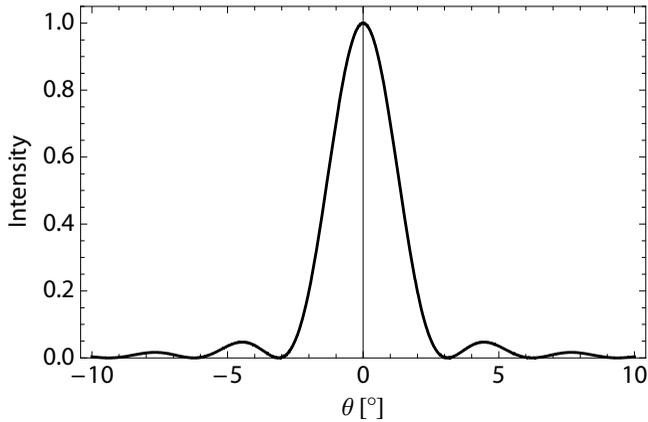
$$E_S(\theta) = E'_0 a \operatorname{sinc} \left[\frac{\pi a}{\lambda} \sin \theta \right]. \quad (8.20)$$

As usual, the measurement quantity is the optical intensity, proportional to the square of the electric field. We may thus finally determine the diffraction intensity pattern, $I_S(\theta)$, to be

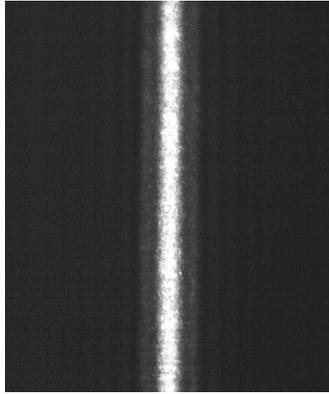
$$I_S(\theta) = I_{S0} \operatorname{sinc}^2 \left[\frac{\pi a}{\lambda} \sin \theta \right] \quad (8.21)$$

where I_{S0} is a constant, but one which varies as the spacing between aperture and image (R or z_i) changes; it is usually normalized to unity at the intensity maximum.

The distribution of the intensity as a function of θ is given in Figure 8.5. As is typical for the sinc function, the central peak seen in Figure 8.5(a) is surrounded by considerably smaller satellite peaks, the first 0.047, the second 0.017 times smaller than the maximum. As can be determined from Equation 8.20, the satellite peaks are found at angles



(a) Calculated diffraction intensity for a $10\ \mu\text{m}$ wide slit as a function of θ , in the x direction, for illumination at $543\ \text{nm}$.



(b) Photograph of the diffraction intensity as a function of x . Image courtesy of Bernd Aatz.

Figure 8.5 Calculated and measured diffraction from a single slit illuminated at $\lambda = 543\ \text{nm}$.

corresponding to $\arcsin(\pm n\lambda/a)$, such that the spacing between peaks decreases as the slit width a increases (or, equivalently, the wavelength decreases for fixed slit width).

The diffraction pattern as imaged is shown in Figure 8.5(b). The broad first satellite peak is clearly seen, and hints of the more distant ones are seen further from the peak. The image has a constant intensity in the vertical (y) direction, since the slit is much larger than the wavelength in that dimension. It is apparent that, for a fixed wavelength, diffraction effects become more prominent as aperture size decreases; equivalently, for a given aperture size, decreasing the wavelength results in more pronounced diffraction. Thus in this case, diffraction is only seen *normal* to the slit direction.

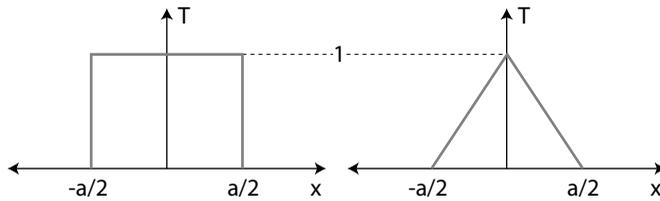


Figure 8.6 Plot of the transmittance T as function of x for a clear rectangular aperture of width a left and one with a graded transmittance right.

Example 8.2: The utility of using the Fourier transform in determining the diffraction from an aperture may be illustrated by analyzing the single slit purely mathematically and comparing the result with that for an aperture with graded transmittance. The two cases are shown in Figure 8.6: both apertures have a width a , in the interval $-a/2 \leq x \leq +a/2$, but the clear aperture has a transmittance of unity in this interval whereas the graded aperture has a linearly varying value, peaking at the center.

The clear aperture may be expressed by a sum of step functions, $v(y)$, as

$$A_C(x) = v\left(x + \frac{a}{2}\right) - v\left(x - \frac{a}{2}\right)$$

where $A = 1$ implies unity transmittance. The Fourier transform of this sum is easily⁴ determined to be

$$\text{FT}[A_C(x)] = a \text{sinc}(ak_x),$$

not surprisingly the same result as in the previous section.

The graded aperture may be represented as a convolution of the sum of step functions with itself, or

$$A_G(x) = A_C(x) * A_C(x),$$

the Fourier transform of which is given by

$$\text{FT}[A_G(x)] = a \text{sinc}^2(ak_x),$$

which we have normalized with a factor $1/a$ to yield a maximum value of unity at $x = 0$. These Fourier transforms give the form of the electric field, and we recall again that the measured intensity is proportional to the square of these values.

The predicted diffraction patterns may thus be found with a few lines of math. By comparing the diffraction characteristics for these two apertures (see Problems), we see that the $\text{sinc}^4(ak_x)$ characteristic for the diffraction intensity of the graded aperture has considerably smaller diffraction maxima next to the primary peak than does the $\text{sinc}^2(ak_x)$ characteristic of the clear aperture. Configuring the aperture to remove

⁴ For the student irritated by this adverb: this and the following Fourier transform represent some of the first examples found in most textbooks on the subject. See for example (Råde and Westergren, 1999, Chapter 13) or (Papoulis, 1962, Chapter 4).

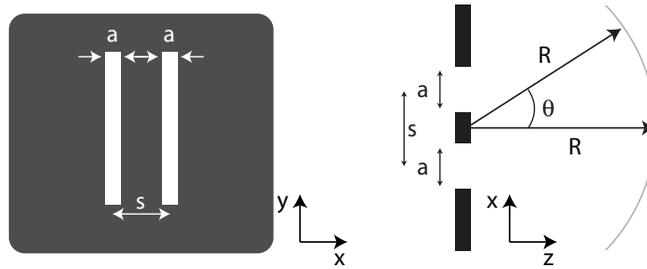


Figure 8.7 Two slits of width a separated by s (center-to-center) seen from the front *left* and top *right*; the axes correspond to those in Figure 8.3. The image plane is again the exaggerated curved surface a distance R from the slit.

these neighboring peaks is an example of apodization, which, despite meaning “foot removal”, does not refer to the awful fate of Cinderella’s step-sisters, but instead to the use of a filter by which the ringing or side-lobes found at the edge of an optical (or electrical) signal are reduced.

8.2.2 Dual slit

We now consider two slits of the form just discussed, spaced a distance s (center-to-center) from each other, as shown in Figure 8.7. This configuration is exactly that of the famous double-slit experiment with which Thomas Young added irrefutable support to the wave theory of light in 1801. The geometry of the system is the same as in Section 8.2.1, save that we now need to add two fields, one emitted from each slit. We take the origin ($x = 0$) to be positioned at the center of the lower slit in the top view of Figure 8.7.

Since the electric fields add, the diffraction pattern is given by the sum of the integrals

$$E_D(x_i) = E'_0 \int_{-\frac{a}{2}}^{+\frac{a}{2}} e^{-jk_x x_a} dx_a + E'_0 \int_{-\frac{a}{2}+s}^{+\frac{a}{2}+s} e^{-jk_x x_a} dx_a \quad (8.22)$$

assuming that the intensities in the two slits are identical and homogeneous. Using the same argumentation as above, these integrals again represent the Fourier transform, albeit with differing limits of integration, and the sum evaluates to

$$E_D(x_i) = 2E'_0 a \operatorname{sinc} \left[\frac{ak_x}{2} \right] \cos \left(\frac{sk_x}{2} \right) e^{-j\frac{k_x s}{2}} \quad (8.23)$$

$$= E'_D(z_i) a \operatorname{sinc} \left[\frac{ak_x}{2} \right] \cos \left(\frac{sk_x}{2} \right) \quad (8.24)$$

where we have defined a new constant pre-factor,

$$E'_D = 2E'_0 e^{-j\frac{1}{2}k_x s}. \quad (8.25)$$

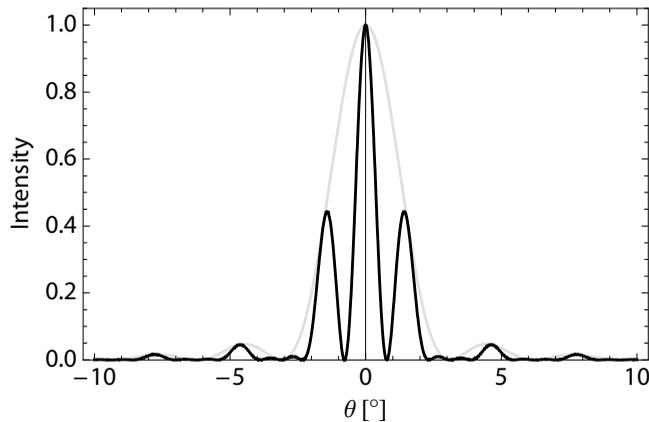


Figure 8.8 Diffraction from two slits with width $a = 10 \mu\text{m}$ and spacing $s = 20 \mu\text{m}$ illuminated with $\lambda = 543 \text{ nm}$. The light gray curve shows the envelope formed by the diffraction from a single slit, Figure 8.5(a).

Explicitly writing the variation with angle θ , we obtain

$$E_D(\theta) = E'_D a \operatorname{sinc} \left[\frac{a\pi}{\lambda} \sin \theta \right] \cos \left(\frac{s\pi}{\lambda} \sin \theta \right) \quad (8.26)$$

and considering the intensity, we have, analogously to the single-slit case,

$$I_D(\theta) = I_{D0} \operatorname{sinc}^2 \left[\frac{a\pi}{\lambda} \sin \theta \right] \cos^2 \left(\frac{s\pi}{\lambda} \sin \theta \right). \quad (8.27)$$

The dual-slit intensity diffraction pattern is then that shown in Figure 8.8. The diffraction from a single slit forms an envelope for this pattern, as shown by the light gray curve. It is clearly seen that the interference between the fields emitted from the two slits leads to a diffraction pattern with satellite peaks of variable intensity. These peaks are found at angular positions $\arcsin(\pm n\lambda/s)$, such that increasing the spacing between the slits, s , (for fixed wavelength) decreases the peak spacing.

Example 8.3: If two slits already gives rise to the rich set of diffraction peaks seen in Figure 8.8, what happens if add more slits, until we have N of them?

As in the previous case, we can add the fields from the individual slits to determine the overall diffraction pattern; we now have a more complicated summation. Taking the origin again to be at the center of the lowest slit, we have

$$\begin{aligned} E_N(x_i) = & E'_0 \int_{-\frac{a}{2}}^{+\frac{a}{2}} e^{-jk_x x_a} dx_a + E'_0 \int_{-\frac{a}{2}+s}^{+\frac{a}{2}+s} e^{-jk_x x_a} dx_a \\ & \dots + E'_0 \int_{-\frac{a}{2}+(N-1)s}^{+\frac{a}{2}+(N-1)s} e^{-jk_x x_a} dx_a. \end{aligned}$$

which we may neatly express as the sum

$$E_N(x_i) = E'_0 \sum_{i=0}^{N-1} \int_{-\frac{a}{2}+is}^{+\frac{a}{2}+is} e^{-jk_x x_a} dx_a.$$

We realize that the integral again has the same form as Equation 8.15, which, when considering the slightly more complicated integration limits, leads to

$$\int_{-\frac{a}{2}+is}^{+\frac{a}{2}+is} e^{-jk_x x_a} dx_a = 2e^{jk_x s} \frac{\sin\left(\frac{ak_x}{2}\right)}{k_x}$$

and if we apply the summation to the previous result, we obtain

$$E_N(x_i) = E'_0 a \frac{e^{jNk_x s} - 1}{e^{jk_x s} - 1} \operatorname{sinc}\left[\frac{ak_x}{2}\right].$$

This last expression can be manipulated into a more useful form by realizing that

$$\begin{aligned} \frac{e^{jNk_x s} - 1}{e^{jk_x s} - 1} &= \frac{e^{jN\frac{k_x s}{2}}}{e^{j\frac{k_x s}{2}}} \left(\frac{e^{jN\frac{k_x s}{2}} - e^{-jN\frac{k_x s}{2}}}{e^{j\frac{k_x s}{2}} - e^{-j\frac{k_x s}{2}}} \right) \\ &= e^{j(N-1)\frac{k_x s}{2}} \frac{\sin\left(N\frac{k_x s}{2}\right)}{\sin\left(\frac{k_x s}{2}\right)} \end{aligned}$$

such that we can write the total field as

$$E_N(x_i) = E'_N a \operatorname{sinc}\left[\frac{ak_x}{2}\right] \frac{\sin\left(N\frac{k_x s}{2}\right)}{\sin\left(\frac{k_x s}{2}\right)}$$

where we define yet another pre-factor,

$$E'_N = E'_0 e^{j(N-1)\frac{k_x s}{2}}.$$

Finally, writing the field as a function of angle, we obtain

$$E_N(x_i) = E'_N a \operatorname{sinc}\left[\frac{a\pi}{\lambda} \sin\theta\right] \frac{\sin\left(N\frac{s\pi}{\lambda} \sin\theta\right)}{\sin\left(\frac{s\pi}{\lambda} \sin\theta\right)}$$

and we recall that the intensity is proportional to the square of this last expression.

Figure 8.9 shows an example of the diffraction pattern generated by multiple slits, in this case five 10 μm slits spaced 20 μm apart. The distinct maximum at $\theta = 0$ is surrounded by a number of satellite peaks of varying intensity. The positions of these peaks are given by the width and spacing of the slits; increasing the number of slits decreases the magnitude of the modulation between the larger peaks, but does not cause a shift in the peak position. A large value for N , then, is useful to generate a set of sharply-defined peaks with little intensity in the off-peak positions. We will re-visit this issue when we discuss gratings in Section 8.3 below.

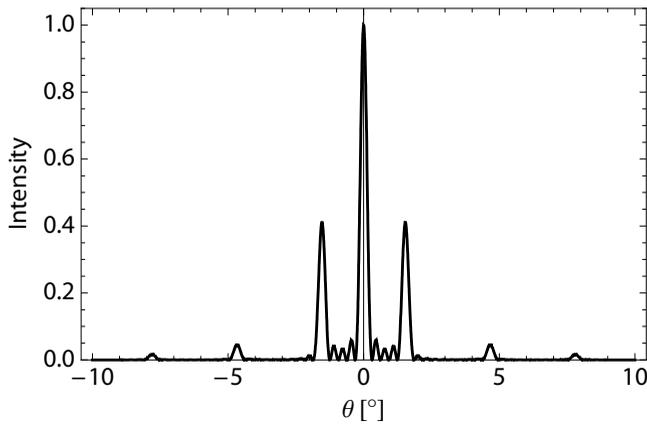


Figure 8.9 The diffraction pattern generated by five slits with width $a = 10 \mu\text{m}$ and spacing $s = 20 \mu\text{m}$ illuminated with $\lambda = 543 \text{ nm}$.

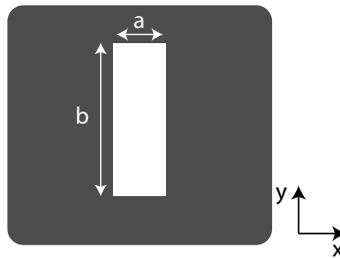


Figure 8.10 A rectangular aperture with both dimensions of the same order of magnitude, with width a in the x direction and width b in the y direction.

8.2.3 Rectangular aperture

The previous examples considered “slits”, apertures much longer in one direction than in the other, so that diffraction was only appreciable in one (the x) dimension and negligible in the other. We now consider a rectangular aperture, such as that shown in Figure 8.10, where both dimensions are of similar magnitude.

The procedure to determine the diffraction pattern is the same, save that we now must return to a two-dimensional surface integral over x and y . We thus have

$$E(x_i, y_i) = E'_0 \iint e^{-j(k_x x_a + k_y y_a)} dx_a dy_a \tag{8.28}$$

$$= E'_0 \iint e^{-jk_x x_a} e^{-jk_y y_a} dx_a dy_a \tag{8.29}$$

which evaluates to the product of the Fourier transforms in the two directions. The diffraction pattern due to a rectangular aperture, E_R , is therefore

$$E_R(x_i, y_i) = E'_0 ab \operatorname{sinc} \left[\frac{ak_x}{2} \right] \operatorname{sinc} \left[\frac{bk_y}{2} \right] \tag{8.30}$$

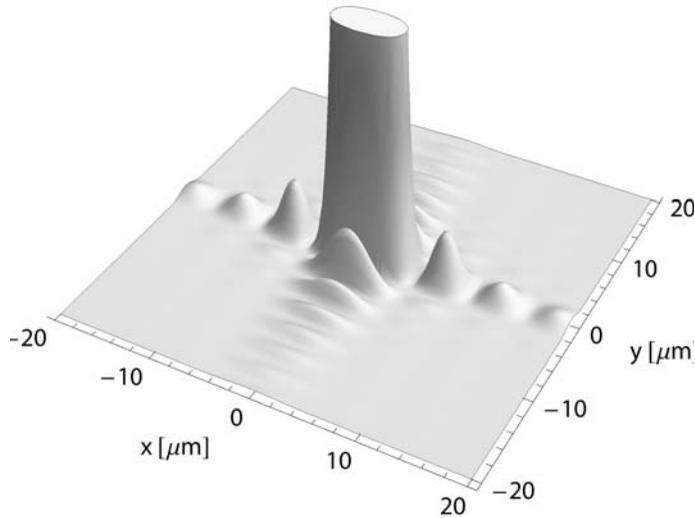


Figure 8.11 Calculated diffraction pattern for a centered rectangular aperture illuminated at $\lambda = 543 \mu\text{m}$; aperture dimensions are $a \times b = 5 \times 10 \mu\text{m}^2$. The main peak has been decapitated to show the diffraction features more clearly.

or, as a function of angle,

$$E_R(\theta) = E'_0 ab \operatorname{sinc} \left[\frac{a\pi}{\lambda} \sin \theta_x \right] \operatorname{sinc} \left[\frac{b\pi}{\lambda} \sin \theta_y \right] \quad (8.31)$$

for θ_x and θ_y the angles in the \mathbf{xz} and \mathbf{yz} planes, respectively. We saw an identical result for rectangular micromirrors in Section 6.4.3, and the diffraction pattern generated by a reflector of small dimension is the same as that generated by a clear aperture of the same size.

As is shown in the calculated diffraction pattern of Figure 8.11, the diffraction features can be clearly seen on both sides of the aperture. Recall that this is the intensity of a diffraction pattern, not a shadow; for that reason, the peak is wider in the direction of the *smaller* part of the aperture, in this case the x direction with $a = 5 \mu\text{m}$. The spacing between the satellite peaks is also larger in this dimension, as can be seen by directly comparing the spacing and shape of the peaks in the x and y directions.

Diffraction from a rectangular aperture neatly summarizes some of the important basics of diffraction. As we mentioned previously and saw in the examples above, diffraction effects become stronger as aperture dimensions decrease and the diffraction peaks are found in the direction of the restricted aperture. The onset of diffraction in an optical system, then, is always determined by the smallest feature size.

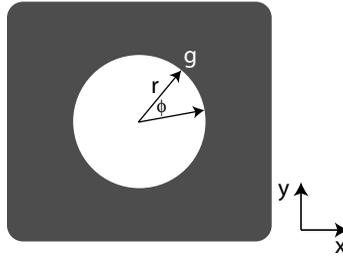


Figure 8.12 A circular aperture with a radius g .

8.2.4 Circular aperture

Most microlenses are round and thus the diffraction which results from a circular aperture is often of interest. Considering the structure of Figure 8.12, a cylindrical coordinate system using variables r and ϕ is advantageous for evaluating the diffraction integral. We may thus either evaluate the Fresnel-Kirchhoff integral over the circular aperture directly (Hecht, 2002, Chapter 10.2.5) or consider the Fourier transform in cylindrical coordinates (Cormack, 1957).

Taking the latter approach, for a planar aperture (i.e., no z -dependence of the field in the aperture), the Fourier transform in cylindrical coordinates is then a transformation from (r_a, ϕ_a) , the aperture, to (r_i, ϕ_i) , the image. For an aperture function $A(r_a, \phi_a)$, its Fourier transform $A(r_i, \phi_i)$ may be written as

$$A(r_i, \phi_i) = \int_0^\infty \int_0^{2\pi} r_a A(r_a, \phi_a) e^{-j2\pi r_i r_a \cos(\phi_a - \phi_i)} d\phi_a dr_a. \quad (8.32)$$

The function $A(r_a, \phi_a)$, since it is periodic in ϕ_a with period 2π , may be expanded into a Fourier series,

$$A(r_a, \phi_a) = \sum_{n=-\infty}^{\infty} A_n(r_a) e^{jn\phi_a} \quad (8.33)$$

with

$$A_n(r_a) = \frac{1}{2\pi} \int_0^{2\pi} A(r_a, \phi_a) e^{-jn\phi_a} d\phi_a. \quad (8.34)$$

If we now apply this to the special case of a clear circular aperture with radius g , as shown in Figure 8.12, where $A(r_a, \phi_a) = 1$ for $r_a \leq g$ and $A(r_a, \phi_a) = 0$ for $r_a > g$, the Fourier transform becomes (Cormack, 1957)

$$A(r_i, \phi_i) = 2\pi \int_0^g r_a J_n(k_r r_a) dr_a \quad (8.35)$$

where

$$k_r = \frac{kr_i}{z_i} = \frac{2\pi r_i}{\lambda z_i} \quad (8.36)$$

and the function

$$J_n(\zeta) = \frac{J^{-n}}{2\pi} \int_0^{2\pi} e^{J(n\zeta' + \zeta \cos \zeta')} d\zeta' \quad (8.37)$$

is a Bessel function of the first kind of order n . For the circularly symmetric case we are considering here (no dependence of the aperture characteristics on ϕ_a), all terms with $n \neq 0$ in Equation 8.37 are zero, so that we are left with a zero-order Bessel function and the Fourier transform simplifies to

$$A(r_i) = 2\pi \int_0^g r_a J_0(k_r r_a) dr_a. \quad (8.38)$$

We may thus finally write the electric field distribution of the diffraction in the image plane due to the clear circular aperture as

$$E(r_i) = 2\pi E'_0 \int_0^g r_a J_0(k_r r_a) dr_a. \quad (8.39)$$

To evaluate this integral, we may employ the recurrence relation for Bessel functions, which states that

$$\frac{d}{d\zeta} [\zeta^m J_m(\zeta)] = \zeta^m J_{m-1}(\zeta) \quad (8.40)$$

or equivalently, when applying this relation to the integral above,

$$\int \zeta J_0(\zeta) d\zeta = \zeta J_1(\zeta). \quad (8.41)$$

The diffraction field thus takes the form

$$E(r_i) = 2\pi E'_0 \frac{g}{r_i} J_1(k_r r_i) \quad (8.42)$$

which we can write more practically as

$$E(r_i) = E'_{C0} \frac{J_1(k_r r_i)}{k_r r_i} \quad (8.43)$$

with a slightly revised constant pre-factor E'_{C0} . Recalling the definition of k_r , we may again, as for the previous cases in which we used Cartesian coordinates, write the diffraction field as a function of angle θ , namely

$$E(\theta) = E'_{C0} \frac{J_1(kg \sin \theta)}{(kg \sin \theta)}. \quad (8.44)$$

We recall that the resultant optical intensity is then proportional to the square of the last expression, so that

$$I_C(\theta) = I_{C0} \left[\frac{J_1(kg \sin \theta)}{(kg \sin \theta)} \right]^2. \quad (8.45)$$

A one-dimensional cross-section of the characteristic of Equation 8.45 is shown in Figure 8.13. When compared to the diffraction cross-section of a 10 μm rectangular slit, also shown in the figure, we see that the diffraction peak for the circular aperture is sharper, decreases from its central peak more rapidly and has smaller satellite maxima.

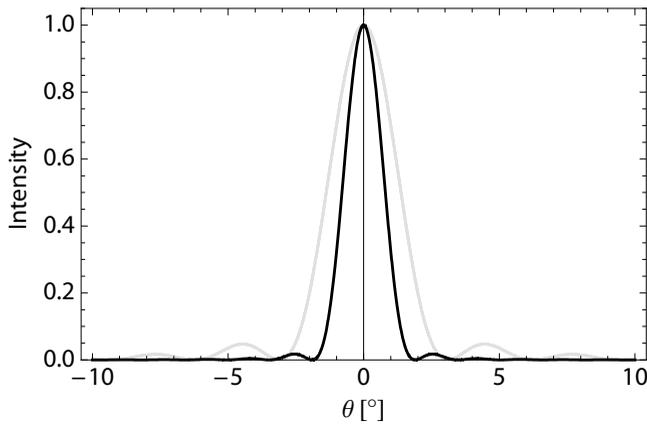
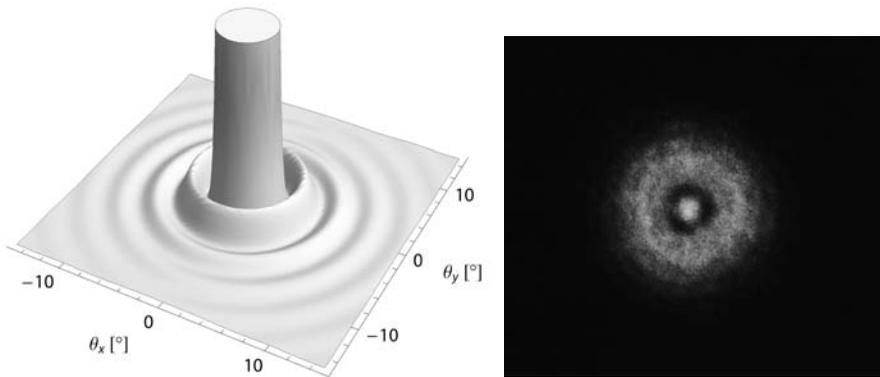


Figure 8.13 The diffraction intensity from a 10 μm radius circular aperture illuminated at 543 nm as a function of angle in the image plane is shown by the black curve. For comparison, the gray curve traces the diffraction intensity of a 10 μm wide one-dimensional slit (Equation 8.20).



(a) Calculated diffraction intensity for a circular aperture with 10 μm radius. The central peak has been decapitated to show the details of the neighboring maxima. (b) Photograph of the diffraction intensity from a circular aperture of 1 mm diameter. Image courtesy of Bernd Aatz.

Figure 8.14 Diffraction from a circular aperture illuminated at $\lambda = 543$ nm.

Figure 8.14 shows a detailed view of the satellite maxima and a measured so-called Airy disk. We have seen this function before: the diffraction pattern for reflection from a circular mirror has the same form, as we saw in Figure 6.15 of Section 6.4.3. The same diffraction effects result if we reflect from or transmit through a given aperture.

Airy disk

The central peak of the circular diffraction pattern, as seen in Figure 8.14(b), is surrounded by a dark ring; a more detailed characteristic would show a series of bright and dark rings, as calculated in Figure 8.14(a). The central bright region is termed the Airy disk and the half-width of this disk (the radial distance from the center to the middle of the first dark ring), r_{Airy} [m], is given by

$$r_{Airy} = 1.22 \frac{R\lambda}{2g} \quad (8.46)$$

for spacing between aperture and image R , corresponding to the first zero of the Bessel function. The last expression may be generalized for the case where the circular aperture is a lens, such that for the image at a distance equal to the focal length f , the Airy disk radius may be written as

$$r_{Airy} = 1.22 \frac{f\lambda}{2g} = 1.22 \frac{f\lambda}{D} \quad (8.47)$$

for lens aperture diameter D .

For a circular diffraction pattern of this form, 84% of the total optical intensity is found within the Airy disk; 91% is found within the disk and the first bright ring. When a lens is thus used to generate a focused spot, the intensity distribution, due to diffraction which is always present, will be that of an Airy disk. Most of the intensity will be in the central disk, but its width and the prominence of the satellite rings will be a function of the relative size of the aperture and the wavelength.

The Airy disk pattern always results in optical systems with a restricted circular aperture, and we saw in Section 7.2.3 that the diffraction (or Rayleigh) limit for a refractive lens is reached when the intensity distribution at the focus takes on this form and all other aberrations are negligible in comparison. At the diffraction limit, the central Airy disk thus contains 84% of the total optical energy and we also recall that the Strehl ratio defines the relationship between the measured height of the Airy function and its theoretical maximum.

Example 8.4: The size of the Airy disk generated due to diffraction from a circular aperture defines a limit for the resolution of optical systems. As we briefly mentioned in Section 7.2.3, two focussed spots may be considered distinguishable if the maximum (Airy disk peak) of one is spaced farther than the first diffraction pattern minimum (the first dark ring) of the other. Conversely, an optical imaging system can distinguish between two points if the diffraction patterns of their images are spaced farther than this value.

Using the relationships we derived above, the resolution, Δr_{min} [m], of an optical system is thus given by

$$\Delta r_{min} = 1.22 \frac{f\lambda}{D}$$

Table 8.1. Resolution (in μm) of lenses used for optical storage, considering solely diffraction, for various aperture diameters. Calculated for the CD, DVD and Blu-ray disc wavelengths, for a lens with $f = 0.75$ mm.

aperture [mm]	CD ($\lambda = 780$ nm)	DVD ($\lambda = 650$ nm)	BD ($\lambda = 405$ nm)
0.1	7.14	0.28	0.14
2.5	5.94	0.24	0.12
5.0	3.71	0.15	0.07

where D represents the aperture diameter and f the focal length of the lens or lens system; we may alternatively, as in Equation 7.33, express this relationship as

$$\Delta r_{min} = 1.22 \frac{\lambda}{NA}$$

for numerical aperture NA . High resolution is thus achieved with short wavelengths and focal lengths, and large apertures. The diffraction limit is a fundamental one: improving the (refractive) quality of a lens that is already diffraction limited will not improve the resolution.

From the above relationships, we see why high-resolution optical systems (such as lithography lenses) use short wavelengths and have such massive diameters. Let us compare some typical values from the realm of optical data storage (Song et al., 2007): using the laser wavelengths employed by standard CD, DVD and Blu-ray discs (BD), 780 nm, 650 nm and 405 nm, respectively, we can estimate the resolution due solely to diffraction by applying the above expression for a focal length of 0.75 mm and a variety of aperture diameters.

Summarizing these calculations in Table 8.1, we see that resolution scales directly with wavelength, explaining why considerable industrial effort was expended to develop a shorter wavelength optical storage system: higher resolution means that smaller features are readable and thus more data may be packed onto the 12 cm polycarbonate disc. Increasing the lens aperture diameter would also help to improve resolution, but a larger lens is heavier, moves more slowly for a given actuation force, and may thus limit access speed in an optical data storage system. The typical Blu-ray disc system lens has a diameter of about 2.5 mm and $NA = 0.85$, compared with $NA = 0.60$ for standard DVD optics.

8.3 Gratings

One of the most useful structures in diffractive micro-optics is the diffraction grating. Consisting of a periodic series of reflecting surfaces or transmissive slits, the grating

results in a diffraction pattern of sharp peaks, whose angular position is a function of the wavelength and the grating period. As we saw in the historical introduction of Chapter 1, the diffraction grating is likely the oldest micro-optical component, first analyzed by Joseph von Fraunhofer, and gratings were originally fabricated using the massive and complex ruling engines developed, among others, by Albert Michelson.

Frequently fabricated today using the holographic and lithographic techniques we will discuss in Chapter 11, and manufactured in large quantities using replication techniques, gratings are now one of the most widely employed micro-optical components, predominantly used as dispersive elements, i.e., for separating a spectrum into its discrete wavelengths. The literature is correspondingly extensive; see for example (Hutley, 1982) or (Loewen and Popov, 1997).

8.3.1 Grating functionality

The functionality of a diffraction grating is shown schematically in Figure 8.15. In the example of Figure 8.15(a), a transparent substrate has been patterned with a regular corrugation of period⁵ Λ [m]. We show a rectangular grating profile here, although a sinusoidal, triangular or virtually any other shape will work as well, albeit with varying efficiency. The broad incident beam illuminates a significant number (typically ten or more) of grating periods and the diffraction from each grating period is summed to an overall diffraction pattern leading a number of discrete intensity maxima at precisely defined angles to the incident beam and the substrate. No refraction effects are considered in analyzing a grating; we may consider it as a thin, purely diffractive structure. Gratings also operate in reflection, as shown schematically in Figure 8.15(b).

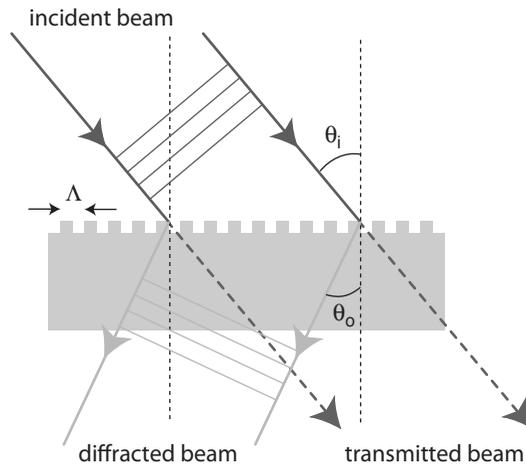
The schematic sketch of Figure 8.15(a) shows two intensity maxima transmitted through the transparent grating substrate. The angular relationship between the incident and diffracted beams is given by the grating equation,

$$\Lambda \sin \theta_i + \Lambda \sin \theta_o = m\lambda, \quad (8.48)$$

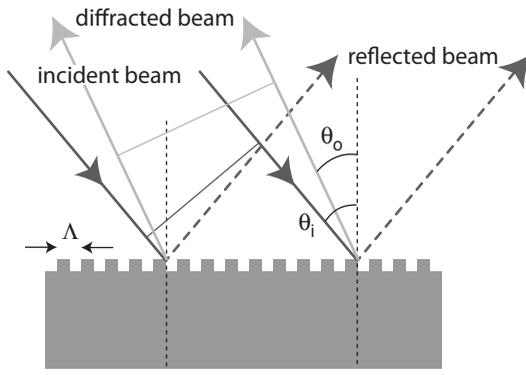
which may be derived by stipulating that diffraction maxima occur at angles for which the path length differences between wavefronts diffracted from each grating corrugation (i.e., each grating “tooth”) are integral numbers of wavelengths long. In the previous expression, we have employed the grating period Λ , wavelength λ and incident and diffraction angles θ_i and θ_o , respectively⁶. The index m [], with integer values $m = 0, \pm 1, \pm 2, \dots$, is known as the mode index, such that, for a given diffraction peak maximum, we speak of the “ m^{th} order” peak.

⁵ Classically, grating periods are given in *grooves per millimeter* [g/mm], and most commercially available gratings are specified in this manner. Many engineers with a microsystems background prefer to think in the inverse, namely period, Λ [μm], the convention which we will employ here.

⁶ The form of Equation 8.48 employs the convention that θ_i and θ_o on the same side of the grating normal have the same sign. Other angle sign conventions, such as that angles in the first and third quadrants are positive and those in the second and fourth quadrants are negative (Hutley, 1982, Chapter 2), may lead to a minus sign in the expression, thus $\Lambda \sin \theta_i - \Lambda \sin \theta_o = m\lambda$; if the angles are consistently defined, such a convention is as applicable as the one we employ here.



(a) Transmission grating: a beam incident at angle θ_i with respect to the normal of a transparent substrate, on which a corrugation with period Λ has been etched, is partially diffracted at an angle θ_o . The transmitted beam emerges at the incident angle θ_i ; refraction has been ignored.



(b) Reflection grating: a beam incident at angle θ_i with respect to the normal of a reflecting substrate, on which a corrugation with period Λ has been etched, is partially diffracted at an angle θ_o . The zero order, undiffracted reflected beam emerges at a reflection angle equal to the incident angle θ_i .

Figure 8.15 Diffraction by transmission and reflection gratings.

Using trigonometric identities, the grating equation may be written in an alternative form,

$$2\Lambda \sin\left(\frac{\theta_o + \theta_i}{2}\right) \cos\left(\frac{\theta_o - \theta_i}{2}\right) = m\lambda \tag{8.49}$$

which expresses the difference between input and diffraction angles explicitly. We will derive the grating equation in the next section.

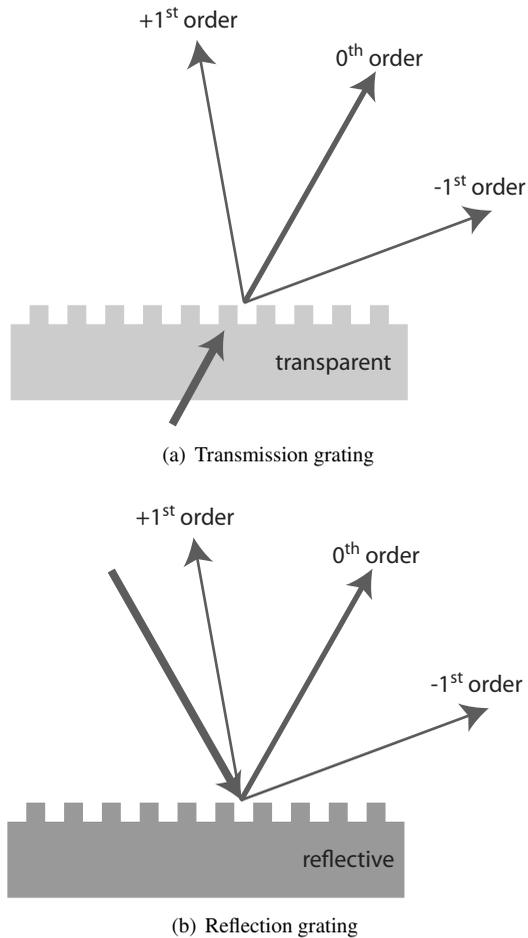


Figure 8.16 The 0th and ± 1 st order diffraction peaks for operation in transmission and reflection. The arrows represent the optical beams schematically, and in reality illuminate a large number of grating periods, as shown in Figure 8.15.

Diffraction orders

As we saw in Figure 8.15, diffraction gratings may be conceived in transmission or reflection, depending on the substrate. The first order diffraction peaks (with $m = \pm 1$), as well as the 0th order peak, are shown for both cases in Figure 8.16; the values of the diffraction angles are determined using the grating equation, Equation 8.48. The 0th order peak for a transmission grating, Figure 8.16(a), is the beam which would result in the absence of a grating and emerges from the substrate at the same angle as the incident beam. Likewise, the 0th order peak for the reflective case, Figure 8.16(b), is that due to specular (mirror-like) reflection, where reflection angle equals incidence angle with respect to the plane of the grating.

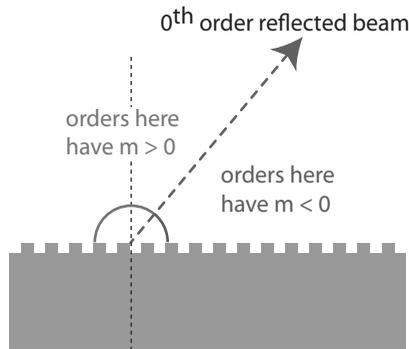


Figure 8.17 Sign convention for the designation of diffracted modes. An order is defined as positive if it is found on the same side of the zero order beam as the surface normal.

The number of diffraction orders generated by a grating depends on the wavelength, grating period and geometry of the incident beam. By solving the grating equation for increasing values of m , the angles of the higher order diffraction peaks may be determined. For angles which place a particular diffraction order inside the substrate, that order then no longer exists for that particular configuration. Whereas the angles of the higher-order diffraction peaks may easily be determined, a much more difficult calculation is that of diffraction *efficiency* for the different diffraction orders, meaning how much intensity is actually diffracted into that order. In general, the higher the diffraction order, the lower the efficiency, so that the intensity of the diffraction peaks decreases as m increases.

The sign convention for the diffraction order is shown in Figure 8.17. By definition, diffracted orders are positive if they are found on the same side of the zero order mode as the normal to the grating surface. Thus, in the figure, all modes positioned counter-clockwise from the specularly reflected (zero order) beam are positive, all orders positioned clockwise are negative. For symmetric systems with normal incidence, mode polarity is arbitrary, but as soon as the substrate is rotated with respect to the incident beam, this sign convention holds.

8.3.2 Grating diffraction

The functionality of a grating can be described mathematically using a basic analysis of diffraction from a periodically repeating arrangement of diffracting elements (Hutley, 1982, Chapter 1). Using the approach we applied to a variety of apertures in Section 8.2, we can represent a grating as a linear arrangement of N slits with a period Λ . The total electric field in an image plane, $E_G(x_i)$, may thus be expressed as a sum of the contribution of all N diffracting elements, which generate a field $E_G(x_a)$, namely

$$E_G(x_a) = E_0 \sum_{n=0}^{N-1} \delta(x_a - n\Lambda); \quad (8.50)$$

we have used a delta function to describe emission from a single period of the grating. Based on the discussion of Section 8.2, we recall that the Fraunhofer diffraction field may be found by taking the Fourier transform of the emitters, so that the diffraction image is given by

$$E_G(x_i) = E'_0 \sum_{n=0}^{N-1} e^{-jn k \Lambda \frac{x_i}{z_i}} \quad (8.51)$$

for propagation constant $k = 2\pi/\lambda$ and aperture/image (grating/image) spacing z_i . As in the previous calculations, we can express this result as a function of the angle θ between z_i and the aperture (grating) plane as

$$E_G(\theta) = E'_0 \sum_{n=0}^{N-1} e^{-jn k \Lambda \sin \theta}, \quad (8.52)$$

which is a more useful form for most grating arrangements in micro-optical systems.

The sum in the previous expression may be evaluated to yield

$$E_G(x_i) = E'_0 \frac{e^{-jNk\Lambda \sin \theta} - 1}{e^{-jk\Lambda \sin \theta} - 1} \quad (8.53)$$

$$= \frac{\sin\left(\frac{1}{2}Nk\Lambda \sin \theta\right)}{\sin\left(\frac{1}{2}k\Lambda \sin \theta\right)} e^{-\frac{1}{2}j(N-1)k\Lambda \sin \theta} \quad (8.54)$$

such that the diffraction intensity as a function of angle may be found from

$$I_G(\theta) = \frac{I_0}{N^2} \frac{\sin^2\left(\frac{1}{2}Nk\Lambda \sin \theta\right)}{\sin^2\left(\frac{1}{2}k\Lambda \sin \theta\right)}; \quad (8.55)$$

the factor $1/N^2$ derives from a normalization of the intensity that yields $I_G(\theta = 0) = 1$, since the intensity of that central peak is N^2 times the intensity from a single emitter. As we will see in the example which follows, this last expression for the diffraction intensity yields sharp maxima at certain angles, whose value is given by the relationship between the wavelength (in $k = 2\pi/\lambda$) and grating period Λ .

Non-normal incidence

The previous derivation assumed implicitly that the incident field which generated the diffraction pattern impinged normally onto the grating aperture. We can expand the analysis to the arbitrary case of non-normalized incidence by replacing

$$\sin \theta \rightarrow \sin \theta + \sin \theta_i \quad (8.56)$$

for incidence angle θ_i . The diffraction intensity distribution then becomes

$$I_G(\theta) = \frac{I_0}{N^2} \frac{\sin^2\left(\frac{1}{2}Nk\Lambda(\sin \theta + \sin \theta_i)\right)}{\sin^2\left(\frac{1}{2}k\Lambda(\sin \theta + \sin \theta_i)\right)}. \quad (8.57)$$

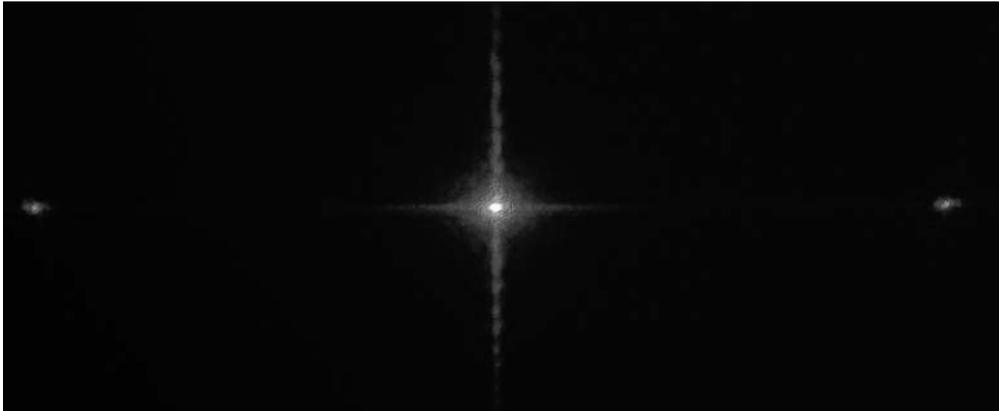


Figure 8.18 Diffraction pattern generated by illumination of a linear transmission diffraction grating with a HeNe laser emitting at 633 nm normal to the grating plane. The bright central peak is the zero order, the weaker peaks at the left and right the ± 1 order maxima. Photo courtesy of Bernd Aatz.

The function described by the previous expression reaches a maximum when numerator and denominator are both zero, so thus when

$$\sin^2 \left[\frac{1}{2} k \Lambda (\sin \theta + \sin \theta_i) \right] = 0 \quad (8.58)$$

implying

$$\frac{1}{2} k \Lambda (\sin \theta + \sin \theta_i) = \pm n \pi \quad (8.59)$$

or

$$\frac{\Lambda}{\lambda} (\sin \theta + \sin \theta_i) = n \quad (8.60)$$

for index $n = 0, \pm 1, \pm 2, \dots$. Comparing this last expression with Equation 8.48 shows a marked similarity, which is of course no accident since we have just derived the grating equation relating the angle of incidence (θ_i) to the diffraction angle (θ or θ_o). A simpler alternative derivation is the subject of one of the problems at the end of the chapter.

For normal incidence, $\theta_i = 0$, the angular positions of the diffraction peaks are given by the relatively simple expression

$$\theta_{peak} = \arcsin \left[\frac{n \lambda}{\Lambda} \right]. \quad (8.61)$$

This last relationship yields the peaks at angles easily seen when illuminating a cheap replicated transmission grating with a laser pointer, as seen in Figure 8.18. For this example, a linear transmission grating is illuminated with a HeNe laser at normal incidence. The bright central peak is the undiffracted zero order mode, and the ± 1 order diffraction modes are seen at the left and right. The difference in intensities is due to

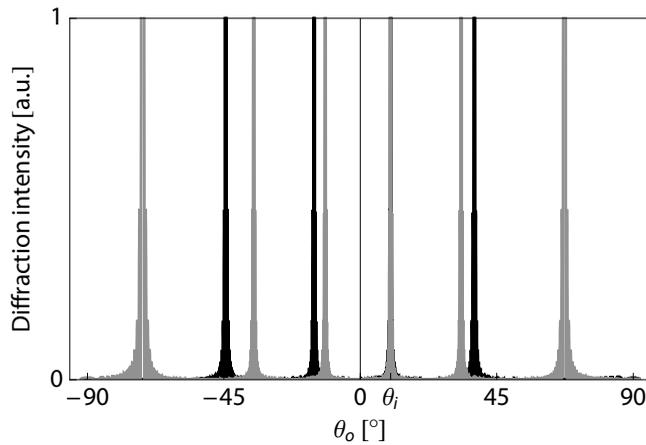


Figure 8.19 The calculated diffraction peaks generated by a grating with $\Lambda = 1.45 \mu\text{m}$ illuminated at an incidence angle $\theta_i = 10^\circ$ with two HeNe wavelengths, 633 nm *black* and 543 nm *gray*. The zero order at θ_i shows only the gray peak, but both overlap at this position.

the relative efficiencies of the diffracted modes; the ± 2 order peaks, which would be further to the left and right, would be even weaker.

Example 8.5: Given two HeNe lasers, one emitting at 633 nm and the other at 543 nm, let us derive the diffraction angles and plot the diffraction pattern which would be generated if these are incident at $\theta_i = 10^\circ$ onto a grating with $\Lambda = 1.45 \mu\text{m}$.

We can use the grating equation in the form we just derived it, Equation 8.60, and solve for θ_o . Doing so, and evaluating for a range of indices $m = 0, \pm 1, \pm 2, \dots$, we find, for the red wavelength:

$$\theta_o = -44.4^\circ, -15.2^\circ, 10.0^\circ, 37.6^\circ$$

and for the green,

$$\theta_o = -71.8^\circ, -35.1^\circ, -11.6^\circ, 10.0^\circ, 33.2^\circ, 67.3^\circ.$$

From this calculation, we consider two implications: 1) the longer wavelength has three diffraction peaks (excluding the zero order), whereas the shorter has five; and 2) the positions of the diffraction peaks are not symmetric about θ_i .

We can see these considerations clearly by examining the plotted intensity patterns in Figure 8.19; we employ Equation 8.57, using the numerical values given above. The diffraction peaks for the shorter wavelength are clearly more closely spaced than for the longer wavelength, with negative diffraction orders more numerous than the positive orders. An experimental implementation of this type of pattern will be seen in Figure 8.23(b), albeit for a two-dimensional grating with a longer Λ ; the relative positions of the peaks for the two wavelengths are the same, however.

8.3.3 Diffraction characteristics

We have seen how a grating generates strongly wavelength-dependent diffraction peaks at well-defined angular directions from the corrugated surface. A number of further parameters are important for characterizing the optical performance.

Dispersion

The function of a grating is usually to disperse the incoming optical spectrum into its wavelength components. The dispersion for a grating is then defined as the variation of the diffraction angle θ_o with wavelength λ and can be found by taking the derivative of the grating equation with respect to diffraction angle, thus

$$\frac{\partial}{\partial \theta_o} [\Lambda \sin \theta_i + \Lambda \sin \theta_o] = \frac{\partial}{\partial \theta_o} [m\lambda] \quad (8.62)$$

which leads to

$$\Lambda \cos \theta_i \frac{\partial \theta_i}{\partial \theta_o} + \Lambda \cos \theta_o = m \frac{\partial \lambda}{\partial \theta_o}. \quad (8.63)$$

For a single, non-varying incidence angle, $\partial \theta_i = 0$ and we are left with

$$\frac{d\theta_o}{d\lambda} = \frac{m}{\Lambda \cos \theta_o} \quad (8.64)$$

for diffraction order m . By applying the grating equation, Equation 8.48, we can also write the dispersion in the form

$$\frac{d\theta_o}{d\lambda} = \frac{\sin \theta_i + \sin \theta_o}{\lambda \cos \theta_o}. \quad (8.65)$$

Alternatively, a bit of trigonometry for the case of normal incidence leads to another useful form for dispersion, namely

$$\frac{d\theta_o}{d\lambda} = \frac{1}{\sqrt{\frac{\Lambda^2}{m^2} - \lambda^2}} \quad (8.66)$$

which is only a function of period, wavelength and diffraction order, for $\theta_i = 0$. For all three expressions, the units of $d\theta_o/d\lambda$ are rad/m.

We thus see that high dispersion results for higher-order diffraction modes, but, as we will see below, efficiency decreases as m increases. As a result, high dispersion is achieved in practice by decreasing the grating period Λ and working in as high an order mode as possible, recalling, however, that higher order modes have lower efficiency than lower order modes. For short wavelengths (in the visible and UV, for example), fabrication capabilities may limit the achievable periods, so that dispersion for higher order m may need to be considered.

Example 8.6: We can examine the variation of dispersion with wavelength and diffraction order using the expressions above. If we treat the case of normal incidence, $\theta_i = 0$, we can use either Equation 8.64 or Equation 8.66; the results will be the same. Let

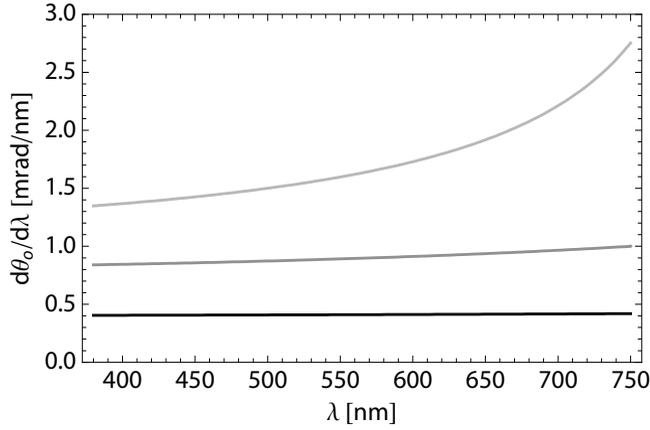


Figure 8.20 Calculated dispersion of a grating with period $\Lambda = 2.5 \mu\text{m}$ for $m = 1$ bottom, $m = 2$ middle, and $m = 3$ top.

we consider a grating with period $\Lambda = 2.5 \mu\text{m}$ for use in the visible wavelength range, $380 \text{ nm} \leq \lambda \leq 750 \text{ nm}$. Plotting the dispersion, in units of mrad/nm, for $m = 1, 2, 3$, in Figure 8.20, we see that the dispersion increases with mode index, as we mentioned above. In addition, although the curves for $m = 1, 2$, at this scale, look almost flat, there is an increase in dispersion for longer wavelengths.

Free spectral range

Since the same wavelengths diffracted in different orders may, depending on the grating characteristics and illumination conditions, overlap at the same θ_o , the free spectral range (FSR) defines the wavelength interval over which the wavelengths from an adjacent order do not overlap; we already employed the concept of FSR when analyzing the spacing between resonance peaks in a Fabry-Perot cavity, in Section 5.2.5. Defining λ_{long} and λ_{short} as the longest and shortest wavelengths, respectively, of the wavelength range of interest, obviating an overlap implies

$$m\lambda_{long} \geq (m+1)\lambda_{short} \quad (8.67)$$

for order m . Defining $\Delta\lambda_{FSR}$ as the free spectral range,

$$\Delta\lambda_{FSR} = \lambda_{long} - \lambda_{short}, \quad (8.68)$$

FSR may be expressed as

$$\Delta\lambda_{FSR} = \frac{\lambda_{long}}{|m|+1} \quad (8.69)$$

Thus the free spectral range decreases as diffraction order increases. For $m = -1$, for example, $\Delta\lambda_{FSR} = \lambda_{long}/2$, such that the FSR is given by $\lambda_{short} \leq \lambda \leq 2\lambda_{short}$.

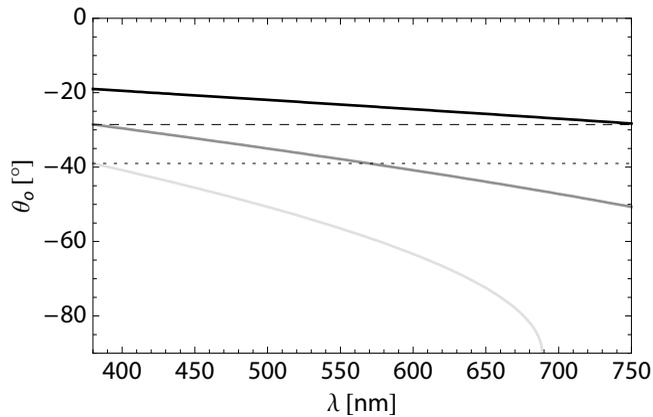


Figure 8.21 Calculated diffraction angles as a function of wavelength in the visible range for a grating with period $\Lambda = 2.5 \mu\text{m}$ and an incidence angle of $\theta_i = 10^\circ$, for $m = -1$ top, $m = -2$ middle, and $m = -3$ bottom. The dashed line represents the value of θ_o for $m = -2$ at $\lambda = 380 \text{ nm}$; the dotted line, the value of θ_o for $m = -3$ at $\lambda = 380 \text{ nm}$.

Example 8.7: Given a grating with period $\Lambda = 2.5 \mu\text{m}$ with the optical field incident at $\theta_i = 20^\circ$, what is the FSR for the visible wavelength range, $380 \text{ nm} \leq \lambda \leq 750 \text{ nm}$, for the first three diffraction orders?

We can estimate the values for FSR using the expression above, $\lambda_{short} \leq \lambda \leq 2\lambda_{short}$, which would yield simply $380 \text{ nm} \leq \lambda \leq 760 \text{ nm}$ for $m = -1$ and $380 \text{ nm} \leq \lambda \leq 570 \text{ nm}$ for $m = -2$. The magnitudes of the FSR are then 380 nm and 190 nm , respectively.

The relationships between the diffraction angles of the various orders can be seen more clearly, perhaps, by plotting the diffraction angle θ_o as a function of wavelength in the complete range, for the first three orders, as shown in Figure 8.21. We see that the magnitude of diffraction angle, $|\theta_o|$, increases with wavelength, and the point at which the values overlap defines the FSR. The dashed horizontal line shows the value at which the $m = -2$ order appears, which would intersect the $m = -1$ order at 760 nm , just off the horizontal axis. The dotted horizontal line shows the equivalent for the $m = -3$ order, which intersects the $m = -2$ order at 570 nm , thus defining the FSR for $m = -3$ as $380 \text{ nm} \leq \lambda \leq 570 \text{ nm}$.

Stray light and ghosts

Diffractive gratings do not direct all the incident light into diffraction orders; due to manufacturing inaccuracies and defects, some of the light is scattered into undesired directions. Two optical effects result, which can limit the performance of grating-based systems. Stray light is unfocused light, which may propagate in virtually any direction from the grating surface. Defects, dirt and other non-regular features give rise to the

scattering which manifests itself as stray light and this tends to increase, particularly for ruled gratings, as Λ decreases.

Ghosts, on the other hand, result from parasitic periodic structures within the grating, or possibly ruling errors which result in super-periodicities or other regularly repeated features. In contrast to stray light, ghosts are typically focussed along with the desired diffracted orders, such that they can give rise to spurious signals in an optical system.

Efficiency

In principle, any regularly repeated profile will give rise to a diffraction pattern, provided the period Λ is uniform. Grating efficiency is a measure of how much of the optical input intensity is diffracted into a given mode, such that high efficiency is typically desirable. Generally, all else being equal, efficiency decreases with increasing mode number, such that the amount of energy in higher order modes decreases as mode number increases.

Two definitions of efficiency, η [], are generally applied to gratings. Absolute efficiency, η_{abs} , is that percentage of incident optical energy which is diffracted into a particular order. In contrast, relative efficiency, η_{rel} , is the ratio of diffracted energy to that which would be reflected if the surface were a mirror of the same material. η_{rel} is always larger than η_{abs} , so, particularly for the use of commercial gratings in an optical design, it is important to verify which efficiency is given in a specification.

Optimization of diffraction efficiency is usually a primary goal of grating design but is unfortunately not trivial to calculate (Turunen, 1998). Efficiency is the function of a number of factors, including grating shape, complex refractive index of the surface material, groove depth and period, as well as polarization and wavelength of the incident field. TM polarization, for which the electric field is perpendicular to the grating grooves, leads to particularly complex efficiency behavior, with a rapid change of η with wavelength typically resulting, leading to features commonly known as Wood's anomalies⁷ (Wood, 1902) in the efficiency spectrum.

For a given grating, efficiency is typically given as a plot of η as a function of λ , for a particular diffraction order, and data sheets of commercial products usually include these. For a microfabricated grating, the engineer usually has to measure η herself. Sinusoidal gratings are typically more efficient than ruled gratings and the former can have $\eta > 0.5$; these values can be reached if a sinusoidal grating with $\Lambda \leq 1.25\lambda$ (for which only the 0 and -1 modes exist) has its groove depth optimized for the $m = -1$ order, for example. For a rectangular grating, on the other hand, the efficiency may be varied by changing the duty cycle (the ratio of high to low portions in a grating period). Blazed gratings, as we will see below, can have efficiency values $\eta \geq 0.7$ for the blaze wavelength; multi-level gratings can achieve $\eta = 0.99$ (Destouches et al., 2005).

⁷ Robert W. Wood (1868–1955), professor of physics at Johns Hopkins University in Baltimore, greatly improved the performance of diffraction gratings but also developed animated films, discovered electron field emission and wrote books of nonsense verse to entertain his children.

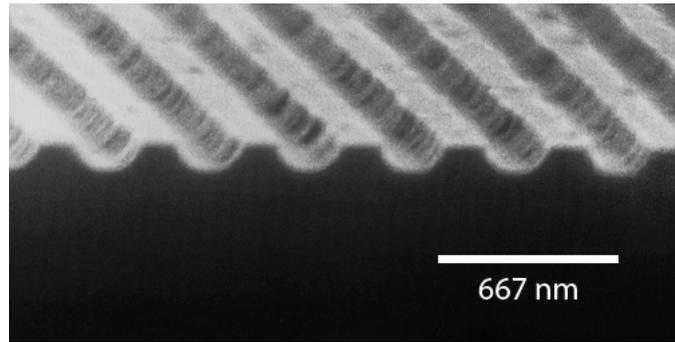


Figure 8.22 SEM photograph of a roughly rectangular, holographically-defined diffraction grating, etched into GaAs. This view is a detail of the grating used as a wavelength-selective mirror for a DBR laser, shown in Figure 8.30(b).

8.3.4 Grating forms

In the schematic drawings of gratings we saw above, we used rectangular grating profiles, sometimes known as lamellar gratings. Such a profile is practical from the point of view of a geometrical analysis, in that the path length difference between the upper and lower part of the grating can be used to determine the interference intensity at an arbitrary point. In addition, a rectangular grating usually suppresses the $m = \pm 2$ mode, leading to higher intensity in the $m = \pm 1$ order mode.

Most real gratings, however, have considerably less ideal shapes, as we may see in the example of Figure 8.22. This linear grating, with a period of about 500 nm, is etched into a GaAs surface after being defined by interference lithography, a fabrication technique which we will encounter in Section 11.3.4. Whereas the grating profile looks like a somewhat rounded rectangle, or a perhaps a decapitated sinusoid, and the sidewalls in the direction of the grating length are rough (on the scale we are examining them), the period is highly uniform over a long distance normal to the grating length.

Types

Most gratings fall into one of three categories:

Ruled gratings are the original form of grating, mechanically machined into a substrate, usually glass (Verrill, 1982). The shape is generally roughly rectangular or blazed (Hutley, 1982, Chapter 4).

Holographic gratings are fabricated using large-area interference lithography (Wu and Glytsis, 2003), which we will discuss in Chapter 11. The profile is sinusoidal or triangular and very uniform.

Blazed gratings will be discussed in detail in Section 8.3.5 and have a triangular or sawtooth profile. They are designed for highest efficiency at a particular operating wavelength.

The type of grating to be employed in a micro-optical system is a function of the application and its demands. Ruled gratings, which are usually distinctly not micro-optical components, but are often used in optical microsystems, are usually relatively inexpensive and often optimal for wavelengths longer than the mid-IR, $\lambda \geq 1.2 \mu\text{m}$; they often have good performance for large Λ . Ruled gratings are generally the optical workhorses in macroscopic optical systems.

Holographic gratings, on the other hand, are to be preferred for laser illumination, and for visible to UV wavelengths where short grating periods are required (below $1 \mu\text{m}$) (Naulleau et al., 2001). Sinusoidal gratings are very uniform over a large area, and may be fabricated using interference lithography on substrates exceeding 1 m^2 in size (Gombert et al., 2004). Blazed gratings, finally, may be fabricated either by holographic, lithographic processes or by means of ruling machines, and, as we will see below, are configured for a particular design wavelength.

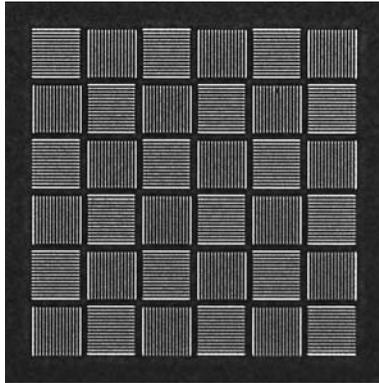
Chirped, curved and 2-dimensional gratings

In some situations, non-regular gratings may be desirable. A so-called chirped grating⁸ has a slowly spatially-varying period, so that $\Lambda \rightarrow \Lambda(x)$. As a result, the position of a diffraction peak for a given wavelength will vary as the beam is scanned in the x direction across the grating. Alternatively, a non-scanning but expanded beam which illuminates a large portion of a chirped grating will have a spatially varying (but stable) distribution of diffraction maxima, an occasionally useful optical arrangement.

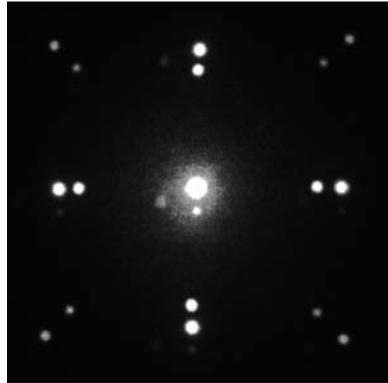
Alternatively, a grating may be curved (as opposed to linear) in the direction normal to the grating period. A circular grating, for example, may be used to diffract and focus a certain wavelength simultaneously, allowing a two-dimensional modification of the diffraction pattern beyond that of single peaks (Miler and Skalsky, 1980). This type of function approaches that of diffractive lenses, which we consider in Section 8.4 below. For macroscopic gratings, fabrication of the periodic structure on a spherically curved surface (Hutley, 1982, Chapter 7) is also a means to simultaneously disperse and focus an optical field.

In general, diffractive grating structures may be expanded and combined in two dimensions. Two orthogonal linear gratings may be superimposed, for example, forming a rectangular arrangement of diffracting elements. In general, periodic structures may be combined at any angle, leading to a combined periodicity at certain angles and thus more complex diffraction patterns. Higher efficiency may be achieved by alternating orthogonal gratings at different positions, as shown in Figure 8.23. In the chessboard-like arrangement of Figure 8.23(a), the linear orientation changes by 90° for every other square, such that the diffraction pattern, an example of which is seen in Figure 8.23(b), has intensity maxima in two orthogonal directions. This diffraction example was done at two wavelengths, showing the wavelength-dependence of diffraction angles explicitly.

⁸ The designation chirp refers to a sweep in frequency, spatial frequency in this case. The love songs of chirping birds generally include a sweep through a range of auditory frequencies, hence the allusion.



(a) A two-dimensional arrangement of linear grating fields, oriented orthogonally to one another. The $50 \times 50 \mu\text{m}$ fields have linear gratings with $\Lambda = 3.3 \mu\text{m}$. Structure and photo courtesy of David Kallweit.



(b) Diffraction pattern of a two-dimensional linear grating, with two HeNe wavelengths for illumination; the inner dots of the diffraction spot pairs are green ($\lambda = 543 \text{ nm}$) and the outer are red ($\lambda = 633 \text{ nm}$). Photo courtesy of Bernd Aatz.

Figure 8.23 A two-dimensional arrangement of orthogonal linear diffraction gratings and its associated diffraction pattern.

There is a variety of further grating arrangements for specialized applications. Notable are concave gratings, fabricated on a curved substrate, which allow simultaneous dispersion and focussing of an optical field, or gratings fabricated directly on the surface of curved refractive lenses (Traut et al., 2000).

8.3.5 Blazed gratings

A special and very useful case of grating structure is the blazed grating, sometimes referred to as a sawtooth or echelette grating (Barrekette and Christensen, 1965; Kallioniemi et al., 2000). As we see clearly in the diffraction characteristics of Figures 8.18 and 8.23(b), the zero order mode is always the brightest, since it has a higher efficiency than all higher order modes, for a rectangular grating. However, the angular position of the $m = 0$ mode has no wavelength dependence, such that the diffraction effect of the grating is unrealized for a large portion (in many cases most) of the energy in the optical field.

Diffraction into higher order modes may be optimized by using the blazed grating structure shown in Figure 8.24. Developed along with advanced grating fabrication techniques at Johns Hopkins University in the early twentieth century, blazed gratings have the prismatic shape shown in Figure 8.24(a), characterized by a blaze angle α [rad] as defined in Figure 8.24(b). Typical values for the blaze angle are in the range 5° to 60° . Specular reflection combined with interference from the surfaces angled at α with respect to the plane of the grating (i.e., the substrate, not the individual grooves) result in an increase in the intensity of the first order mode, to the detriment of the zero order

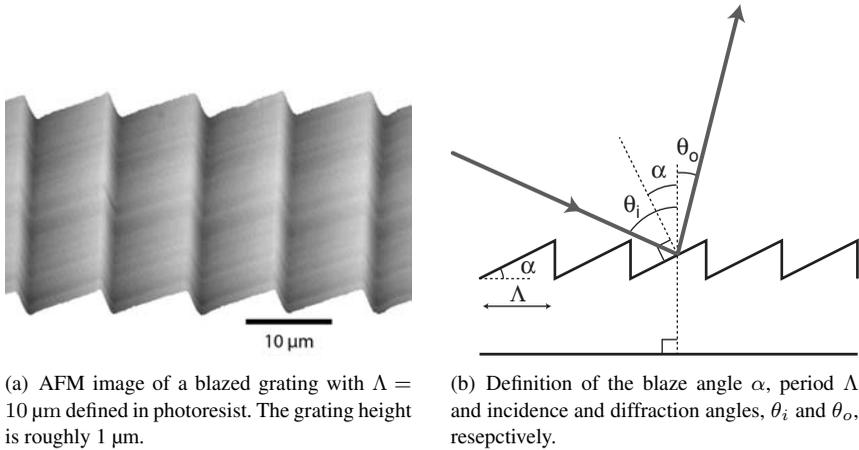


Figure 8.24 The blazed grating results in greater efficiency for the diffracted modes.

mode, thereby significantly increasing the diffraction intensity in the $m = 1$ mode for a narrow range of wavelengths.

Blaze condition

By examining the relationships between the angles of the incident and specularly reflected beams on the grating surface, as defined in Figure 8.24(b), we see that θ_i and θ_o are still defined relative to the grating substrate, but that the equality of angles of incidence and of reflection is satisfied with respect to the normal of the grating surface, tilted at α with respect to the substrate. As a result, we may define the blaze condition,

$$\theta_i - \alpha = -\theta_o + \alpha. \quad (8.70)$$

where we define the polarities,

- $\theta_i > 0$ if to the *left* of the normal to the grating surface;
- $\theta_o < 0$ if to the *right* of the normal to the grating surface.

For a blazed grating, therefore, the form of the grating equation depends on how the grating is mounted in the optical system. The Littrow mount, which we will consider in detail presently, is likely the most popular. For blazed gratings used in transmission, a normal mount (the plane of the substrate normal to the incident beam) is frequently used to avoid additional refraction at the rear substrate surface.

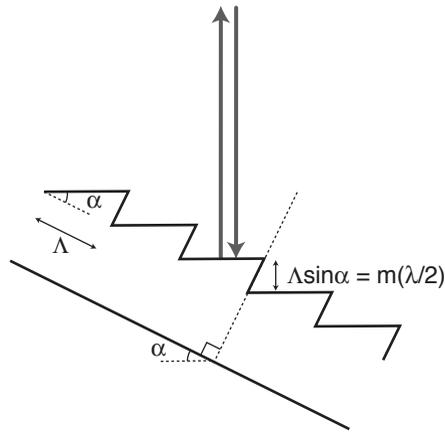


Figure 8.25 A blazed grating in the Littrow configuration; the grating substrate is tilted at α such that the vertical incident and diffracted beams are in autocollimation.

Littrow mount

For the Littrow mount⁹, the grating is placed into the optical field such that the planes of the grating are normal to the incident optical field, as shown in Figure 8.25. In this case, the incident and diffracted rays are in autocollimation, implying that $\theta_i = \alpha$ and $-\theta_o = -\alpha$. The grating equation, Equation 8.48, becomes

$$2\Lambda \sin \alpha = m\lambda \quad (8.71)$$

where the wavelength which satisfies this equation, for a given Λ and α , is termed the blaze wavelength, λ_B [m]. We see from the geometry of Figure 8.25 that the vertical spacing between the grating periods is then given by

$$\Lambda \sin \alpha = \frac{m\lambda_B}{2}, \quad (8.72)$$

thus an integral number of half-wavelengths of λ_B . The “steps” between adjacent grating periods thus result in constructive interference between the reflections from all grating periods at the blaze wavelength.

A Littrow-mounted blazed grating therefore gives rise to high efficiency diffraction at the blaze wavelength, since the efficiency of the first order mode is greatly enhanced and that of the zero order mode relatively depressed. A blazed grating is thus generally specified by its blaze wavelength, λ_B , that for which the highest efficiency is reached (Popov et al., 1995). As a general rule of thumb, for illumination in a wide wavelength range, the diffraction efficiency is greater than 0.41 of the peak efficiency in the wavelength range

$$\frac{2m}{2m+1}\lambda_B \leq \lambda \leq \frac{2m}{2m-1}\lambda_B \quad (8.73)$$

⁹ Joseph Johann von Littrow (1781-1840), Austrian aristocrat and astronomer, was director of the observatory in Vienna and was said (perhaps apocryphally) to have proposed setting fire to large kerosene-filled canals configured as geometric shapes in the Sahara, to signal extraterrestrial civilizations.

for diffraction order m . The efficiency decreases more rapidly for shorter wavelengths, so the blaze wavelength should not necessarily be placed at the center of the spectrum of interest. Efficiency calculations for blazed gratings become complicated for blaze angles larger than a few degrees due to polarization effects, but holographic blazed gratings with high efficiency in a wider wavelength range have been designed (Davidson et al., 1992) as have wavelength-tunable blazed gratings (Li et al., 2006a). We will see in the case studies of Section 8.5.1 below that the Littrow mount is the preferred manner of coupling a grating into an external cavity tunable laser system.

The dispersion of a blazed grating mounted in the Littrow configuration, finally, can be found by a simplification of Equation 8.65, using the Littrow condition of Equation 8.71, as

$$\frac{d\theta_o}{d\lambda} = \frac{2}{\lambda} \tan \theta_o, \quad (8.74)$$

thus simply expressed as a function of wavelength and diffraction angle.

8.4 Diffractive microlenses

We have seen how diffraction effects result from finite apertures, yielding wavelength-dependent intensity patterns, and how diffraction gratings employ these effects to efficiently disperse an optical field. Diffraction may also be used to form an optical field, almost arbitrarily. As the size of optical elements decreases, diffractive effects become increasingly important, and at sufficiently reduced dimensions, *only* diffraction is relevant. As a result, there are roughly three physical regimes: pure refraction, a mixture of refraction and diffraction, and pure diffraction.

The function of any optical element is to generate a desired phase profile in an optical field; diffractive lenses accomplish this using diffraction, or a combination of refraction and diffraction. In this section, we will consider diffractive optical components which can be used to shape an optical field (Herzig, 1998). We examine diffractive microlenses, diffractive and refractive hybrids, and diffractive optical elements in general.

8.4.1 Fresnel diffraction

The discussion of diffraction to this point, particularly in Section 8.2, has almost exclusively been for that of the far-field, or Fraunhofer, regime, thus that for which Equation 8.1 holds. This regime is that which is probably most frequently encountered in micro-optical systems, since the diffraction patterns generated a larger distance from an aperture or other diffracting structure are stable with distance, just changing in size. Luckily for us, calculation of far-field diffraction patterns is relatively easy and can often be done by hand for simple structures, as we saw above.

The near-field, or Fresnel, regime is more complicated, but is often that encountered when designing or analyzing diffractive lenses or optical elements, as we will do in the

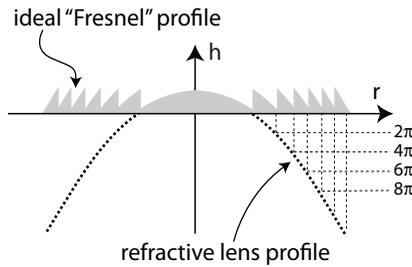


Figure 8.26 Construction of a diffractive Fresnel-like lens; a new zone is created whenever the height of the equivalent refractive profile increases by an integer multiple of full wavelengths, thus a phase shift of 2π or $2n\pi$.

next two sections. The near-field diffraction patterns are a strong function of spacing and their determination is often non-trivial. In particular, the obliquity (the angles of the waves emitted by the Huygens point sources, only in the forward direction) needs to be considered, making calculations more involved. We will now consider the diffraction of a number of optical elements without applying rigorous diffraction theory, since an in-depth discussion of Fresnel diffraction would go beyond the scope of this text; the interested reader is referred to excellent discussions in the literature, for example (Goodman, 2004), (Hecht, 2002, Chapter 10.3) and, for the calculation of diffractive optics in general, (Turunen and Wyrowski, 1997).

8.4.2 Diffractive “Fresnel” lenses

The Fresnel lens was presented as a special case of micro-optical lenses in Section 7.6.4. We saw there that the concept underlying the Fresnel lens was that the curved surface of a refractive lens could be deconstructed into a flatter equivalent, with the different zones maintaining the curvature of the original thick lens; the result was a purely refractive optical element.

The Fresnel concept may also be applied to a type of diffractive lens, wherein diffraction can be used to focus an optical field. Whereas the form of such a lens looks much like a classical Fresnel lens, the lenses we consider here rely on a combination of diffraction and refraction or, in the limiting case, solely diffraction. Thus we cautiously denote the lenses in this section “Fresnel-like” lenses, to distinguish them from their refractive relatives.

A diffractive Fresnel-like lens may be conceived using the construction of Figure 8.26, which is not un-similar to Figure 7.44. As the sizes of the zones in this lens decrease to the order of several wavelengths, diffraction effects begin to play an increasingly important role. We can thus fold back the profile of an ideal refractive lens whenever its height has changed by an integer multiple of full wavelengths in the material, thus representing a phase shift of 2π . Each zone maintains the curvature of its corresponding portion of the refractive profile, but the phases are matched so that the overall phase front corresponds to that which would result from the macroscopic refractive profile. The transition between a purely refractive and a purely diffractive lens is a fluid one,

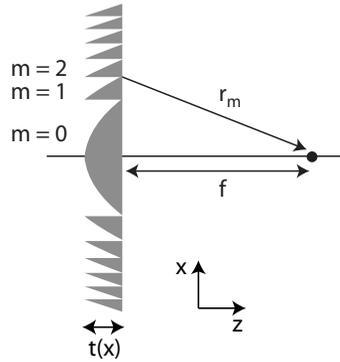


Figure 8.27 Focussing an optical field using a Fresnel-like diffractive lens. The spacing between zone m and the focus f is an integral number of wavelengths longer than the focus, $r_m = f + m\lambda_0$.

such that for this refractive/diffractive hybrid lens, we can calculate the optical effect of the lens by determining the total phase shift imparted to the wavefront, including all interference effects, by the lens structure.

Consider the diffractive Fresnel-like lens used for focussing a collimated beam, with the infinite conjugate to the left, in Figure 8.27. We use a lens material with refractive index n_L , operated in air, in which the optical field with wavelength λ_0 has a propagation constant k_0 . The required phase function of the lens as a function of dimension x , $\Phi(x)$ [rad], leads to a thickness variation, $t(x)$ [m], which we can write in the form

$$t(x) = t_{max} \left(\frac{\Phi(x)}{2\pi} + 1 \right) \quad (8.75)$$

where t_{max} [m] is the maximum lens thickness (at the center, for $x = 0$). The phase function represents the phase difference between a plane wave and one which is focused to the focal point f , and, for a spherical profile, is given by

$$\Phi(x) = k_0 n_L \left(f - \sqrt{x^2 + f^2} \right) \quad (8.76)$$

where we limit the phase to (modulo 2π),

$$-2\pi \leq \Phi(x) \leq 0. \quad (8.77)$$

This last criterion allows us to define the individual zones of the lens, as indicated in the figure, such that we discretize x to fall in the range

$$x_m \leq |x| \leq x_{m+1}. \quad (8.78)$$

In the last expression, m [] is an integer numerical index representing the zone number, and the start of zone m is at position

$$x_m = \sqrt{\frac{2m\lambda_0 f}{n_L}}. \quad (8.79)$$

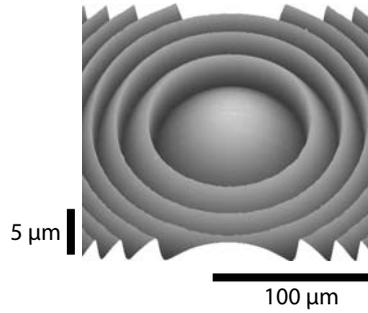


Figure 8.28 A Fresnel-like microlens with continuous-profile zones defined in photoresist. Photo courtesy of Markus Rossi.

We can simplify the above expressions using the approximation

$$\Phi(x) = k_0 n_L \left(f - \sqrt{x^2 + f^2} \right) \approx - \left(\frac{k_0 n_L}{2f} x^2 \right) \quad (8.80)$$

such that the phase may be written in the more convenient form

$$\Phi(x) \approx 2m\pi - \left(\frac{k_0 n_L}{2f} x^2 \right) \quad (8.81)$$

and finally

$$t(x) \approx t_{max} \left[(m+1) - \left(\frac{k_0 n_L}{4\pi f} x^2 \right) \right]. \quad (8.82)$$

From Equations 8.79 and 8.82 and the other definitions above, we can then define the required shape for a Fresnel-like microlens, and directly calculate the required widths of the zones and their profile. Thus a Fresnel-like microlens may be easily designed, given the required focal length and material refractive index; an example of such an element, defined in photoresist, is shown in Figure 8.28.

Operation

In the simple design methodology above, we implicitly used constructive interference from the distributed diffracting elements of the lens to generate a maximum of constructive interference (and thus intensity) at the focus. As a result, the radial distances r_m , between each segment m and the focus f as shown in Figure 8.27, are integer multiples of the wavelength λ_0 . In the paraxial region, the focal length may be expressed as

$$f_0 = \frac{n_0 r_m^2}{2m\lambda_0} \quad (8.83)$$

such that the focal length for a given wavelength λ is given by

$$f(\lambda) = f_0 \frac{\lambda_0}{\lambda} \quad (8.84)$$

where λ_0 is the design wavelength and n_0 is the refractive index of the surrounding medium typically air.

Equation 8.84 clearly shows that a Fresnel-like microlens of this type, in contrast to a “real” macroscopic refractive Fresnel lens, is highly dispersive, meaning its operation varies strongly with wavelength. High dispersion is typical for diffractive microlenses, since their operation essentially relies on the design of an interference pattern and this is typically defined for a single design wavelength. We will discuss the resulting chromatic aberration in Section 8.4.4 below.

8.4.3 Fresnel zone plates

As the lens dimensions of Fresnel-like microlenses shrink, the ability to define the curved surface of each of the zones becomes more limited. As we will discuss in detail in Section 11.3.3, the continuous profile of a Fresnel lens (or other continuous-relief structure) may be approximated by a stepped profile, a design and fabrication concept known as binary optics. Binary optical structures generally have numerous vertical steps, or phase levels.

When a Fresnel-like lens is fabricated using only two phase levels (i.e., two steps), the result is known as a Fresnel zone plate, sometimes referred to simply as a zone plate. Fresnel zone plates (an example is shown in the bottom sketch of Figure 11.7) are used with technologies which do not allow multi-level profiles or those which use only alternating transparent and opaque regions on a substrate and may also be fabricated with macroscopic dimensions.

Fresnel zone plates are useful diffractive lenses for two primary reasons: 1) they may be designed and manufactured relatively easily, since only two levels are required in the pattern; and 2) they are useful optical elements for wavelengths for which suitable refractive materials do not exist, particularly in the x-ray regime. For the latter application, extreme designs may involve 9 mm diameter lenses with up to 44 000 zones, the smallest of which has dimensions of 54 nm, using Ni as an absorbing material (Anderson et al., 2000). Alternative approaches use arrays of switchable Fresnel zone plates for nanolithography (Smith, 1996). Micro-optical zone plates have also been used to image beams of atoms (Carnal et al., 1991) and electrically-tunable zone plates using liquid crystals have been demonstrated (Ren et al., 2003).

8.4.4 Aberrations

As for refractive lenses, diffractive microlenses are subject to aberrations which limit their imaging quality. The origins and means for correction of these aberrations are different than for refractive optics.

Chromatic aberrations

Diffractive lenses, like all the diffractive structures we have seen in this chapter, are very strongly wavelength-dependent; diffractive imaging or beam-shaping structures are typically designed for operation at a single wavelength, and are thus most suitable for laser illumination. Equation 8.84 shows the wavelength-dependence of focal length for a Fresnel-like diffractive lens explicitly, and we recall that this relationship was derived

from the structure of the lens itself. We see thus that the focal length scales as λ^{-1} for a diffractive lens.

In contrast, as we learned in Section 7.5, the chromatic aberration of a refractive lens is due to the dispersion of the refractive material, to which we were introduced in Section 3.2, which results in a wavelength-dependent refractive index, $n(\lambda)$. Slightly re-writing Equation 7.19, the focal length of a refractive lens operated in air may be expressed as

$$f(\lambda) = R \left(\frac{1}{n(\lambda) - 1} \right) \quad (8.85)$$

for a single curved surface with radius R . Since, for the visible wavelength range, $n(\lambda)$ varies relatively slowly with wavelength (see the variation plotted in Figure 3.5 for SiO_2), the change in focal length with wavelength, and thus the chromatic aberration, for refractive lenses is considerably weaker than for diffractive components.

Abbe numbers

Just as we used the Abbe number to characterize the dispersive power of a refractive lens using the definitions of Section 3.2.2, we can similarly define an Abbe number for diffractive microlenses, ν_D []. Recalling that in the case of diffraction, focal length is inversely related to wavelength, $f \propto \lambda^{-1}$, we can write

$$\nu_D = \frac{\lambda_d}{\lambda_F - \lambda_C} \quad (8.86)$$

where the standard wavelengths are again those defined in Table 3.1, recalling that $\lambda_C > \lambda_d > \lambda_F$. We see from the previous expression that Abbe numbers for diffractive lenses, in contrast to the refractive case, are always negative. Typical values are $\nu_D \approx -3.5$ for the visible wavelength range and $\nu_D = -2.5$ for the IR range, between 8 and 12 μm . We emphasize that, in marked contrast to the refractive case, the Abbe numbers for diffraction are *not* a function of the material.

Negative Abbe numbers imply that focal length for longer wavelengths is smaller than for shorter wavelengths (red focusses closer to the lens than blue), whereas positive values for ν imply the opposite, as we saw for geometric (refractive) optics (blue focusses closer to the lens than red). This contrast in behaviors implies that we can combine diffractive and refractive structures to reduce overall chromatic aberration, as we will see presently in Section 8.4.5.

Primary aberrations

Diffractive lenses are also subject to some of the primary (non-chromatic) aberrations which we saw for geometric optics in Section 7.4 (Buralli and Morris, 1991). Petzval curvature and distortion are zero for diffractive lenses, but the remaining three aberrations are present and can be characterized in the same manner as for refractive lenses (Morris and McIntyre, 1997, Chapter 3.2).

As for refractive optics, aberrations for diffractive lenses may be minimized using properly placed stops and by restricting the aperture. In addition, in the diffractive case,

spherical aberration for on-axis, rotationally-symmetric objects can be eliminated for a single wavelength and can be reduced by bending the substrate or using “aspheric” shapes for the individual zones. For imaging using equal conjugates, an equi-convex lens structure is optimal to reduce SA, and, in the case of one infinite conjugate, the curved side should face that collimated side, just as for the refractive case.

In diffractive lenses, coma is zero for 1:1 imaging and can be eliminated for the other cases if an appropriately curved substrate for the lens is used, possible using flexible materials. Astigmatism, finally, can be reduced using an aperture stop, but frequently at the expense of introducing excess spherical aberration.

8.4.5 Diffractive/refractive hybrids

One useful application of diffractive optics is in improving the performance of refractive components. Diffractive/refractive hybrid lenses combine the characteristics of diffraction with the optical power of refraction (Behrmann and Mait, 1998). Both micro-optical lenses as well as macroscopic components may benefit from diffractive effects and certain functions, such as achromaticity in the IR range (Wood, 1992), can only be practically accomplished using this approach.

We consider here athermal lenses (athermats, with reduced sensitivity to temperature changes), achromatic lenses (achromats, with reduced chromatic aberration) and diffractively-corrected microlenses.

Athermal hybrids

We saw in Section 3.2.3 that the refractive index of a material varies with temperature. The astute physics student will recall that virtually all materials, including the glasses we saw in Section 3.4.3, also expand or contract with temperature, this expansion given by the thermal coefficient of expansion (see for example (Ralls et al., 1976, Chapter 24.2)). As a result, changes in ambient temperature will result in a change of focal length for both refractive and diffractive lenses; an athermal lens is one which compensates for and thereby reduces these changes to the greatest extent possible (Londoño et al., 1993).

The linear thermal coefficient of expansion of a material, α [K^{-1}] is given by

$$\alpha = \frac{1}{L} \left(\frac{\partial L}{\partial T} \right) \quad (8.87)$$

for length L and temperature T . Typical values for α at 20°C are about $8.5 \times 10^{-6} \text{ K}^{-1}$ for glass, $0.6 \times 10^{-6} \text{ K}^{-1}$ for quartz and $3 \times 10^{-6} \text{ K}^{-1}$ for silicon. We may combine this parameter with the temperature variation of refractive index and write the change in focal length of a refractive lens with temperature as (Jamieson, 1981)

$$\frac{\Delta f}{\Delta T} = \left[-\frac{1}{n-1} \frac{dn}{dT} + \alpha \right] f = \xi f \quad (8.88)$$

for focal length f and material refractive index n ; the term in the square brackets is defined as the opto-thermal expansion coefficient ξ [K^{-1}]. Depending on the material, ξ can be positive or negative, and may vary widely in magnitude.

For diffractive lenses, we saw that the critical dimension defining the focus is the spacing between a zone m and the focal point, r_m , as defined in Figure 8.27. The temperature variation of this spacing is given by

$$r_m(T) = r_m + \alpha r_m \Delta T \quad (8.89)$$

so that, for a diffractive lens,

$$\frac{\Delta f}{\Delta T} = 2\alpha f \quad (8.90)$$

assuming that the temperature variation of ambient refractive index (typically air), $\Delta n_0/\Delta T$, is negligible. $\Delta f/\Delta T$ is always positive for diffractive lenses, so that $\xi > 0$ for the diffractive case.

For an athermal lens, we need to obtain a value as close to $\Delta f/\Delta T = 0$ as possible. Since we saw that the values of ξ for refractive lenses can be positive or negative and, most importantly, potentially (depending on the material) of opposite sign for refractive and diffractive lenses, we realize that combining refractive and diffractive power in a single lens can lead to athermalization. In micro- and macro-optics, this combination is usually realized by making a refractive/diffractive doublet lens or etching a diffractive lens directly onto a refractive surface.

Designing such an athermal doublet requires that the optical powers ($1/f$) are divided into a refractive and a diffractive part, such that the sum is then the required optical power of the lens. Expressing this requirement in terms of focal length, we have

$$\frac{1}{f} = \frac{1}{f_r} + \frac{1}{f_d} \quad (8.91)$$

for required total focal length f , and the refractive and diffractive contributions to the desired focal length, f_r and f_d , respectively. The total opto-thermal expansion coefficient for the doublet, ξ_T , is then given by

$$\xi_T = \xi_r \frac{f}{f_r} + \xi_d \frac{f}{f_d}. \quad (8.92)$$

By dividing the optical power of the lens into a refractive and diffractive part, then, the total thermal expansion coefficient can be made zero, such that, in a given temperature range, the optical properties of the lens do not vary with temperature. Some further constraints may be applied to the design, such as that the total linear coefficient of expansion of the lens should match that of the mount or substrate material, to prevent the generation of mechanical stresses as the temperature varies.

Example 8.8: To see how the balance of refractive and diffractive power can lead to athermalization, with proper choice of lens material, we can consider a lens made using SF11 glass. For this material, we have (Behrmann and Mait, 1998, Table 10.3)

$$\left. \frac{\Delta f}{\Delta T} \right|_{\text{refractive}} = (-10.4 \text{ ppm}/^\circ\text{C}) f$$

$$\left. \frac{\Delta f}{\Delta T} \right|_{\text{diffractive}} = (+11.3 \text{ ppm}/^\circ\text{C}) f$$

such that the values for ξ_r and ξ_d are closely matched in magnitude, but opposite in sign.

Let us design a hybrid lens with $f = 5 \text{ mm}$ using this material. Dividing the optical powers as described above, we need to solve Equations 8.91 and 8.92 simultaneously for f_r and f_d . Given $\xi_r = -10.4 \times 10^{-6} \text{ K}^{-1}$ and $\xi_d = 11.3 \times 10^{-6} \text{ K}^{-1}$ and aiming for $\xi_T = 0$, we find that

$$f_r = 9.6 \text{ mm} \quad \text{and} \quad f_d = 10.4 \text{ mm}.$$

Therefore, by designing the refractive lens to have a focal length of 9.6 mm and combining this in a doublet with a diffractive lens whose focal length is 10.4 mm, we expect a lens with no thermal variation of focal length.

There are numerous other examples of practical athermal hybrids using a variety of materials. As a further example, using semiconductors, an athermal microlens has been fabricated by etching the diffractive structure into InP on the rear of a laser substrate, providing an operating temperature regime of -20° to $+70^\circ \text{ C}$ with a shift in focus length of only $0.5 \mu\text{m}$ in that range (Fu and Bryan, 2002).

Achromatic hybrids

We saw in the previous section that chromatic aberration for diffractive lenses is of opposite polarity than that for refractive lenses: the former have $\nu_D < 0$, whereas for the latter, $\nu > 0$. As a result, combining refractive and diffractive structures in a single optical element also suggests new means for fabricating achromatic microlenses (Davidson et al., 1993).

Again, a refractive/diffractive doublet can be designed in which the optical power is shared between the two, but the CA of one is compensated for by the other. Relatively weak diffractive optical power is required, but the residual chromatic aberration in the visible wavelength range is larger than that achievable using the classical purely-refractive Fraunhofer doublet we saw in Section 7.5.2. For IR wavelengths, however, due to the materials available, a refractive/diffractive hybrid is comparable in performance to a Fraunhofer doublet. For the hybrid case, moreover, spherical aberration varies with wavelength.

Optically-corrected hybrids

The combination of refractive and diffractive effects can be used to conceive of a wide variety of further optical elements, with a combination of features and optimized parameters. Achromatic athermats, athermal DOEs for beam shaping, and triplets using two different materials and a diffractive structure have been demonstrated.

In addition, compensation for aberrations is possible using a hybrid lens. Spherical aberration may be corrected, as may field curvature using an etched binary “staircase”

structure, without adding any optical power (Sasian and Chipman, 1993). As a further example, a microlens compatible with both DVDs and Blu-ray discs has been developed using superimposed diffractive structures: the first diffraction order is used for the DVD, whereas the second order is for the Blu-Ray wavelength (Tanaka et al., 2004). This two-element lens, one of which is diffractive, is also designed for reduced chromatic aberration.

8.4.6 Diffractive optical elements

The Fresnel-like microlens was one example of employing a diffractive structure to perform a desired optical function, in this case focussing or imaging. Diffractive structures may also be used to generate virtually arbitrary intensity distributions in an image plane: it is “merely” a question of calculating the necessary phase function and then fabricating it. Such diffractive structures are generally referred to as diffractive optical elements (DOEs) or holograms¹⁰; as a special case, a two-dimensional diffractive optical element which is based only on the modification of the phase of an incoming field is often referred to as a kinoform, a structure which is optimized for diffraction into a single order, which may have a lensing or a more complex optical function (Lesem et al., 1969). As the reader probably suspects, a calculation of the form of DOEs is generally not trivial. For that reason, the journal and conference literature is replete with examples of diffractive optical elements and the means for their calculation, and these components are used to generate a wide variety of optical functions.

Calculation

The calculation of the two-dimensional profile of a DOE intended to transform the phase of an incoming optical field can be relatively complex and we only provide a heuristic view here; excellent in-depth studies are available to the interested student (Wyrowski, 1993; Goodman, 2004). Using the notation and approach of (Herzig, 1998, Chapter 1.2), the design of a DOE is based on a two-dimensional phase function, $\Phi(x, y)$ [rad], required to generate the desired optical intensity distribution in the image plane. Whereas a change in the intensity (through absorption) in the optical field might also be used in combination with a phase distribution to achieve a certain intensity distribution, the most efficient DOE structures are phase-only.

$\Phi(x, y)$ represents a phase added to the incoming optical field by the DOE, and may be generated, for example, by an optically transparent material whose thickness $t(x, y)$ [m] varies across the area of the element, giving rise to an optical path length distribution $nt(x, y)$ and thus a phase distribution, $k_0nt(x, y)$. As we saw for the Fresnel-like microlens above, we can wrap the phase to multiples of 2π , such that the diffractive microstructures can be thin and thus compatible with micro-optical process technolo-

¹⁰ The designation hologram in this case implies an optical structure which stores intensity and phase information. The popular (and original) types of holograms which generate three-dimensional images store the phase information about an illuminated object, the image of which is then reconstructed under illumination of the hologram (Gabor, 1948).

gies. As a result, we can write the fabricated phase function $\Psi(x, y)$ [rad] as

$$\Psi(x, y) = [\Phi(x, y) + \phi_0] \pmod{2\pi} \quad (8.93)$$

for a phase offset ϕ_0 . The layer thickness $t(x, y)$ which then has to be fabricated is given by

$$t(x, y) = \frac{\lambda_0}{n(\lambda_0) - 1} \frac{\Psi(x, y)}{2\pi} \quad (8.94)$$

for operation at the design wavelength λ_0 . The function $\Psi(x, y)$ is designed to yield the required phase shift of the wavefront $\Phi(x, y)$ in the first diffraction order, and, as we saw for the Fresnel-like lens, we can approximate an ideal profile using the techniques of binary optics; different kinoforms with varying numbers of steps can generate the same $\Psi(x, y)$, albeit with varying efficiency. The case $\Psi(x, y) = \Phi(x, y)$ implies that no phase wrapping is done, and the manufactured profile is exactly the design profile; for the Fresnel-like lens, this case would imply that the original refractive surface (without steps) is fabricated.

To generate an arbitrary intensity pattern, an arbitrary phase function must be determined, and this is usually given as a polynomial expansion of the form

$$\Phi(x, y) = C_0 \sum_m \sum_n a_{mn} x^m y^n \quad (8.95)$$

for C_0 a constant, such that the designer has to determine the coefficients a_{mn} . Whereas the relatively simple case of the Fresnel-like lens used only diffraction from the first order, arbitrary DOEs generally combine diffraction from several orders and analytic solutions are typically impossible. The iterative solutions required (see for example (Gale et al., 1993; Wyrowski, 1993)) usually start with a given desired intensity distribution and a guess for the phase at the image, then calculate the field in the plane of the DOE through a Fourier transform, add fabrication constraints (such as limits in DOE film height or achievable phase shift), then calculate a new image field, again by FT. This last field is compared with the ideal, and the process repeats. An example of the type of pattern which results is shown in Figure 8.29.

Applications

Diffraction optical elements have a very wide range of applications, since virtually any intensity and phase distribution can be generated using the appropriate $\Phi(x, y)$. Applications are, however, limited by the spectrum employed; since the DOE is designed for operation at a certain wavelength, efficiency can decrease rapidly as wavelength shifts.

Two areas in which DOEs are particularly useful are for beam shaping and generation of fan-outs. Beam shaping is most often used in conjunction with laser beams, in which a non-ideal intensity profile can be optimized (Leger, 1998). A Gaussian beam, for example, can be formed into a flat-topped profile with steep sides to allow homogenous illumination (Romero and Dickey, 1996). In addition, beam-flattening can be used for optimizing the profile of the beam inside a Nd:YAG laser cavity (Kermene et al., 1992);

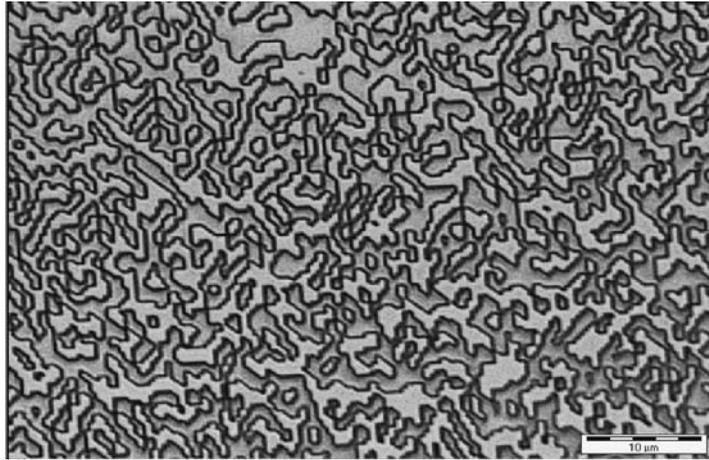


Figure 8.29 Portion of a binary diffractive optical element. Photo courtesy of Heptagon Oy.

the DOEs are placed inside the cavity, next to or etched onto the laser mirrors, and result in an emitted beam with considerably reduced divergence.

Fan-out elements are those which convert a point source or planar wavefront into a controlled number of one- or two-dimensionally distributed intensity peaks. Continuous-relief diffractive optics have been developed generating two-dimensional intensity arrays (Gale et al., 1993) and a variety of kinoforms for fan-outs have been designed and employed (Herzig et al., 1990; Prongué et al., 1992; Stigwall and Bengtsson, 2004).

8.5 Diffractive micro-optics: case studies

We have seen in the previous sections that diffraction is a multi-faceted phenomenon with a wide variety of manifestations: whereas diffraction gives rise to limitations on the sizes of apertures and the corresponding quality of an optical image, it may also be beneficially used to perform a wide variety of micro-optical functions.

Some of the applications have been mentioned in the previous sections. Although it is difficult to rank these in order of relevance or application, micro-optics benefits greatly from the ability of using diffractive optical elements, or holograms, for the generation of virtually arbitrary intensity profiles (Dai et al., 2003). A further important area, both academically and industrially, is the use of binary phase gratings, often referred to as Dammann gratings (Dammann and Görtler, 1971) in honor of their inventor, for beam splitting and generation of multiple intensity peaks from uniform or point illumination sources (Walker and Jahns, 1990; Mait, 1997).

The spectrum of other uses for diffractive optics is large. MEMS-based micro-optical switches using movable gratings have been demonstrated for use in optical communications systems (Liu et al., 2000), and the so-called grating light valve has been extensively developed for optical modulation and displays (Solgaard, 2009, Chapters 10 and

11). We have seen how micromachined grating structures can be used for monitoring position or movement using, for example, features etched into scanning mirrors (which we saw in Figure 6.11), as discussed in Section 6.6.2. Finally, in contrast to their utility in the formation of images, diffractive optics may also be used in optical analysis, since a wide variety of filters for image processing may be conceived (Schwarzer and Teiwes, 1997).

We take a small sampler here, considering two applications as case studies to provide the reader with a taste for what is possible: diffractive wavelength-selective mirrors and grating-based microspectrometers.

8.5.1 Wavelength-selective mirrors

The diffraction grating of Section 8.3 is possibly the most generally useful and widely applied diffractive structure and provides the basis for a host of micro-optical devices. The most widely applied of these is likely to be the wavelength-selective mirror, which forms the functional basis for entire families of semiconductor lasers.

We already saw the wavelength-selective mirror in the discussion concerning optical multilayers. Recall that in Section 5.3.5, Bragg filters and mirrors were presented, based on a stack of thin film layers with periodically varying refractive index. This so-called Bragg stack may be considered to be a one-dimensional grating, for which propagation is perpendicular to the plane of the substrate. We saw in Section 5.3 that the reflectance is a function of wavelength and is defined by the periodicity of the Bragg layers.

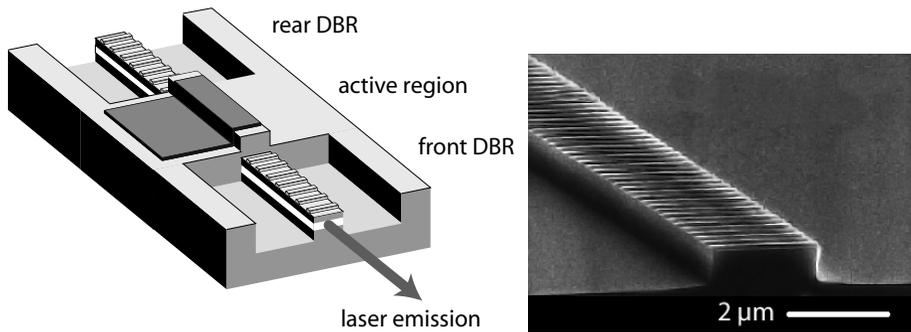
Distributed feedback lasers

Distributed feedback lasers represent a family of semiconductor lasers which use periodic grating structures to generate optical feedback into the laser cavity; rather than classical mirrors, they use Bragg reflectors as wavelength-selective mirrors (Agrawal and Dutta, 1986; Morthier and Vankwikelberge, 1997; Carroll et al., 1998). We will examine semiconductor lasers in more detail in Chapter 10, and the waveguides on which they are based in Chapter 9, but we consider the application of the diffractive microstructure as a laser mirror as an example of the application of diffraction here.

The distributed Bragg reflector (DBR) laser, shown schematically in Figure 8.30(a), confines the optical field in a waveguide onto which a diffraction grating structure has been etched, as seen in the SEM photo of Figure 8.30(b). The optical field in the waveguide couples to the periodic grating structure, and it can be shown (Zappe, 2004, Chapter 4.4.2) that this coupling results in the generation of two modes, one which propagates forward in the waveguide and the other backwards; this latter mode is the reflected field which provides feedback into the cavity.

Based on the same principles as the periodic multi-layer interference filters of Section 5.3.5, the reflection peak for a DBR grating with period Λ occurs at a wavelength

$$\lambda_{DBR} = \frac{2N\Lambda}{m} \quad (8.96)$$



(a) A DBR laser shown schematically, with the grating on either side of the central active region. Not to scale. (b) SEM photograph of the grating on the waveguide of a DBR laser. A magnified view of this structure was shown in Figure 8.22.

Figure 8.30 The DBR laser uses a diffraction grating etched onto the laser waveguide as a wavelength-selective mirror.

where $m = 1, 2, 3, \dots$ is the order of the grating and N [] the effective waveguide index, to which we will be introduced in the next chapter. Thus, in contrast to the standard mirror with a broad, flat reflectance spectrum, the DBR mirror reflects only in a narrow band of wavelengths, thereby assuring single-mode operation and freedom from mode-hops.

The distributed feedback (DFB) laser uses the same principle, but in a different configuration. Rather than employing a pair of gratings as wavelength-selective mirrors at either end of the active region, the DFB laser combines the active region and grating reflector as one. The same form of modal coupling results in two counter-propagating modes, providing the feedback necessary for laser action.

The grating structures employed for DBR lasers are typically configured as shown in Figure 8.30. First order ($m = 1$) gratings, etched into the compound semiconductor materials generally employed, typically have a period of about 100 nm (for the near infrared (NIR) region, around 800 nm), to up to about 230 nm for the 1.55 μm wavelength regime. The grating depth varies, but is typically about 100 nm (as seen in Figure 8.22). The length of the grating section for the DBR laser is on the order of 50 – 200 μm , and for the DFB, the combined grating/active region is generally 300 – 600 μm long. The patterns are usually defined holographically (Hofstetter et al., 1994) or by e-beam. More advanced structures using curved grating reflectors have been used for focussing an vertically-emitted beam or even for generating squeezed light (Uemukai et al., 2005).

VCSELS

The vertical-cavity surface-emitting laser (VCSEL) is also based on a Bragg mirror for wavelength-selective reflection into the active region (Li and Iga, 2003). As seen in Figure 8.31, the laser cavity is now turned onto its side and the Bragg mirrors are formed by pairs of alternating materials with different refractive indices.

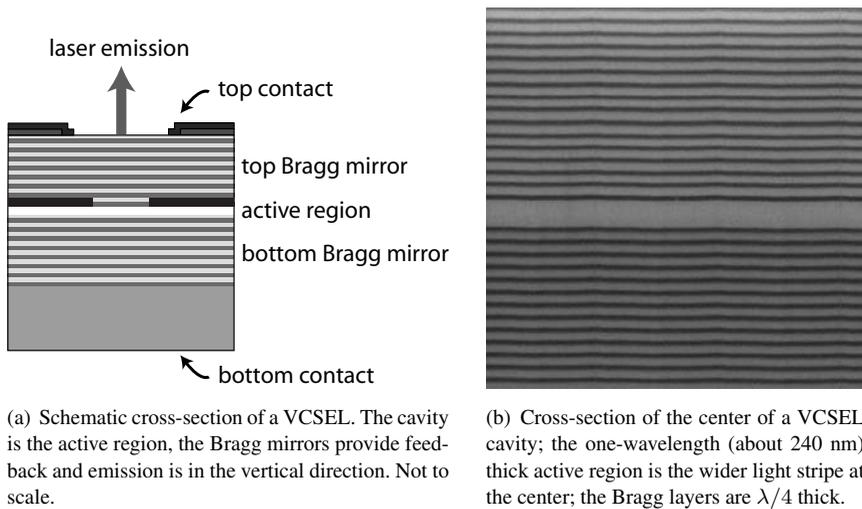


Figure 8.31 The VCSEL is based on two epitaxially grown Bragg mirrors, with an active region (in which light is generated) sandwiched in between. A typical laser has 40 mirror pairs below and 25 pairs above the active region.

The VCSEL cavity is composed of two Bragg mirrors, with individual layers $\lambda/4$ thick (recalling that $\lambda = \lambda_0/n$ for refractive index n of the semiconductor material, where typically $n \approx 3.5$ for the III-V semiconductors employed) and a one-wavelength thick active region, in which light is generated (Rakic and Majewski, 2003). Due to the small active volume, a very large mirror reflectance at the lasing wavelength is required, such that the VCSEL mirrors, like the thin-film interference filters, have a large number of mirror pairs to provide a sharp wavelength resonance.

Analogously to Equation 5.113, the Bragg reflectance of the VCSEL mirror can be written in the form (Wilmsen et al., 2001, Chapter 3)

$$R_{Bragg} = \left[\frac{1 - \left(\frac{n_H}{n_L}\right)^{2p}}{1 + \left(\frac{n_H}{n_L}\right)^{2p}} \right]^2 \quad (8.97)$$

for p mirror pairs, with high and low refractive index n_H and n_L , respectively. We can see from the previous expression that the reflectance increases and slowly saturates with an increasing number of mirror pairs. The limiting factor in pair number for the VCSEL is generally the total mirror thickness, since the pump current for the laser needs to traverse the mirrors to reach the active region, and a thick mirror leads to a high ohmic resistance.

The Bragg mirror is not necessarily the only portion of the VCSEL which takes advantage of diffractive effects. Since most VCSELs have a circular aperture, as we will see in Chapter 10, the polarization direction of the emission is not defined, or, even worse, can rotate or jump spontaneously. For polarization-stable operation, it has

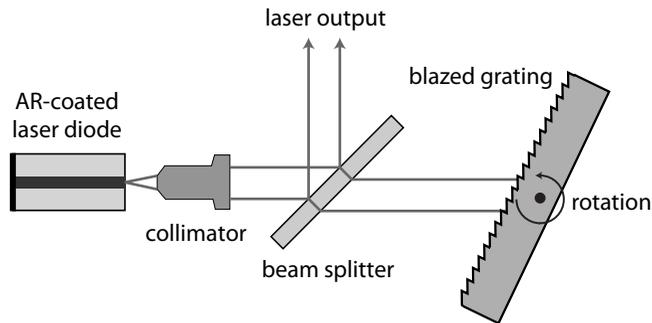


Figure 8.32 The external cavity tunable laser employs a laser diode, for which one facet (which normally acts as a mirror) has been AR-coated, and reflection into the active region (the laser diode itself) is from the Littrow-mounted, rotatable grating. As the grating rotates, the wavelength reflected back into the cavity changes, resulting in tuning of the laser frequency.

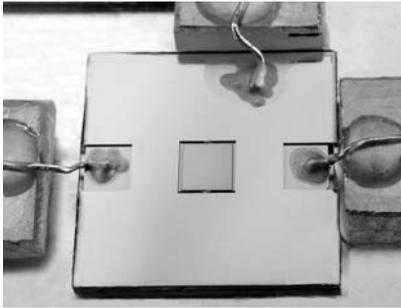
been shown that etching a diffraction grating structure onto the emission facet of the VCSEL itself (thus, on the top surface, where the laser emission takes place, as seen in Figure 10.17(a)) can result in a defined and stable polarization direction (Debernardi et al., 2005; Chung et al., 2008); spurious polarization modes can be suppressed by up to 20 dB (Haglund et al., 2006). Taking the technique one step further, a single sub-wavelength high-contrast grating can replace the entire top Bragg mirror (Huang et al., 2007), providing efficient feedback into the laser cavity as well as polarization stability.

Micro-external cavity lasers

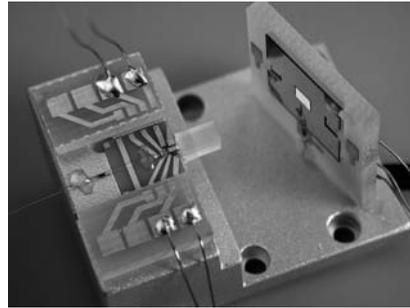
We saw in Section 8.3.5 that a blazed grating, especially when mounted in the Littrow configuration, provides an efficient, wavelength-dependent reflection of an optical field, back into the direction in which it came. This functionality has long been applied to the fabrication of macroscopic external-cavity tunable lasers, in which one mirror of an active laser cavity is replaced by a Littrow-mounted grating. By rotating the grating, the reflected wavelength changes, allowing a tuning of the resonator and thus the lasing mode; such an arrangement is shown schematically in Figure 8.32. We thus have an example of a wavelength-selective mirror which is tunable through physical movement, in this case rotation.

Employing developments in micro-optics and microsystems engineering, the concept of the external cavity laser has been extended to MEMS-based systems (Liu and Zhang, 2007). Whereas macroscopic concepts have dimensions in the centimeter (or larger) range, micro-optical external-cavity lasers can be reduced in size to just larger than the laser chip (several hundred micrometers) and can have cavity lengths down to several $10\ \mu\text{m}$, the latter implying a large mode spacing, not achievable with classical assembly methods.

Although MEMS-based external-cavity lasers can employ standard Fabry-Perot cavities, whose length is changed by micromechanical means, the arrangement which best results in stable, single-wavelength, and narrow-linewidth operation is the grating-based



(a) Detail of the electrostatically actuated silicon-based tiltable grating, with contact wires; the dimension of the grating, the square at the center of the substrate, is $2 \times 2 \text{ mm}^2$.



(b) The grating is mounted in the Littrow configuration at the right of the setup; on the left is the laser diode, coupled with a cylindrical lens to its right. The output optical fiber is seen emerging from the left of the laser chip.

Figure 8.33 A miniaturized external-cavity tunable laser based on an electrostatically-actuated silicon grating, operating in the $2.1 \mu\text{m}$ wavelength range. Photos courtesy of Eva Geerlings and the Fraunhofer IAF.

approach shown in Figure 8.32. A blazed grating provides the highest efficiency, but the appropriately shaped grating can only be practically fabricated in silicon by generating the grating vertically and operating in the plane of the wafer (i.e., a beam at normal incidence onto the grating is in the plane of the substrate). However, using cleverly-designed mechanical linkages and electrostatic actuation, such a 12th-order Littrow-mounted blazed grating has been demonstrated in a MEMS-based system, resulting in a tuning range of 120 nm in the $1.55 \mu\text{m}$ wavelength range (Syms and Lohmann, 2003; Lohmann and Syms, 2003). The grating “teeth” extend into the substrate and are about $150 \mu\text{m}$ long; the grating shape, with a blaze angle of 45° , may thus be defined photolithographically.

Alternatively, a hybridly-assembled microsystem using out-of-plane incidence onto a scanning silicon micro-grating has demonstrated a large tuning range for a micro-external-cavity laser, using a rectangular grating also mounted in the Littrow configuration (Geerlings et al., 2006, 2008). Since the grating “teeth” are in this case defined in the plane of the substrate (i.e., a beam at normal incidence onto the grating is normal to the plane of the substrate), holographic exposure could be used to generate a regular grating pattern, with $\Lambda = 3.3 \mu\text{m}$ and a depth of 400 nm, over an area of $2 \times 2 \text{ mm}^2$, as seen in Figure 8.33(a). Designed for operation in the $2.1 \mu\text{m}$ wavelength range, the rectangular (unblazed) grating still provided an efficiency of $\eta \approx 0.81$ in the -1 order.

The microsystem was assembled in the configuration seen in Figure 8.33(b); the Littrow-mounted grating is at the right, onto which the emission from an AR-coated

GaSb-based laser diode chip at the center is coupled, using an anamorphic¹¹ cylindrical lens for beam collimation. Laser output is at the left, using an optical fiber. The grating consists of a spring-mounted membrane with dimensions of $2 \text{ mm} \times 2 \text{ mm} \times 5 \text{ }\mu\text{m}$; its angle with respect to the optical axis can be scanned by electrostatic actuation using electrodes defined on the Pyrex substrate on which the silicon wafer is mounted. Using an actuation voltage of $\pm 165 \text{ V}$, a tilt angle range of $\pm 3.1^\circ$ could be achieved. In the $2.1 \text{ }\mu\text{m}$ wavelength range at which the laser was operated, this grating movement provided a tuning range of 152 nm , and was limited by the gain spectrum of the laser diode, not the grating movement. A tunable laser in this wavelength range is of considerable utility for non-invasive measurement of physiological parameters, in particular blood glucose levels, and provides an excellent example for the combination of MEMS, diffractive- and micro-optics technologies for the conception of miniaturized optical systems.

8.5.2 Mini-spectrometers

One of the primary reasons that the classical macroscopic diffraction grating was so avidly developed was for its potential use in spectroscopy; grating-based spectrometers are today the workhorses of most optics laboratories. Using advanced manufacturing techniques, mini-spectrometers have also been developed and a variety of miniaturized versions are now commercially available. Developments in microsystems fabrication technologies have given rise to numerous forms of micro-spectrometer concepts, which use microfabricated gratings and either micromechanical actuation or assemblies which do not require any component movement. The most interesting concepts are based on silicon or polymers.

Resolution

The resolution of a spectrometer, R [], is given by

$$R = \frac{\lambda}{\Delta\lambda} \quad (8.98)$$

where $\Delta\lambda$ [m] is the FWHM of the minimum spectral width which can be resolved at a wavelength λ . High-end spectroscopy applications, such as for chemical analysis or astronomy, require $R > 10^5$; micro-spectrometers cannot approach this range.

The reason for the resolution limit is the device size. The spectral resolution of a grating is proportional to the number of illuminated grating periods, N_i [], so that $R \approx mN_i$ for mode index m (Born and Wolf, 1980, Chapter Chapter 8.6). This approximation assumes, however, that the detector is in the far-field (Fraunhofer regime, such that Equation 8.1 holds) with respect to the grating. As a result, for the illumination of N_i

¹¹ Anamorphic optical systems or components are those with different magnifications in the two directions normal to the optical axis. These are frequently required for collimation of emission from laser diodes, since they have strongly elliptical emission.

periods of a grating with period Λ at wavelength λ , the detector must be spaced at least

$$L_{det} = \frac{(N_i \Lambda)^2}{\lambda} \quad (8.99)$$

from the grating. We can thus estimate the resolution of a micro-spectrometer to be

$$R \approx m N_i \approx m \frac{\sqrt{\lambda L_{det}}}{\Lambda}. \quad (8.100)$$

Using values typical for micro-spectrometers, the previous expression predicts values of R in the range 5 to 10, a far cry from 10^5 .

There are significant application areas, however, for which $R < 100$ is sufficient, and these are regimes in which microfabricated spectrometers play a useful role; typical uses include environmental sensing, color analysis or industrial monitoring. The limitations due to reduced resolution are offset by the advantages of extremely small size and, due to their structure, very robust nature.

Grating micro-spectrometers

Numerous types of micro-spectrometers have been developed, using various principles (Wolffenbuttel, 2004). Microfabricated and actuated scanning Fabry-Perot interferometers (Tran et al., 1996) and scanning Fourier Transform spectrometers (Manzardo et al., 1999; Wallrabe et al., 2005) have experienced some degree of popularity. Grating-based designs typically employ microfabricated diffraction gratings in arrangements which do not require movement, and instead use the dispersive properties of the grating to distribute the optical field across a photodetector array.

One example uses a fixed grating mounted over an array of thermopiles, efficient detectors for the mid-IR range, 2 – 5 μm (Kong and Wolffenbuttel, 2005). The structure, with an overall size of 5 \times 5 \times 1 mm^3 , is fabricated using two bonded Si wafers and demonstrated a resolution of $R \approx 10$ in this wavelength regime. For visible wavelengths, a 3 \times 3 \times 11 mm^3 sized module employed two diffraction gratings, one for dispersion, the other for focussing the spectrum onto a CCD chip (Grabarnik et al., 2007); 3 nm resolution was achieved in the wavelength range $450 \leq \lambda \leq 750 \text{ nm}$.

An in-plane design, using SiON/SiO₂ slab waveguides, took advantage of the fact that propagation of the field in the plane of the substrate allows definition and fabrication of highly accurate blazed gratings, since the gratings are oriented normal to the surface (Sander and Müller, 2001). In this case, the optical field is coupled into a waveguide, which is transparent for the visible wavelengths for which the spectrometer is designed, $350 \leq \lambda \leq 650 \text{ nm}$; the blazed grating, with a feature size on the order of 1 μm , implying ease of fabrication using standard lithography techniques, has a high efficiency in the -1 order and leads to a wavelength resolution of about 9 nm. The light path through the chip is complicated, using multiple reflections from curved and focussing mirror facets, and emerges, spectrally dispersed, onto a photodiode array.

A variety of other grating-based, microfabricated spectrometer concepts have been demonstrated, including the use of movable lamellar gratings as a means of realizing a Fourier-transform spectrometer (Hongbin et al., 2008) or liquid crystals as a dynamically-reconfigurable diffractive element (McMurdy et al., 2006). Customized

multi-level gratings which simulate absorption spectra of certain chemical species have been proposed for optical chemical sensors (Sinclair et al., 1997) and programmable and movable diffractive gratings for spectral analysis as well as displays have been demonstrated (Sagberg et al., 2004). Finally, gratings have been etched directly onto refractive microlenses (Traut et al., 2000), resulting in hybrid optical elements which accomplish dispersion and focussing simultaneously.

Problems

1. Plot the intensity of single-slit diffraction patterns as a function of diffraction angle in the range $-25^\circ \leq \theta_o \leq +25^\circ$ for three slits, with $a = 2, 4, 16 \mu\text{m}$, illuminated with a HeNe laser emitting at 633 nm.
2. Plot the intensity of single-slit diffraction patterns as a function of diffraction angle in the range $-25^\circ \leq \theta_o \leq +25^\circ$ for a single slit, with $a = 5 \mu\text{m}$, illuminated at wavelengths of 400, 500 and 600 nm.
3. Compare the diffraction patterns generated by two two-slit configurations, both illuminated with a green HeNe laser emitting at 543 nm and with slit widths of $5 \mu\text{m}$, but one with a slit spacing of $7.5 \mu\text{m}$ and the other with $15 \mu\text{m}$.
4. Plot and compare the diffraction intensities for the clear and shaded apertures of Example 8.2 and verify apodization.
5. Show that evaluation of the integral sum in Equation 8.22 leads to Equation 8.23.
6. Compare the diffraction patterns generated by two multiple-slit apertures, both with $a = 10 \mu\text{m}$ and $s = 25 \mu\text{m}$ and illuminated at 633 nm, one with $N = 3$ and other with $N = 15$ slits. For sharp diffraction peaks, do we prefer few or many slits? What happens if we illuminate 100 slits?
7. Design a multiple-slit aperture for use at 633 nm where the angular spacing between the principle and first satellite peak is 4° and its intensity is 80% of the principle peak. Describe qualitatively the effect of each variable parameter on the diffraction pattern.
8. Calculate the period Λ in micrometers for gratings which are characterized by 30 grooves/mm (typical for commercial blazed gratings), 1 500 grooves/mm and 5 800 grooves/mm (about the highest line density manufacturable using ruling).
9. Derive the expression for the Airy disk radius, Equation 8.46. Is the factor 1.22 exact?
10. Sketch the wavefronts incident onto a grating and derive the grating equation, Equation 8.48, from the resultant geometry, realizing that diffraction peaks occur at angles for which constructive interference takes place.
11. Derive and plot the variation of diffraction angle with period (in the range $0.5 \mu\text{m} \leq \Lambda \leq 5.0 \mu\text{m}$) for the ± 1 orders generated by a 543 nm HeNe laser incident onto a reflection grating at 22.5° . For what value of the period do the diffraction orders disappear?
12. Derive and plot the variation of diffraction angle with wavelength over the entire visible range for light incident onto a grating with $\Lambda = 1.2 \mu\text{m}$ at an angle of -15° , for the ± 1 orders.

13. In the previous problem, is the diffraction angle shift with wavelength the same for the two orders?
14. Consider a transmission grating with a period of $\Lambda = 1.5 \mu\text{m}$ onto which a 633 nm HeNe laser beam is incident at 35° . Determine the number of visible diffraction orders and calculate their spacing from the zero order when imaged on a screen 20 cm from the grating.
15. We illuminate a transmission grating with emission from a discharge lamp containing an unknown substance. The grating has a period of $\Lambda = 6.5 \mu\text{m}$, and on a screen placed 1 cm behind the grating, the second order diffraction spot is found at a spacing of 1.847 mm from the zero order. What is probably the substance in the lamp?
16. Show that the two expressions for grating angular dispersion, Equations 8.64 and 8.66, are equivalent.
17. Given a grating with period $\Lambda = 5 \mu\text{m}$ illuminated with a 690 nm laser diode, calculate and compare the angular dispersion for $m = 3$ using Equation 8.66 and that which results by taking the difference in diffraction angle which results when varying the wavelength a small amount around 690 nm.
18. Calculate and plot to scale the profile of the first four zones of a Fresnel-like microlens fabricated in photoresist ($n = 1.56$) which has a maximum thickness of $2.5 \mu\text{m}$, designed for a focal length of $500 \mu\text{m}$ for operation at the red HeNe wavelength, 633 nm.
19. Design an athermal refractive/diffractive doublet lens with a focal length of 10 mm using polycarbonate, given that $\xi_r = 246 \times 10^{-6} \text{K}^{-1}$ and $\xi_d = 131 \times 10^{-6} \text{K}^{-1}$.
20. How many layer pairs of SiO_2 and Si_3N_4 are required for a Bragg stack to achieve $R \geq 0.99$? What is the thickness of this stack if we operate it at $\lambda = 1.55 \mu\text{m}$?
21. Using a grating spectrometer, you would like to distinguish between the sodium D1 and D2 lines (589.592 nm and 588.995 nm, respectively). Given an arrangement for which the input beam impinges onto the grating at 20° , and the detector is angled 45° from the input, what is the required grating period for operation in the first diffraction order? How many periods of the grating need to be illuminated to distinguish the two lines?
22. For the previous problem, by how much does the grating have to be rotated such that the $m = 2$ order is directed onto the detector?
23. Still considering the arrangement of the two previous problems, determine the FSR of the grating spectrometer for visible wavelengths for $m = 1, 2$.

9 Guided-wave micro-optics

*La musique est l'arithmétique du son, comme l'optique est la géométrie de la lumière*¹

Claude Debussy

The optical components and systems we have considered to this point belong to the realm of *free-space optics*, meaning the optical field is transmitted through a homogeneous medium (such as vacuum or air) without being guided in any manner. The lenses, mirrors and diffractive structures we have studied are intended to modulate the unguided beams of free-space optics.

A significant branch of optics, however, uses structures to guide an optical field using reflections at an optical interface. The optical structures which accomplish this, waveguides and optical fibers, may be considered to be a form of “optical wire”, in which the field is confined to a small area in two dimensions and transmitted along the third. Waveguides are micro-optical components for light transmission which are typically fabricated on glass, plastic or semiconductor substrates, giving rise to the concept of *integrated optics*, analogously to integrated electronics (Miller, 1969b). When combined with integrated active optical devices, such as lasers, modulators or detectors, also waveguide-based, such guided wave chips form the basis for the field of *photonics*. Finally, optical fibers span several orders of magnitude in their dimensions, with micrometer-sized guiding regions and hundred kilometer lengths; they are also useful as light guides in hybrid micro-optical systems.

We consider guided-wave optics in this chapter, beginning with an overview of how light is transmitted in a waveguide, and introduce the discrete propagation modes which result. Slab and channel waveguides, including their propagation characteristics, structures and means for characterization, will be extensively discussed, as will optical fibers, conceptually quite similar to waveguides. We will see how waveguides and fibers play an important role in small- as well as large-scale systems by concluding the chapter with a look at a few micro-optical case studies, giving examples of the use of guided wave optics in telecommunications and sensor applications.

The literature offers numerous in-depth treatments of integrated optics, including the classic (Marcuse, 1974), the didactically excellent (Hunsperger, 2002) or (Iizuka,

¹ “Music is the arithmetic of sound, as optics is the geometry of light.” Claude Debussy (1862–1918), French composer, was inspired to this analogy by the music of Jean-Philippe Rameau.

2002b), as well as a focus on semiconductor integrated optics (Zappe, 1995) and integrated optical circuits (Nishihara et al., 1989). Optical fibers are the subject of (Derickson, 1998) and (Bass and van Stryland, 2001), with consideration of their use in optical communications in (Rogers, 2001) and subject to an engaging historical survey in (Hecht, 1999).

9.1 Waveguides: ray optic model

Guided-wave optics relies on reflection from closely-spaced surfaces. Whereas reflection may be achieved using mirror surfaces, such as we studied in Chapter 6, most waveguides and fibers use total internal reflection, to which we were introduced in Section 4.4.2. As a result, numerous concepts we introduced in Chapter 4 will become important in the current discussion.

We will begin with an analysis of waveguides based on a ray-optic model, which treats the optical field propagating through the guide as composed of rays, which move in straight lines and reflect at interfaces, as we used in ray tracing through refractive lenses or the in analysis of mirrors. We ignore the electromagnetic fields until Section 9.2, where we will employ an electromagnetic model to reach the same conclusions. The ray-optic model is conceptually simple, and thus useful for a first analysis, but somewhat limited in physical rigor.

9.1.1 Guided waves

Guiding light using a solid transparent structure, a light-guide, is an effect which has been well-known for hundreds of years. A light-guide is a structure as shown on the left of Figure 9.1: a long, transparent rod with cross-sectional dimensions typically in the millimeter range. If the rod is not bent too tightly, light coupled into one end is confined inside the material by total internal reflection at the inside interface. As a result, a light-guide may be used to guide and transmit optical power; for suitably flexible materials (plastics, or glass if sufficiently thin), the light may thus be guided around corners. Whereas the concept of light-guiding was analyzed scientifically as early as 1842 (Colladon, 1842), the eighteenth-century applications of light-guides were predominantly for light shows and other forms of entertainment (Hecht, 1999, Chapter 2). The light-guide concept is still important today as the basis for specialized forms of illumination, for example in endoscopy (Maeda et al., 1996).

If we shrink the lateral dimensions of the light guide, it eventually becomes a waveguide. As we will see in detail in the following sections, the physics of light propagation changes as the transverse dimensions of the light-guide decrease to the order of the wavelength of light, as seen at the right of Figure 9.1. Total-internal reflection still confines the optical field inside the transparent waveguide, but propagation takes place in discrete modes and has a strong wavelength dependence. This means of optical propagation is the essence of guided wave optics, and we will consider its physical basis and implications in the next few sections.

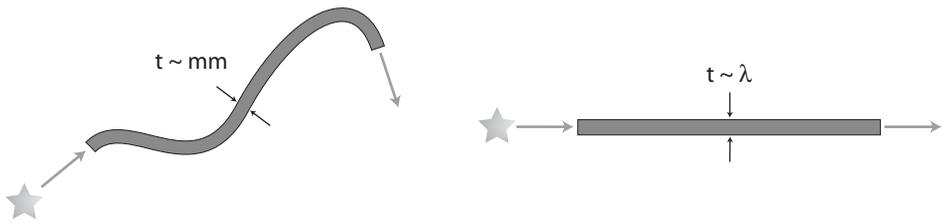


Figure 9.1 A light-guide *left* is a transparent rod, with transverse dimensions in the millimeter or larger range, whereas a waveguide *right* has transverse dimensions on the order of the wavelength, typically micrometers. Both structures are typically considerably longer than they are wide.

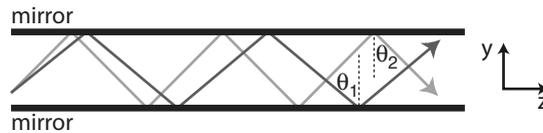


Figure 9.2 Two-dimensional cross-section of a mirror-based waveguide. Light propagates in the z direction by multiple reflections from the top and bottom mirrors.

9.1.2 Mirror waveguide

A first analysis of propagation in a waveguide can be done using a simplified representation, namely that of two parallel mirrors as shown in Figure 9.2. Although this is not a terribly realistic structure, it allows a simplified analysis of guided wave propagation. In this cross-sectional view, the mirrors are closely-spaced in the y direction, and we allow light to propagate as shown in the z direction. Since both the x and z directions extend to infinity, waveguiding is in an infinite two-dimensional slab limited only in the y direction, hence the designation “slab waveguide.” We will consider the “channel waveguide,” in which light is guided along a single direction, in Section 9.3.

If we allow optical fields to propagate from the left, as shown for two multiply-reflecting wave vectors in the figure, a first important question becomes: even though we recall that, for perfect mirrors, reflection angle equals incidence angle, are any values of these angles, θ_1 or θ_2 , allowed in a waveguide? To analyze this propagation condition, we must consider in detail what happens when the waves undergo multiple reflections, which we may do with the aid of Figure 9.3. We have sketched a series of parallel wavefronts moving at an angle to the z direction and two wave vectors normal to these.

Consider the wave vectors propagating from the positions 1 and 2 in the figure; at the positions A and B on the wavefront, these are in phase. The upper beam (from position 1) is then reflected at the upper and lower mirrors (both times at angle θ) and then forms a new wavefront at position C. The lower (from position 2) propagates without reflection and also forms part of the wavefront at C. For propagation in the waveguide to take place, these two beams must be *in phase* at the new wavefront at C, so that the path length difference of the two vectors must be an integral multiple of 2π . Were this not

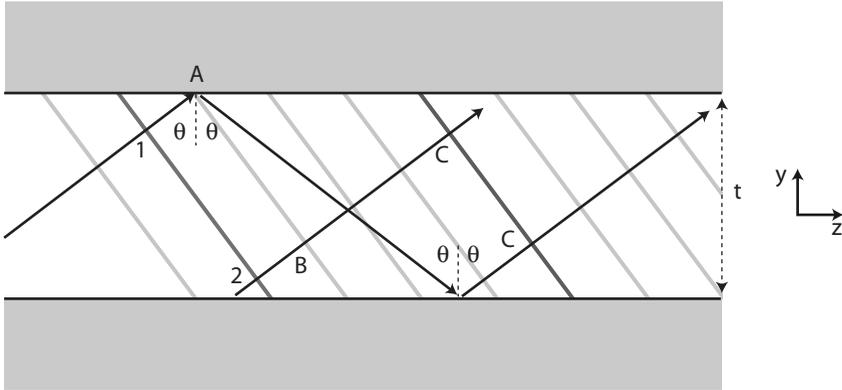


Figure 9.3 Wave propagation in a waveguide: the parallel wavefronts *gray* are normal to the wave vectors originating at points 1 and 2 *black*. For phase matching during propagation, the phase along any wavefront must be constant or shifted by integer multiples of 2π .

the case, destructive interference between reflected and un-reflected beams in the waveguide would result in the degradation of the propagating field after a few reflections.

The path-length difference between wave vectors 1 and 2, ΔL [m], is then given by

$$\Delta L = AC - BC \quad (9.1)$$

corresponding to $\Delta L/\lambda_0$ wavelengths, assuming vacuum between the mirrors. For mirror spacing t [m], a bit of geometry shows that

$$\Delta L = AC - BC = 2t \cos \theta \quad (9.2)$$

such that multiplying by the propagation constant $k_0 = 2\pi/\lambda_0$ yields the phase shift due to the longer path length of beam 1,

$$k_0 \Delta L = 2k_0 t \cos \theta = \frac{4\pi t}{\lambda_0} \cos \theta. \quad (9.3)$$

We recall furthermore from Section 4.2.3 that a phase shift of π occurs upon reflection from a mirror surface, such that beam 1 experiences an additional 2π of phase shift due to the two reflections. For the two waves to be in phase on the wavefront C, we thus require that the condition

$$\frac{4\pi}{\lambda_0} \cos \theta - 2\pi = 2m\pi \quad (9.4)$$

holds; $m = 0, 1, 2, \dots$ is an integer, usually referred to as the mode number.

Discrete propagation

Equation 9.4 is the so-called characteristic equation of a mirror waveguide; if Equation 9.4 is satisfied, the optical field propagates through the waveguide. We see, however, that this restriction implies that only certain values of reflection angle θ are allowed,

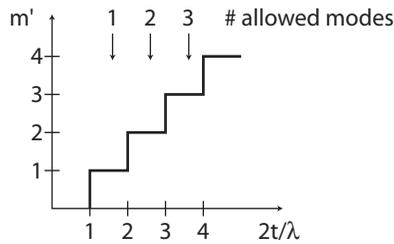


Figure 9.4 The number of allowed modes, given by the index m' , as a function of the ratio $2t/\lambda$. For $2t/\lambda < 1$, no mode is allowed and the waveguide is cutoff.

namely the discrete values

$$\theta = \arccos\left(\frac{\lambda_0 m'}{2t}\right) \quad (9.5)$$

where we define for convenience $m' = m + 1 = 1, 2, 3, \dots$; this last equation was simply derived by solving Equation 9.4 for reflection angle. Not only is propagation restricted to discrete θ , there is also a maximum value for wavelength, for a given spacing t , since the argument of the inverse cosine must be less than or equal to unity, implying

$$\lambda_0 \leq \frac{2t}{m'}. \quad (9.6)$$

Alternatively, for a given wavelength, we can express this relationship as a minimum value for t , namely

$$t \geq \frac{m' \lambda_0}{2}. \quad (9.7)$$

We see from the relationships above that propagation in a waveguide is considerably different from propagation in free space. Most importantly, propagation in a waveguide is not continuous, with all wavelengths and wave vectors allowed, but instead takes place in discrete modes; only certain combinations of reflection angle, wavelength and waveguide thickness are allowed. The number of allowed modes for a certain waveguide depends on the ratio $2t/\lambda$, as shown in the sketch of Figure 9.4. As $2t/\lambda$ increases, more modes allowed, and the mode number m' defines the m^{th} order mode.

We see from the figure that for $2t/\lambda < 1$, no mode is allowed. This situation corresponds to cutoff for a waveguide, implying that there is no propagation at all. Cutoff occurs if either, for a given waveguide thickness, the wavelength is too long or, conversely, for a given wavelength, if the waveguide is too thin. This type of behavior is quite distinct from that which we experience in free space.

Example 9.1: Let us consider a mirror waveguide with a spacing of $t = 0.5 \mu\text{m}$, with vacuum in the separation. How many modes propagate if we couple in a blue laser at 405 nm? A red laser at 690 nm? An infrared laser at 1.55 μm ?

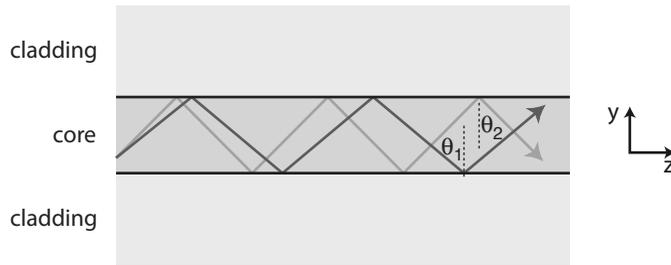


Figure 9.5 Cross-section through a dielectric slab waveguide showing the waveguiding core region at the center surrounded at the top and bottom by cladding layers. Guiding is by total internal reflection at the core/cladding boundaries.

By calculating $2t/\lambda$ for these three cases, we determine values of 2.47 for 405 nm, 1.45 for 690 nm and 0.65 for 1.55 μm . Since the number of allowed modes is the integer value of these results, we see that two, one and zero modes, respectively, propagate at those wavelengths. For the infrared laser, the waveguide is completely cutoff since no modes are allowed.

Summary

In summary, we have seen that propagation in a waveguide takes place in a fundamentally different manner than in bulk material or free space. Guided wave propagation takes place in discrete modes, defined by a mode number m' and these have individual values of wavelength, propagation constant and effective index. We will now extend this treatment to the more physically-realistic dielectric waveguide.

9.1.3 Dielectric waveguide

The mirror waveguide represents a simple model which allows us to understand intuitively why guided-wave propagation takes place in discrete modes, and how we characterize these; it is, however, not entirely physically realistic. An optical waveguide structure as it exists in reality is generally a sandwich of transparent dielectric layers, as shown in Figure 9.5. The optical field is guided in the center core layer and confined by the cladding layers on either side. We continue to consider the slab waveguide here, meaning that light is confined to a two-dimensional slab extending infinitely into the x and y directions.

The core layer has a higher refractive index than the cladding layers, such that reflection at the boundaries is due to total internal reflection, to which we were introduced in Section 4.4.2. The reliance on TIR represents an important distinction between this and the idealized mirror waveguide. We will see a number of important further differences: the wave extends partially beyond the interface and into the cladding, the phase shift is angle-dependent, the optical field experiences losses due to absorption in the cladding, and unguided modes can leak into the cladding layers.

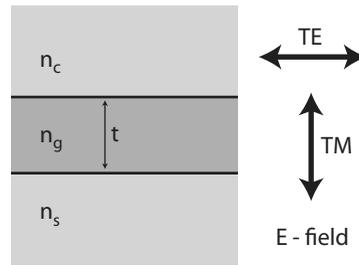


Figure 9.6 Cross-section through a generic dielectric waveguide. The core (or guiding) region with thickness t has refractive index n_g , the upper cladding (or cap) region has n_c and the lower cladding (or substrate) region has n_s . The direction of the electric field for the TE and TM polarizations is also shown.

Total internal reflection

We recall from Section 4.4 that TIR takes place for angles of incidence onto the interface larger than the critical angle,

$$\theta_c = \arcsin \frac{n_{cladding}}{n_{core}} \quad (9.8)$$

where we stipulate that $n_{core} > n_{cladding}$. For angles of incidence $\theta < \theta_c$, TIR is not satisfied and a portion of the field is transmitted across the boundary; as a result, the optical field slowly leaks from the core and guiding does not take place.

In general, the upper and lower cladding regions may be made of different materials, with differing refractive indices, as we see in Figure 9.6. Since the reflection angles at the top and bottom interfaces are identical, and the critical angle becomes larger as the index difference across the interface becomes smaller, the limiting interface in a general waveguide structure is that with the *smaller* index difference. Thus, if $n_s > n_c$ in Figure 9.6, then θ_c at the lower boundary is the critical angle which limits the propagation in the waveguide.

Characteristic equation

The phase matching condition which we derived above for explaining propagation in a mirror waveguide also applies to the dielectric waveguide. We can consider the general structure shown in Figure 9.6. We define our notation so that the core (or guiding) region with thickness t has refractive index n_g ; the upper cladding (or cap) region has n_c ; and the lower cladding (or substrate) region has n_s .

Phase matching led to the characteristic equation for the mirror waveguide, Equation 9.4, and since the geometry is the same for the dielectric waveguide, we can write

$$\frac{4\pi t}{\lambda} \cos \theta - \Delta\phi_c - \Delta\phi_s = 2m\pi \quad (9.9)$$

where we recall that λ is the wavelength in the material, $\lambda = \lambda_0/n_g$. In Equation 9.9, the phase shifts upon reflection at the interfaces are given by $\Delta\phi_c$ for the upper (core/cap) interface and $\Delta\phi_s$ for the lower (core/substrate) interface; that these values are not

a constant π is another of the essential differences between the mirror and dielectric waveguides.

We saw in Section 4.4.3 that the phase shift upon total internal reflection is a function of the refractive indices on either side of the interface, the angle of incidence, as well as the polarization. Recalling Equations 4.45 and 4.46, we have,

$$\Delta\phi_{TE} = 2 \arctan \left[\frac{\sqrt{\sin^2 \theta - \left(\frac{n_c}{n_g}\right)^2}}{\cos \theta} \right] \quad (9.10)$$

$$\Delta\phi_{TM} = 2 \arctan \left[\frac{\sqrt{\sin^2 \theta - \left(\frac{n_c}{n_g}\right)^2}}{\left(\frac{n_c}{n_g}\right)^2 \cos \theta} \right]. \quad (9.11)$$

for TE and TM, respectively; recall the definitions in Figure 4.2.1. In the previous expressions, we have considered the core/cap interface (with n_g and n_c), but the expressions equally apply for the core/substrate interface by replacing n_c with n_s . The values we enter for the phase shifts, $\Delta\phi_c$ and $\Delta\phi_s$, are thus a function of the materials on the outside of the waveguide and the polarization of the light.

Symmetric waveguide

Let us consider a special case: a symmetric waveguide structure, for which cap and substrate are of the same material such that $n_c = n_s$; let us also consider a TE-polarized optical field, such that the arrangement is as shown in the cross-section of Figure 9.7. Since both interfaces have the same index step, the total phase shift due to reflection required to attain phase matching is two times the value of Equation 9.10, so that the characteristic equation becomes

$$\frac{4\pi t}{\lambda} \cos \theta - 4 \arctan \left[\frac{\sqrt{\sin^2 \theta - \left(\frac{n_c}{n_g}\right)^2}}{\cos \theta} \right] = 2m\pi \quad (9.12)$$

which can alternatively be re-written in the form

$$\tan \left(\frac{\pi t}{\lambda} \cos \theta - \frac{m\pi}{2} \right) = \frac{\sqrt{\sin^2 \theta - \left(\frac{n_c}{n_g}\right)^2}}{\cos \theta}. \quad (9.13)$$

Equation 9.12 is the characteristic equation for the present case: TE polarization and a symmetric waveguide. In contrast to the characteristic equation which resulted from our simple mirror model, Equation 9.12 cannot be solved analytically for θ , such that either graphical or numerical approaches must be employed. We will undertake these solutions and discuss their implications in Section 9.1.4 below.

If we now consider the case for a TM-polarized field, the phase relationship is similar, save that we need to consider the phase shift upon TIR given for that orientation of the

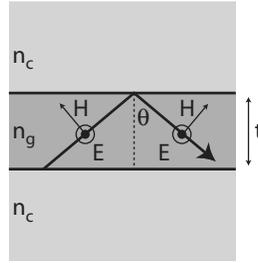


Figure 9.7 A symmetric waveguide with cap and substrate refractive index n_c and a core region of thickness t with refractive index n_g . The orientation of the electric and magnetic fields for TE polarization are also shown; E is always transverse to the interface and the orientation of H changes upon reflection.

field, given by Equation 9.11, which then yields the characteristic equation

$$\frac{4\pi t}{\lambda} \cos \theta - 4 \arctan \left[\frac{\sqrt{\sin^2 \theta - \left(\frac{n_c}{n_g}\right)^2}}{\left(\frac{n_c}{n_g}\right)^2 \cos \theta} \right] = 2m\pi. \quad (9.14)$$

A general optical field may always be decomposed into a sum of TE and TM fields, such that knowing the behavior of these two polarizations allows the determination of the behavior of an arbitrarily polarized optical field.

Asymmetric waveguides

The general configuration for a waveguide employs three different materials for the substrate, core and cap regions; a classic example, which we will consider in the examples, is a silicon nitride waveguide fabricated on a silicon dioxide substrate and covered by air. If $n_s \neq n_c$, an asymmetric waveguide results, where typically $n_s > n_c$. The top and bottom interfaces are thus different, and we see from Equations 9.10 and 9.11 that the resultant phase shift differs if the reflection takes place at the bottom or top interface.

As a result, the characteristic equation for an asymmetric waveguide must take these two phase shifts into account, and accordingly becomes, for TE polarization,

$$\frac{4\pi t}{\lambda} \cos \theta - 2 \arctan \left[\frac{\sqrt{\sin^2 \theta - \left(\frac{n_s}{n_g}\right)^2}}{\cos \theta} \right] - 2 \arctan \left[\frac{\sqrt{\sin^2 \theta - \left(\frac{n_c}{n_g}\right)^2}}{\cos \theta} \right] = 2m\pi \quad (9.15)$$

and for TM polarization,

$$\frac{4\pi t}{\lambda} \cos \theta - 2 \arctan \left[\frac{\sqrt{\sin^2 \theta - \left(\frac{n_s}{n_g}\right)^2}}{\left(\frac{n_s}{n_g}\right)^2 \cos \theta} \right] - 2 \arctan \left[\frac{\sqrt{\sin^2 \theta - \left(\frac{n_c}{n_g}\right)^2}}{\left(\frac{n_c}{n_g}\right)^2 \cos \theta} \right] = 2m\pi. \quad (9.16)$$

Equations 9.15 and 9.16 represent the most general case for the waveguide characteristic equations, and reduce to Equations 9.12 and 9.14, respectively, when the cap and substrate materials have the same refractive index. As for the symmetric case, of course, these equations must be solved numerically, which we will do presently.

9.1.4 Waveguide analysis

Solving the characteristic equation for a particular waveguide structure and polarization leads to a value for reflection angle θ , for a given structure, wavelength and mode number. What useful waveguide propagation characteristics can we derive from knowledge of the allowed values for θ ? Most interesting in many cases are the number of allowed modes, the propagation constants and the effective index.

Supported modes

As for the simple mirror waveguide model, the number of supported modes (those not cutoff) is a function of the waveguide structure. For the dielectric waveguide, the limiting factor is the critical angle: for angles of incidence $\theta < \theta_c$, a portion of the field is transmitted across the interface, and the optical intensity leaks out of the waveguide after a few reflections. As a result, the maximum mode number m_{max} is found from

$$m_{max} = \text{int} \left(\frac{2t \cos \theta_c}{\lambda} \right) = \text{int} \left(\frac{2t}{\lambda} \sqrt{1 - \frac{n_c^2}{n_g^2}} \right) \quad (9.17)$$

where the “int” operator takes the integer value of its argument. The total number of modes is then $m_{max} + 1$, including the $m = 0$ mode.

We see from Equation 9.17 that, as before, the number of modes increases as waveguide thickness increases or wavelength decreases. In addition, a larger refractive index step between core and cladding also leads to a larger number of supported modes, for a given t and λ . In addition, the $m = 0$ mode is always allowed, regardless of the ratio $2t/\lambda$, implying that, for a symmetric dielectric waveguide, the lowest order mode does not experience cutoff.

Propagation constants

We recall that, in free space or in a homogenous material, the propagation constant is a function of the wavelength and the refractive index, meaning that in vacuum we have

$$k_0 = \frac{2\pi}{\lambda_0} \quad (9.18)$$

and in a material with refractive index n ,

$$k = \frac{2\pi}{\lambda} = \frac{2\pi n}{\lambda_0}. \quad (9.19)$$

In a waveguide, since propagation is no longer continuous but in discrete modes, k is also no longer continuous but takes on discrete values since the allowed wavelengths take on discrete values.

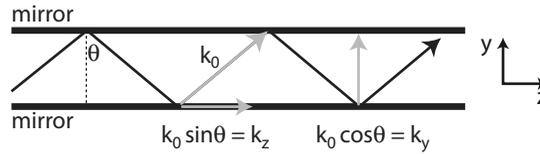


Figure 9.8 Definition of the z and y components of k , k_z and k_y , as related to the free-space value k_0 .

Due to the nature of waveguide propagation, which in the ray-optic model we are employing here means that a ray bounces back and forth between upper and lower interfaces, the general direction of movement is still in the z direction. We can usefully define components of the propagation constants for the components of the movement in the guide; using the convention shown in Figure 9.8, we see that

$$k_z^2 + k_y^2 = k_0^2 \quad (9.20)$$

for the z and y components of k , k_z and k_y , respectively. From rudimentary geometry we find that the vertical (y -directed) motion of the wave is given by

$$k_y = k_0 \cos \theta \quad (9.21)$$

and the longitudinal (z -directed) as

$$k_z = k_0 \sin \theta. \quad (9.22)$$

The z -directed propagation constant is often written as

$$k_z = \beta = k_0 \sin \theta \quad (9.23)$$

using the symbol β [m^{-1}] for simplicity. β may be considered to be an effective propagation constant in the waveguide direction; the ray zigs and zags back and forth, but ultimately moves in the z direction with propagation constant β .

Effective index

If the waveguide core is not vacuum but instead a material with $n > 1$, the effective propagation constant is simply modified by the refractive index, namely

$$\beta = nk_0 \sin \theta = Nk_0 \quad (9.24)$$

where we have defined the effective index of the waveguide N [] as

$$N = n \sin \theta. \quad (9.25)$$

The effective index is a very useful parameter. Related to the physical refractive index of the waveguide material as well as the mode number m , N takes on discrete values and may be interpreted as the refractive index of a material which would lead to the z -directed propagation in the absence of reflection at the boundaries. Since the detailed movement of the rays is unknown to an observer outside the waveguide, propagation takes place as if the waveguide were a bulk slab of material with physical refractive index equal to N . Each mode has a different value for N .

The propagation constant β and the effective index N , closely and simply related parameters, are two of the most important characteristics which result from a waveguide analysis.

Solving a waveguide

How do we proceed in a numerical waveguide analysis? The procedure is straightforward, as we will outline here.

Determine the parameters We first assemble the necessary information about the waveguide structure and optical field: core thickness t , refractive indices (n_s , n_g and n_c), wavelength in free space λ_0 and polarization.

Solve the characteristic equation Based on the structure (symmetric or asymmetric) and polarization (TE or TM), the appropriate characteristic equation (one of Equations 9.12, 9.14, 9.15 or 9.16) is solved for $m = 0$. The solution must be done numerically, using your favorite mathematical program. The result is the allowed value for θ for $m = 0$.

Repeat for higher order modes Set $m = 1$ and repeat the previous step. If a solution is possible, the result is the allowed θ for $m = 1$. Repeat again until no solution is possible. The highest of value of m for which a solution is possible represents the highest mode; all higher-order modes are cutoff. The total number of allowed modes is $m + 1$.

Determine parameters From the expressions above (Equations 9.24 and 9.25), calculate the values for β and/or N for each θ for all possible m . The propagation conditions for all supported modes are then determined.

Further calculations As we will see below, a number of further parameters may be determined from β , including the form of the electric field and evanescent field penetration; we will revisit these in Section 9.2.

Example 9.2: Let us solve a symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4$ waveguide which we illuminate with a red HeNe laser at $\lambda = 633$ nm in the TM polarization. For a 400 nm thick core region, how many modes are allowed, and what are the reflection angles, effective indices and \mathbf{z} -directed propagation constants for the $m = 0$ and highest-order modes? How do these change if we increase the waveguide core thickness to 1.2 μm and calculate the $m = 0$ and highest-order modes again?

We know from the material parameters that SiO_2 has a refractive index of 1.46 and Si_3N_4 about 2.05, at visible wavelengths; thus we use Si_3N_4 as the waveguide core with SiO_2 as the top and bottom claddings. The index of the core region is higher than the cap and substrate, as required for guiding; we can then set $n_s = n_c = 1.46$ and $n_g = 2.05$.

Using Equation 9.17, we start with $t = 400$ nm, and evaluate

$$m_{max} = \text{int} \left[\frac{2t}{\lambda} \sqrt{1 - \frac{n_c^2}{n_g^2}} \right]$$

to find that $m_{max} = 1$, so that two modes (including $m = 0$) are expected in this structure. Since we have a symmetric waveguide and TM illumination, we then employ the symmetric TM characteristic equation, Equation 9.14, namely

$$\frac{4\pi t}{\lambda} \cos \theta - 4 \arctan \left[\frac{\sqrt{\sin^2 \theta - \left(\frac{n_c}{n_g}\right)^2}}{\left(\frac{n_c}{n_g}\right)^2 \cos \theta} \right] = 2m\pi$$

which we solve numerically, again for $t = 400$ nm, first for $m = 0$ and then $m = 1$. Recall that $\lambda = \lambda_0/n_g$ in this expression.

Based on these calculations, we find that, for $t = 400$ nm, the two allowed modes are: for $m = 0$, $\theta = 71.2^\circ$, $N = 1.94$ and $\beta = 19.3 \mu\text{m}^{-1}$; and for $m = 1$, we have $\theta = 53.1^\circ$, $N = 1.64$ and $\beta = 16.3 \mu\text{m}^{-1}$.

Repeating this procedure for a thicker core, $t = 1200$ nm, we find that $m_{max} = 5$, so that 6 modes are supported. For $m = 0$, we determine $\theta = 83.0^\circ$, $N = 2.03$ and $\beta = 20.2 \mu\text{m}^{-1}$. For $m = 5$, on the other hand, we have $\theta = 47.1^\circ$, $N = 1.50$ and $\beta = 14.9 \mu\text{m}^{-1}$.

We see from these calculations how the important parameters vary. For a fixed wavelength, a thicker core allows more modes, as we saw repeatedly above. Higher order modes are closer to cutoff than lower order modes, which we can see from the values for reflection angle: this parameter decreases in magnitude as m increases. We also see that the effective index and propagation constant decrease as mode number increases.

Recalling that wave velocity is related to frequency as $v = \omega/k$ (Equation 2.46), we see that higher order modes have a higher wave velocity than lower order modes. This assertion seems counterintuitive when employing the ray-optic picture of the waveguide (higher order modes zig and zag more), but will become reasonable when we consider the fields in a waveguide, since higher order modes have a greater portion of the evanescent field in the low-index cladding; see Section 9.2.6.

Parameter variations

The propagation characteristics, as we have just seen, vary with the waveguide parameters. It is instructive to see how these adapt as, for example, the structure, wavelength, and mode number change. Let us again consider a symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide, using $\lambda = 633$ nm. As in the example just considered, we have SiO_2 with a refractive index of 1.46 and Si_3N_4 with a refractive index of 2.05, giving $n_s = n_c = 1.46$ and $n_g = 2.05$.

Using the step-by-step procedure we just described, we can calculate how reflection angle θ , effective index N and z-directed propagation constant β vary with waveguide core thickness t ; the results are summarized in Figure 9.9. All of the calculations were done for $m = 0, 1, 2$ for both TE and TM polarizations. If we consider the reflection angle in Figure 9.9(a), we see a number of effects. As t increases, so does θ for a given mode, meaning the angle of the ray becomes increasingly oblique as the core becomes thicker. We also see that the number of modes increases: the $m = 1$ mode is supported

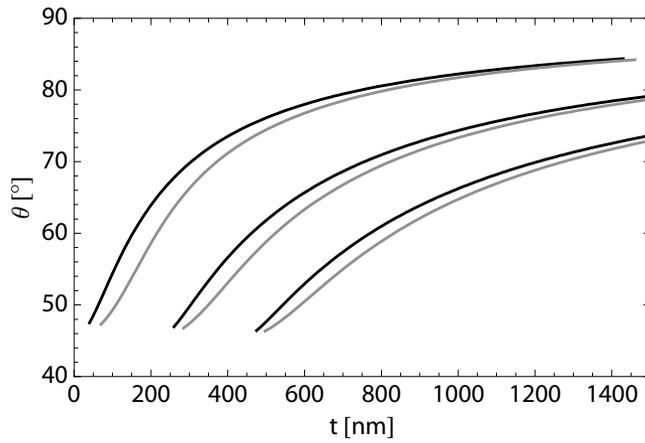
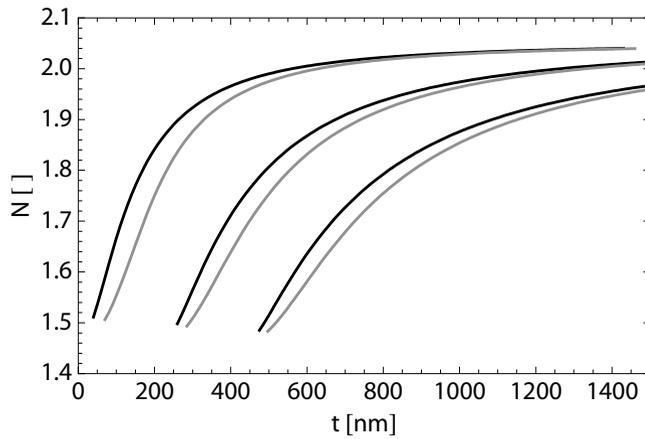
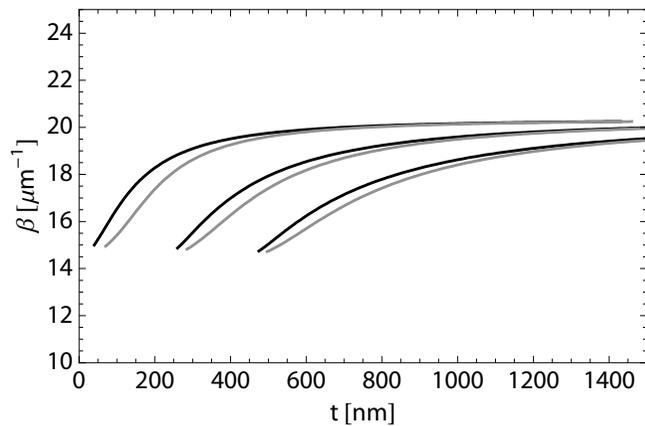
(a) Reflection angle θ (b) Effective index N (c) z-directed propagation constant β

Figure 9.9 Plots of several waveguide characteristics as a function of core thickness t for a symmetric $\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide illuminated at $\lambda = 633$ nm. The calculations are done for $m = 0$ top pair, $m = 1$ middle pair and $m = 2$ bottom pair, in each case for TE black and TM gray polarizations.

for $t > 250$ nm, the $m = 2$ mode for $t > 500$ nm. Higher order modes have a smaller reflection angle (steeper incidence) than lower order modes, and the two polarizations (black and gray for TE and TM, respectively) have slightly different values for θ as well.

These variations are reflected in the values for effective index, seen in Figure 9.9(b), and the propagation constant of Figure 9.9(c). The qualitative behavior of the curves is roughly the same as those for θ . The effective index, N , varies between the refractive index of the cladding (SiO_2 , with $n_s = n_c = 1.46$) for modes close to cutoff and approaches the refractive index of the core (Si_3N_4 , with $n_g = 2.05$) for modes far from cutoff. Indeed, effective index is always found in the range

$$n_g \geq N \geq n_s \quad (9.26)$$

where, for the asymmetric waveguide, the right hand side may be either n_s or n_c , whichever is *larger*.

In addition, typical magnitudes for β are in the tens of inverse micrometers, seen in Figure 9.9(c). Similarly to the effective index, propagation constant varies in the range

$$k_0 n_g \geq \beta \geq k_0 n_s \quad (9.27)$$

with the same caveat as for the asymmetric waveguide. The physical significance of these value ranges will become clearer when we consider the form of the optical field in Section 9.2.

Example 9.3: Given the symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide with 400 nm core thickness we have considered a few times above, how does N vary with wavelength across the visible wavelength range?

We take the same approach as in the examples above: find the correct characteristic equation (symmetric TE, Equation 9.12 in this case), and solve it repeatedly for all possible modes, determine θ and calculate N . Admittedly this is only practical using mathematical software; the results of such a calculation are shown in Figure 9.10.

Three modes are supported through most of the visible wavelength range; we see that the cutoff wavelength for $m = 2$ is about 570 nm. In this range, N decreases monotonically with increasing wavelength, an effect which we saw in the examples above. As the modes approach cutoff (which they do as λ increases), the effective index decreases from its maximum (n_g) to its minimum (n_c or n_s , whichever is larger) value.

Example 9.4: What is the effect of the cladding refractive index value for an asymmetric waveguide? We consider an 800 nm thick Si_3N_4 waveguide deposited on BK7 glass for operation at 543 nm. We thus know that $n_g = 2.05$ and $n_s = 1.512$, and we can let n_c vary from 1 (air) to just below n_g ; any larger and, of course, guiding no longer takes place. Let us use the effective index as an indicative parameter.

The characteristics of Figure 9.11 result from the repeated solution of the asymmetric TE characteristic equation (Equation 9.15). We see that the different modes have

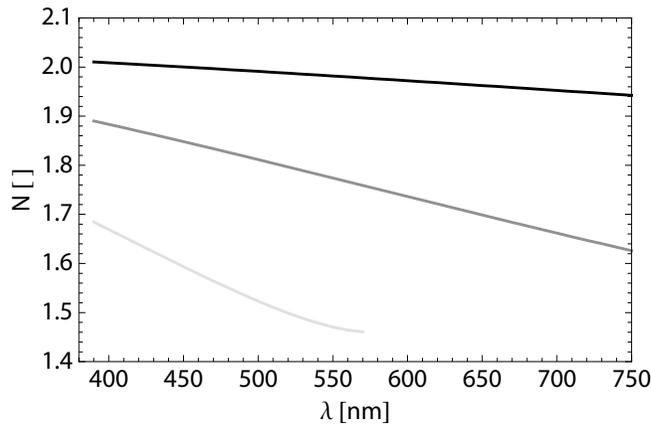


Figure 9.10 Calculated effective index N as a function of wavelength for the visible range for a $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ symmetric waveguide with $t = 400$ nm and TE polarization. The plots are for $m = 0$ top, black, $m = 1$ middle, gray, and $m = 0$ bottom, light gray.

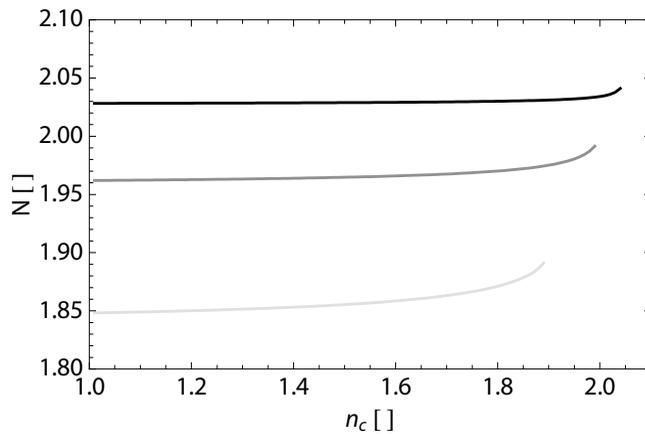


Figure 9.11 Effective index N calculated as a function of cap refractive index n_c for a $\text{Si}_3\text{N}_4/\text{BK7}$ asymmetric waveguide. Core thickness is $t = 800$ nm, wavelength $\lambda = 543$ nm and polarization is TE. The plots are for $m = 0$ top, black, $m = 1$ middle, gray, and $m = 0$ bottom, light gray.

decreasing N with increasing m , as above, but also that N does not vary strongly with varying cap index until the mode comes close to cutoff, for n_c approaching 2.0. In this regime, small variations in cap index can lead to cutoff for higher order modes and a variation in N which is larger for those modes.

In general, waveguides are designed such that the external environment (the cap medium could be air, or water ($n = 1.33$) or another medium) does not affect the propagation conditions; for the $m = 0$ mode in this example, which is far from cutoff, that is more or less the case. Some applications, such as sensors (which we consider in Section

9.7.2), on the other hand, explicitly rely on this variation; in that case, modes close to cutoff are preferable.

9.2 Waveguides: electromagnetic model

The ray optic model for a waveguide has the advantage that it provides a useful heuristic picture for propagation of discrete modes in a guided-wave structure. The results we have obtained from it, particularly mode number, cutoff wavelength or thickness, and effective index, are all dependable values, and allow a reliable waveguide design.

An optical field propagating in a waveguide is still an electromagnetic disturbance, however, so that we can also perform an electrodynamic analysis of the structure. The advantage of an electromagnetic solution is that the distribution of the optical field may be determined, and the analysis may be done rigorously. A field solution allows us to determine the evanescent field, which extends outside the waveguide core, the intensity distribution and the power transmission. For this reason, we repeat our analysis of a slab waveguide using Maxwell's equations as the most important tool.

9.2.1 Wave equation in a waveguide

We recall from Chapter 2 that the propagation of an electromagnetic field is described by the wave equation (Equation 2.27), given in three dimensions as

$$\nabla^2 \mathbf{E} = \frac{1}{v^2} \frac{d^2 \mathbf{E}}{dt^2}. \quad (9.28)$$

for electric field vector \mathbf{E} and velocity v . In the discussion of the Gaussian beam in Section 2.6, we saw that entering field solutions of the form

$$E(r, t) = E(r) e^{j\omega t} \quad (9.29)$$

into the wave equation gave rise to the Helmholtz equation, namely

$$\nabla^2 E(r) + k^2 E(r) = 0 \quad (9.30)$$

which is a three-dimensional eigenvalue equation for electric field E and includes the propagation constant k .

For the special case of a slab waveguide, as shown again in Figure 9.12, propagation is in the z direction and the slab extends infinitely in the x and z directions; the y direction is limited to a thickness $t = 2h$, in the range $-h \leq y \leq +h$. The field may thus be written as

$$E(r, t) = E_0(x, y) e^{-j\beta z} e^{j\omega t}, \quad (9.31)$$

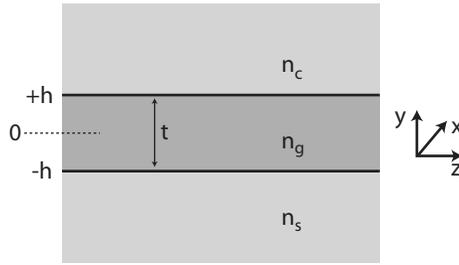


Figure 9.12 Cross sectional view of a dielectric slab waveguide for an electromagnetic analysis.

for z -directed propagation constant β . Since the field has a variation only in the y direction, the Laplacian of the Helmholtz equation becomes

$$\nabla^2 \mathbf{E} \rightarrow \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} = \frac{\partial^2 E}{\partial y^2} - \beta^2 E \quad (9.32)$$

such that the eigenvalue equation takes on the simpler form

$$\frac{\partial^2 E}{\partial y^2} + (n_i^2 k_0^2 - \beta^2) E = 0; \quad (9.33)$$

n_i represents the refractive index in layer i , where $i = s, g, c$ depending on whether we perform the solution in the substrate, core or cap layer, respectively. If we further define a y -directed propagation constant k_{yi} for each region i as

$$k_{yi} = n_i^2 k_0^2 - \beta^2, \quad (9.34)$$

the eigenvalue equation in the waveguide then simplifies to

$$\frac{\partial^2 E}{\partial y^2} + k_{yi}^2 E = 0. \quad (9.35)$$

This simple eigenvalue equation has two types of solutions, depending on the sign of the k_{yi}^2 term. For the case where $k_{yi}^2 > 0$, we obtain harmonic electric field solutions, of the form

$$E_0 e^{\pm j k_{yi} y} \quad (9.36)$$

and if alternatively $k_{yi}^2 < 0$, the electric field solutions are exponential, namely

$$E_0 e^{\pm k_{yi} y} \quad (9.37)$$

where only decaying solutions are physically realistic for the configuration we consider here.

9.2.2 Propagating and decaying solutions

Whether oscillating (Equation 9.36) or exponentially decaying (Equation 9.37) solutions to Equation 9.35 are obtained thus depends on the value of k_{yi}^2 which in turn depends on the relationship between the refractive index in the region in which Equation 9.35 is solved (n_i) and that in the core, through the propagation constant β , given

by Equation 9.24. Oscillating solutions correspond to propagating modes whereas exponentially decreasing solutions are those which decay and disappear.

Regional solutions

We can now solve Equation 9.35 in the substrate, core and cap ($i = s, g, c$) regions and see what types of solutions are obtained. Assuming that $n_g > n_s > n_c$, we find four possibilities:

$\beta > k_0 n_g$: Solutions in all three regions are of the form $E_0 e^{\pm k_y i y}$, where we take the plus or minus sign depending on the polarity of y . We have exponentially decaying solutions in all three regions and thus no guiding or propagation.

$k_0 n_g > \beta > k_0 n_s$: In the core region with n_g , we obtain a propagating solution, of the form $E_0 e^{\pm j k_y i y}$, and the cap and substrate regions still have decaying solutions. This situation corresponds to the ideal case for a waveguide: guiding only in the core and decaying solutions elsewhere.

$k_0 n_s > \beta > k_0 n_c$: Here harmonic solutions result in the core but also in the substrate; the cap has a decaying solution. Thus propagation takes place in both the core and the substrate, such that optical energy is lost into the substrate, giving rise to a substrate radiation mode.

$k_0 n_c > \beta$: For this final case, all three regions have harmonic, propagating solutions. Thus the optical field propagates everywhere and we again have no guiding.

Solving the wave equation in a three-layer waveguide structure thus leads to different types of solutions, depending on the values of the n_s, n_g, n_c and β ; not all of these correspond to an ideal guided mode. A properly operating waveguide is obtained only for the second case above, for $k_0 n_g > \beta > k_0 n_s$, with a propagating solution only in the core, verifying that which we saw for the ray-optic model in Equation 9.27.

9.2.3 Waveguide electric field

By solving the wave equation, we can derive the form of the electric field in the waveguide core and cladding. A well-functioning waveguide has propagating solutions in the core and decaying solutions in the cladding, such that the fields in the three regions may be given by

$$E_x(y)|_{cap} = E_{c0} e^{-k_y c (y-h)} \quad \text{for } y \geq +h \quad (9.38)$$

$$E_x(y)|_{guide} = E_{g0} e^{j k_y g y} \quad \text{for } -h \leq y \leq +h \quad (9.39)$$

$$E_x(y)|_{substrate} = E_{s0} e^{+k_y s (y+h)} \quad \text{for } y \leq -h \quad (9.40)$$

using the notation of Figure 9.12. The harmonic solution, $E_x(y)|_{guide}$, may, by Euler's relation, be written as a sum of sine and cosine functions; without loss of generality, we may express this in the useful form

$$E_x(y)|_{guide} = E_{g0} \cos(k_y g y + \phi) \quad (9.41)$$

where the phase factor ϕ allows the generation of both an even and odd (or mixed) functional form. As we will see in the following example, the phase factor is in general determined by application of the boundary conditions between core and cladding.

An important consideration which results from this analysis is that the field solutions are not zero outside the waveguide core, as we implicitly assumed with our ray-optic model based on total internal reflection, but decays to zero exponentially. Thus a portion of the optical field extends outside the waveguide even for a well-guided mode; we will analyze this quantitatively in the subsequent section.

Propagation constants

We defined the y-directed propagation constant for a waveguide, k_{yi} , in Equation 9.34 above. Based on the results of the previous section, we can write values for this constant in the three regions explicitly. The interesting case is the one for $k_0 n_g > \beta > k_0 n_s$, namely propagation only in the core region; for this situation, we have

$$k_{yc} = \sqrt{\beta^2 - k_0^2 n_c^2} = k_0 \sqrt{N^2 - n_c^2} \quad (9.42)$$

$$k_{yg} = \sqrt{k_0^2 n_g^2 - \beta^2} = k_0 \sqrt{n_g^2 - N^2} \quad (9.43)$$

$$k_{ys} = \sqrt{\beta^2 - k_0^2 n_s^2} = k_0 \sqrt{N^2 - n_s^2} \quad (9.44)$$

for the cap, core and substrate regions, respectively. In the previous three expressions, the square roots all evaluate to real values so that the arguments of the exponentials in Equations 9.38 and 9.40 above are all real.

Example 9.5: What is the typical form of the electric field in a waveguide? Let us take a 800 nm thick symmetric SiO₂/Si₃N₄/SiO₂ waveguide and illuminate it with a 633 nm HeNe laser. We know from our numerous previous analyses that $n_s = n_c = 1.46$ and $n_g = 2.05$, using the structure and notation of Figure 9.12. Just for fun, let's plot the TE mode with $m = 2$.

The waveguide is solved using the symmetric TE characteristic equation, Equation 9.12, from which we find that the reflection angle for $m = 2$ is 61.0°. From this value, we can directly determine the z-directed propagation constant from Equation 9.24 to be $\beta = 17.8 \mu\text{m}^{-1}$ and then, from Equations 9.42 – 9.44, calculate the propagation constants in the cap, core and substrate: $k_{yc} = k_{ys} = 10.3 \mu\text{m}^{-1}$ and $k_{yg} = 9.9 \mu\text{m}^{-1}$.

The one remaining unknown is the phase factor ϕ in Equation 9.41. For even modes ($m = 0, 2, \dots$), we expect an even harmonic function (the cosine) centered at $y = 0$; in contrast, for odd modes ($m = 1, 3, \dots$), we expect an odd function (the sine), again centered at $y = 0$. Whereas in many cases (the one we are presently analyzing included) we can then simply take $\phi = 0$ or π for $m = 0, 2, \dots$ and $\phi = \pm\pi/2$ for $m = 1, 3, \dots$, a general analysis requires that we use the condition that the derivative of the electric field is continuous across the core/cladding boundary to obtain a relationship which allows us to determine ϕ . Such a procedure is absolutely necessary for an asymmetric waveguide, since the phase factor in that case is not a simple multiple of $\pi/2$.

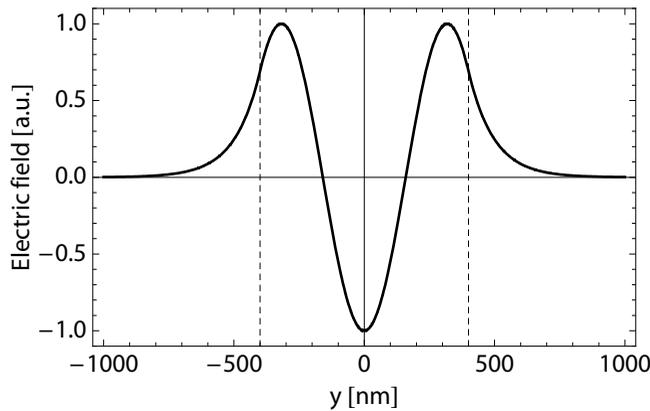


Figure 9.13 The electric field for the $m = 2$ mode in a 800 nm thick $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ illuminated at $\lambda = 633$ nm with TE polarization. The waveguide core is between the vertical dotted lines.

Therefore, taking

$$\frac{d(E_x(y)|_{\text{guide}})}{dy} = \frac{d(E_x(y)|_{\text{cap}})}{dy}$$

at $y = h$, and solving the resultant expression for ϕ , we determine $\phi = -\pi$, no great surprise.

We now have all the information we need. Plugging these numerical values into the field equations for the three regions, Equations 9.38 – 9.40, we obtain the characteristic of Figure 9.13. We see three peaks (recalling that intensity is proportional to E^2) and the expected symmetry about $y = 0$. The oscillating solution in the core becomes exponentially decaying outside the waveguide core (shown by the vertical dotted lines), giving the evanescent field.

It is interesting to compare the characteristics of different modes by examining their intensities, as we have done for the same waveguide structure for $m = 0, 1, 2$ in Figure 9.14. The number of intensity peaks is given by $m + 1$ and we see that the extent of the evanescent field, which we will discuss presently, increases as m increases.

Evanescent field

From Equations 9.38 and 9.40 and the examples we have just calculated above, it is clear that the electric (optical) field decays outside the waveguide core region, with a decay constant given by k_{yc} or k_{ys} . The field outside the core is termed the evanescent field, to emphasize that no energy is transmitted outside the waveguide: energy *exists* outside the core, but there is no propagation in the y direction.

The rate of decay of the evanescent field is given by the magnitude of k_{yc} or k_{ys} . If we consider the field in the cap layer, the propagation constant leads to an exponential

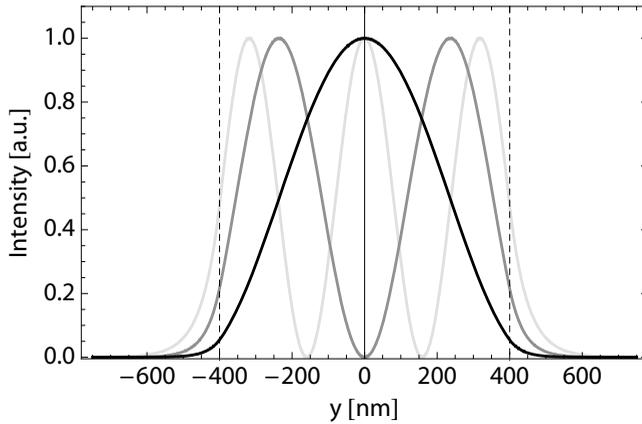


Figure 9.14 Field intensity in a 800 nm thick $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ illuminated at $\lambda = 633$ nm, for $m = 0$ black, $m = 1$ gray and $m = 2$ light gray, all for TE polarization. The waveguide core is between the vertical dotted lines.

function of the form

$$E_x(y)|_{cap} = E_{c0} \exp \left[-k_0 \sqrt{N^2 - n_c^2} \cdot (y - h) \right] \quad (9.45)$$

which, recalling the definition of effective index, may be written as

$$E_x(y)|_{cap} = E_{c0} \exp \left[-k_0 \sqrt{n_g^2 \sin^2 \theta - n_c^2} \cdot (y - h) \right]. \quad (9.46)$$

A smaller value for k_y implies a slower exponential decay and thus a greater extent of the evanescent field into the cap region. From the previous expression, we see that smaller values of reflection angle θ lead to smaller values for k_y and, since modes closer to cutoff have smaller incidence angles than more well-guided modes, higher-order modes generally have larger evanescent fields than lower-order modes. We see this clearly in Figure 9.14, where the field intensity outside the core is distinctly larger for larger values of m .

Considering the waveguide variations we saw in Figure 9.9, it is instructive to see how the magnitude of the evanescent field varies for these different cases. Using the relationship of Equation 9.46, we can define a penetration depth (in units of meters) as that distance from the interface for which the field has dropped to $1/e$ of its interface value; penetration depth is thus given by $1/k_{yc}$ and is a useful parameter for characterizing the evanescent field.

Recall that in Figure 9.9 we analyzed a symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide using $\lambda = 633$ nm, such that $n_s = n_c = 1.46$ and $n_g = 2.05$. Calculating the penetration depth for the various modes and polarizations we considered in our previous calculation leads to the characteristics of Figure 9.15. We see from the calculations that higher order modes have larger penetration depths, that TM generally has a larger penetration depth than TE and that penetration depth increases rapidly as we approach cutoff.

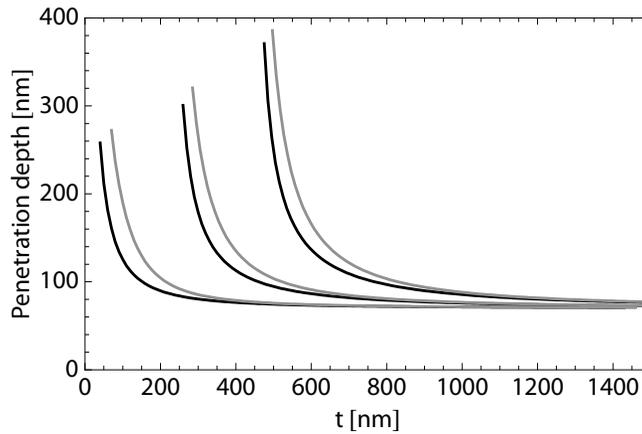


Figure 9.15 Penetration ($1/e$) depth for a $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ symmetric waveguide illuminated at $\lambda = 633$ nm. The plots are for $m = 0$ left pair, $m = 1$ middle pair and $m = 2$ right pair, in each case for TE black and TM gray polarizations. The maximum values are a function of how close to cutoff we calculate: at cutoff, the penetration depth is infinite.

Typical magnitudes of penetration depth for well-confined modes, far from cutoff, are less than 100 nm, and we see that this value increases rapidly as cutoff is approached.

Example 9.6: Let us examine the form of the propagating and evanescent fields for an asymmetric waveguide. Employing a 800 nm thick PVC waveguide core ($n_g = 1.540$) deposited on a BK7 substrate ($n_s = 1.517$) and water as the cap material ($n_c = 1.3318$), we can solve this waveguide for $\lambda = 633$ nm and TE polarization using the mathematical tools we have developed.

Solving Equation 9.15 for $m = 0$, we determine a reflection angle of $\theta = 81.5^\circ$. Employing Equation 9.24, we find that $\beta = 15.1 \mu\text{m}^{-1}$ and then using Equations 9.42 – 9.44, calculate the propagation constants in the cap, core and substrate: $k_{yc} = 7.3 \mu\text{m}^{-1}$, $k_{yg} = 2.3 \mu\text{m}^{-1}$ and $k_{ys} = 1.3 \mu\text{m}^{-1}$.

We then need to determine the phase factor ϕ for the harmonic field solution in the core, Equation 9.41. As in Example 9.5, we take the continuity of the derivative of the electric field across the core/cap (or core/substrate) interface, and find that the resultant equation is solved (at $y = h$) for $\phi = 21.04^\circ$. All necessary parameters are now known.

The functional form of the fields is then calculated from Equations 9.38 – 9.40 and is seen in Figure 9.16, where the waveguide core region is found between the vertical dotted lines. The peak of the field is distinctly not at the center of the core, but is shifted toward the substrate; such behavior is typical for an asymmetric waveguide.

It is also apparent that the substrate, to the left, has a considerably more slowly decaying evanescent field than the cap, to the right. This behavior is expected, since the refractive index difference between core and substrate (1.540 to 1.517) is considerably smaller than between core and cap (1.540 to 1.3318); had we used air ($n = 1$) as the cap, the evanescent field in the cap would have been even smaller. As a result, the critical angle

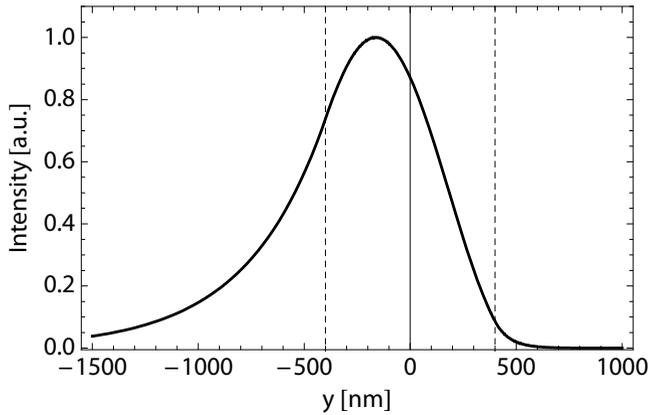


Figure 9.16 The electric field intensity for an asymmetric waveguide consisting of a 800 nm thick PVC core ($n_g = 1.540$) on a BK7 substrate ($n_s = 1.517$) with water as the cap material ($n_c = 1.3318$); illumination wavelength is 633 nm, TE polarization. The core region is between the dotted lines; the substrate is to the left; and the cap on the right. We clearly see the difference in the penetration depths in substrate and cap and that the peak field is shifted toward the substrate from the waveguide center.

for the core/substrate interface is much larger than for the core/cap interface and cutoff will occur when the mode begins to propagate across the lower interface; the penetration depth is thus larger there.

9.2.4 Confinement factor

For many applications of waveguides, it is of interest to have a quantitative measure of how much of the modal energy is transmitted in the core, as opposed to in the evanescent field. Recall we asserted above that no energy is transmitted out of the waveguide by the evanescent field, but that the guided mode has a portion of its z -directed energy transmitted in the cladding, which ultimately has an effect on the propagation conditions.

The confinement factor Γ [] is a measure of the fraction of optical energy found in the core, normalized to the total energy of the field, and is defined as

$$\Gamma = \frac{\int_{-h}^{+h} E_x^2(y) dy}{\int_{-\infty}^{+\infty} E_x^2(y) dy}. \quad (9.47)$$

A perfectly confined mode (with zero evanescent field) would have $\Gamma = 1$ and the smaller the value for Γ , the poorer the confinement. As seen in the plots of Figure 9.17, in which Γ is calculated for a $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide, such that $n_s = n_c = 1.46$ and $n_g = 2.05$, with a wavelength of $\lambda = 633$ nm, confinement factor varies with waveguide geometry, mode number and polarization. Γ increases as the waveguide core

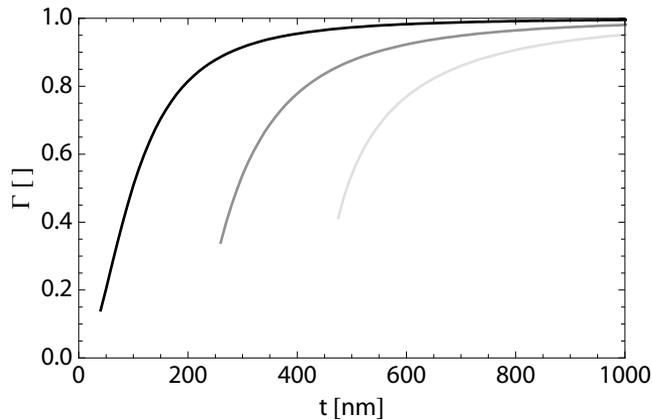


Figure 9.17 Confinement factor in a symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide with $\lambda = 633$ nm, calculated for TE polarization and $m = 0$ black, $m = 1$ gray, and $m = 2$ light gray as a function of waveguide core thickness t .

thickness increases, and, for a given value of t , modes closer to cutoff have poorer confinement than modes of lower order. This effect can be seen, for example, if we consider $t = 300$ nm in the figure: for $m = 0$, $\Gamma = 0.92$, but the $m = 1$ mode, close to cutoff, has $\Gamma = 0.54$. As a mode approaches cutoff, then, its size increases, more and more energy is transmitted in the cladding regions, until, when cutoff is reached, guiding ceases altogether.

The value of confinement factor desired in waveguide structure is a function of its application. Well-guided modes with $\Gamma \approx 1$ typically result in waveguides designed for efficient transmission of an optical field, such as in an integrated optical circuit, or in long-distance fiber optical systems. Good confinement implies that the effects of the external environment which may result in a change in the optical characteristics of the cladding layers have little effect on the optical field in the core.

Some optical sensor applications, on the other hand, rely on the change in propagation conditions which result from variations in the characteristics of the cladding. In Figure 9.16, for example, we considered an asymmetric waveguide for which the cap consisted of water; variations in the composition of this liquid surface layer may result in a change in its refractive index and thus in the propagation conditions of the mode. To optimize the sensitivity of the mode to changes in the cladding, we would prefer Γ to be small, such that the evanescent field extends further into the sensing medium. We note, incidentally, that the structure of Figure 9.16 is not optimal for sensing in the water layer, since the evanescent field extends much further into the substrate than into the cap: a better design would be required here to increase the effect of the cap layer on mode propagation.

9.2.5 Characteristic equation

In Section 9.1.3, we derived the characteristic equation for a dielectric waveguide using arguments based on phase matching after reflection from the boundaries. It is instructive to consider the field distribution in the waveguide, apply the electromagnetic boundary conditions and see what the result predicts about propagation conditions and the existence of modes.

We employ the fields for a guided, propagating mode in the core and cladding given by Equations 9.38 – 9.40, recalling that the propagation constants are defined by Equations 9.42 – 9.44; the structure is still that shown in Figure 9.12. As is well known from basic electromagnetics, and as we discussed in Example 9.5, both the electric field parallel to the interface (E_x in our case) and its derivative must be continuous across an interface. At the core/cap and core/substrate boundaries, continuity of the field implies

$$E_{c0} = E_{g0} \cos(k_{yg}h + \phi) \quad \text{at } y = +h \quad (9.48)$$

$$E_{s0} = E_{g0} \cos(-k_{yg}h + \phi) \quad \text{at } y = -h \quad (9.49)$$

and continuity of the y -derivative, dE_x/dy , yields

$$E_{c0} = E_{g0} \frac{k_{yg}}{k_{yc}} \sin(k_{yg}h + \phi) \quad \text{at } y = +h \quad (9.50)$$

$$E_{s0} = E_{g0} \frac{k_{yg}}{k_{ys}} \sin(k_{yg}h - \phi) \quad \text{at } y = -h. \quad (9.51)$$

The previous four equations allow us to derive eigenvalue equations for the four unknowns. Setting Equation 9.48 equal to Equation 9.50, and then Equation 9.49 equal to Equation 9.51, we obtain the two relationships

$$\tan(k_{yg}h + \phi) = \frac{k_{yc}}{k_{yg}} \quad (9.52)$$

$$\tan(k_{yg}h - \phi) = \frac{k_{ys}}{k_{yg}}; \quad (9.53)$$

taking the inverse tangent of the previous two expressions and adding them, we are left with

$$2k_{yg}h - \arctan\left(\frac{k_{yc}}{k_{yg}}\right) - \arctan\left(\frac{k_{ys}}{k_{yg}}\right) = m\pi \quad (9.54)$$

for integer m . Equation 9.54 is an eigenvalue equation which looks suspiciously similar to the characteristic equation for TE polarization in an asymmetric waveguide, Equation 9.15. Indeed, if we apply the definitions of the propagation constants k_{yc} , k_{yg} and k_{ys} , using Equations 9.42 – 9.44, we can re-write Equation 9.54 as

$$2hk_0 \sqrt{n_g^2 - N^2} - \arctan \sqrt{\frac{N^2 - n_c^2}{n_g^2 - N^2}} - \arctan \sqrt{\frac{N^2 - n_s^2}{n_g^2 - N^2}} = m\pi. \quad (9.55)$$

It is left to the Problems below for the student to show that Equation 9.55 is identical to Equation 9.15 and thus corresponds to the characteristic equation for an asymmetric waveguide operated in TE polarization. The point of this exercise was to show that

the eigenvalue equation defining the allowed propagating modes in a waveguide may also be derived rigorously from consideration of Maxwell's equations in the slab waveguide structure. By verifying the result we obtained using the ray optic model, specular reflection and the critical angle, we show that the physics of the system is completely self-consistent. Not that we had any doubts.

9.2.6 Other aspects of waveguide fields

The physics of waveguiding is quite rich, and numerous further aspects of propagation in slab waveguides may be considered; see for example (Nishihara et al., 1989), (Hunsperger, 2002) or (Iizuka, 2002b). We consider two further topics briefly here: the Goos-Hänchen shift and mode velocity.

Goos-Hänchen Shift

The total-internal-reflection and evanescent field models we considered in the previous sections appear to be mutually exclusive, or at least contradictory. The first says that, upon TIR, nothing is transmitted across the boundary whereas the latter defines an exponentially decaying field on the far side of the boundary. In reality, both models are correct and consistent: there is no *propagation* across the boundary, so energy is not transmitted across it, but the field has non-zero magnitude on the far side of the interface, so part of the modal energy is transmitted (in the z , not the y , direction) in the cladding layer.

If we look in detail at the reflection at the boundary, we see that the penetration of the field beyond the interface is as shown in Figure 9.18. The reflection at the interface is accompanied by a lateral shift in the z direction, known as the Goos-Hänchen shift² (Goos and Hänchen, 1947; Goos and Lindberg-Hänchen, 1949), S_{GH} [], which, for a reflection angle θ , is given by (Lotsch, 1968)

$$S_{GH} = \frac{2}{k_{yc}} \tan \theta, \quad (9.56)$$

for the case of the core/cap interface. If we recall our definitions of z -directed propagation constant β , as well as k_{yg} and k_{yc} , we can alternatively write the Goos-Hänchen shift as

$$S_{GH} = \frac{2\beta}{k_{yc}k_{yg}} = \frac{2N}{k_0\sqrt{N^2 - n_c^2}\sqrt{n_g^2 - N^2}} \quad (9.57)$$

which is only a function of wavelength (through k_0) and the refractive indices of the waveguide. A similar expression (replacing n_c with n_s) can be written for the core/substrate interface.

² Fritz Goos and Hilda Lindberg-Hänchen worked together at the Physikalisches Staatsinstitut in Hamburg, Germany, in the 1940s. Goos also studied the spectra of electrical arcs, discovering the so-called "pole effect", and Hänchen appears to have been his PhD student.

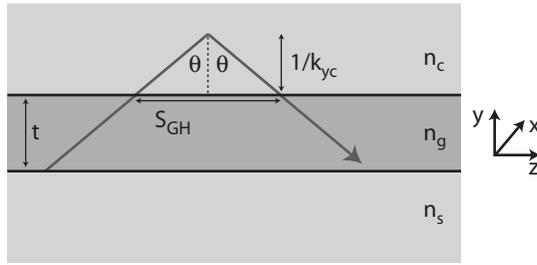


Figure 9.18 Definition of the Goos-Hänchen shift S_{GH} in a waveguide. The extent of the field into the upper cladding region is characterized by the penetration depth, $1/k_{yc}$, and is in reality an exponential decay.

The Goos-Hänchen shift increases in magnitude as mode number increases; thus, higher order modes are subject to a larger lateral shift for each reflection than lower order modes. Since a larger value for S_{GH} implies that the wave effectively spends more of its propagation time in the cladding layer, which has a lower refractive index than the waveguide core, its overall velocity is higher. This picture provides an alternative explanation for why higher order modes have a higher propagation velocity than lower order modes.

Example 9.7: Let's consider the Goos-Hänchen shift of the asymmetric BK7 / PVC / water waveguide we analyzed in Example 9.6. Recall that we had a 800 nm thick PVC waveguide core ($n_g = 1.540$) on a BK7 substrate ($n_s = 1.517$) with water as the cap material ($n_c = 1.3318$), using a TE-polarized HeNe laser ($\lambda = 633$ nm) for illumination.

As previously, we first solve Equation 9.15 for $m = 0$, and determine that the reflection angle has a value of $\theta = 81.5^\circ$. It is then straightforward to calculate that $N = 1.523$, which we can then directly insert into Equation 9.57 from which we find that $S_{GH} = 1.81 \mu\text{m}$ for the core/water interface and $S_{GH} = 9.93 \mu\text{m}$ for the core/substrate interface.

These results are consistent with the evanescent fields we determined for this structure in Example 9.6. S_{GH} is much larger for the core / substrate (PVC / BK7) interface than for the core / cap (PVC / water) interface, due to the closer match in refractive indices for the former. As a result, the evanescent field extends further into the BK7, or, using the alternative view of the lateral shift upon reflection, undergoes a larger Goos-Hänchen shift. Both pictures are equivalent: more of the field extends into and propagates in the substrate than in the cap.

Velocity

A final interesting consideration is that of modal velocity. From the discussion above, we realize that a waveguide is a highly dispersive medium, meaning that its optical

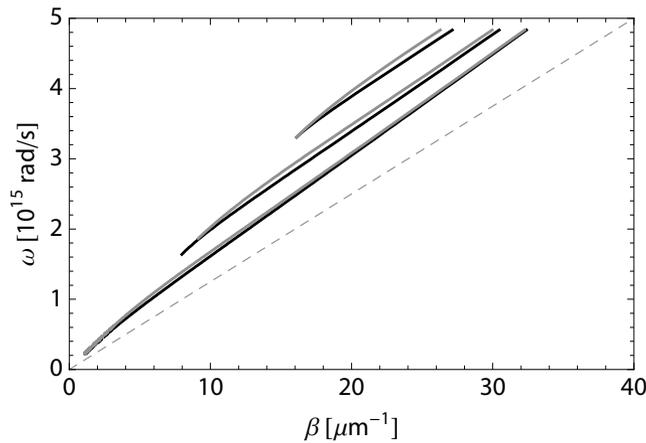


Figure 9.19 Plot of the variation of frequency ω with propagation constant β for a 400 nm thick symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide. The TE *black* and TM *gray* polarizations have been calculated for $m = 0$ *bottom pair*, $m = 1$ *middle pair*, and $m = 2$ *top pair*. The slope $d\omega/d\beta$ corresponds to the group velocity of the mode. The diagonal dotted straight line has been sketched to emphasize the fact that the plotted characteristics are not linear.

characteristics are a strong function of wavelength. Recalling that in Section 2.3 we defined the group velocity of an electromagnetic wave as (Equation 2.47)

$$v_g = \frac{\partial\omega}{\partial k}, \quad (9.58)$$

the velocity of a mode in a waveguide can be found from knowledge of the variation of frequency ω with z -directed propagation constant β . Since, furthermore,

$$k_0 = \frac{2\pi}{\lambda_0} = \frac{2\pi\nu}{c} = \frac{\omega}{c} \quad (9.59)$$

for speed of light in vacuum c , we can solve the characteristic equation of a waveguide for a given λ , thus a given $\omega = 2\pi c/\lambda_0$, and calculate β from the resultant reflection angle. By plotting the variation of ω with β , the modal group velocity can be found from the slope of the characteristic.

We have generated such a characteristic for our favorite symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide, with a 400 nm core, and plotted it in Figure 9.19. Both TE and TM polarizations for $m = 0, 1, 2$ are shown. By comparing with the straight dotted line drawn through the figure, we see that the plotted ω vs. β characteristics are not linear, but slightly curved. Careful examination shows that higher order modes have higher v_g , and TM-polarized modes have a slightly higher velocity than TE-polarized modes.

This result corroborates our previous conclusions based on the extent of the evanescent field or the length of the Goos-Hänchen shift as a function of mode number. Whereas our simple ray-optic waveguide picture gives rise to the impression that high-order modes zig and zag more than lower order modes and thus take longer to move down the waveguide, the more physical electromagnetic picture shows that the evanes-

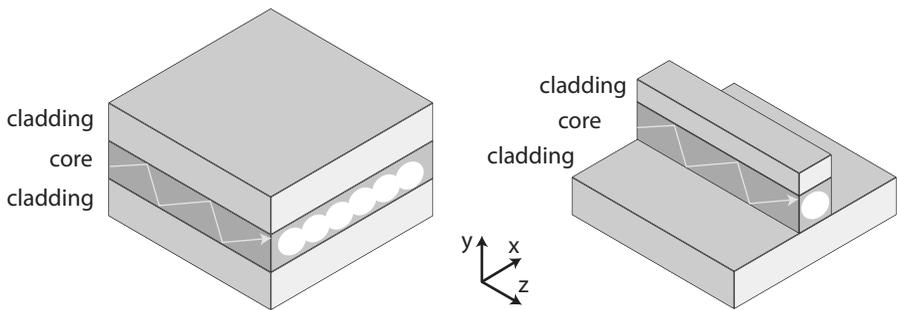


Figure 9.20 Sketch of a one-dimensional slab waveguide *left* and a two-dimensional stripe waveguide *right*, with modal intensity peaks shown schematically. Not to scale

cent field extends more deeply into the cladding for higher order modes, so that the field sees more of the lower-index material in which propagation velocity is higher. The zig-zag picture is rescued by the Goos-Hänchen shift which gives a larger shift down the waveguide per reflection for higher order modes, again resulting in higher modal velocity. We can thus quantify the dispersive behavior of the waveguide explicitly.

9.3 Channel waveguides

The one-dimensional slab waveguide we have considered to this point is a useful structure for understanding guided wave propagation. As shown on the left of Figure 9.20, light in a slab waveguide is guided in the xz plane, propagating in the z direction and confined by the layer structure in the y direction.

Whereas slab waveguide structures are useful in certain sub-disciplines of integrated optics, and form the basis of the broad-area semiconductor lasers we will see in Section 10.3.2, the two-dimensional channel waveguide, shown on the right of Figure 9.20, corresponds more closely to the concept of an “optical wire”, for transmission of photons analogously to electrons in an electrical wire. In a channel (or stripe) waveguide, guiding is in two (the x and y) dimensions, and propagation is in the remaining (z) direction. Channel waveguides, in the various configurations which we will study in this section, form the basis for most integrated optics, optical integrated circuits and integrated photonic devices.

As we see in Figure 9.20, a channel waveguide requires that light is confined in an additional dimension, the x direction in our notation. There are numerous waves of accomplishing this, and the technologies vary strongly with materials and applications; we will look at a sampler here.

9.3.1 Layer structure

With a few prominent exceptions (notably ion exchange or diffusion waveguides), channel waveguides are usually made by further processing of slab waveguides. Layers of

the type that we analyzed thoroughly above are deposited using thin-film or epitaxy techniques, forming a slab waveguide on a suitable substrate. The most important criteria are that the index of the core layer is higher than either cladding and that all (also the cladding layers) are reasonably transparent at the operating wavelength.

Semiconductors

Silicon-based integrated optics is of particular interest for many applications, due to the potential for large-scale integration and the availability of extensive and advanced processing technologies (Hickernell, 1988; Splett and Petermann, 1994; Tsuchizawa et al., 2005; Lipson, 2005); silica (SiO_2) on silicon or SOI (silicon-on-insulator) (Bogaerts et al., 2004) are the most popular variants and were already in use in the 1980s (Soref and Lorenzo, 1986). A popular configuration we have used as an example repeatedly above is Si_3N_4 on SiO_2 with air as the cap layer, or the $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ sandwich; these may be deposited on a silicon substrate, if the lower SiO_2 cladding layer is thick enough to de-couple the waveguide mode from the silicon substrate, which is absorbing at visible wavelengths. Alternatively, glasses on silicon, basically various forms of doped SiO_2 , provide flexibility in routing signals on a Si substrate while also enhancing coupling efficiency to external optical fibers (Henry et al., 1989), as are rare-earth-doped glassy materials (Righini et al., 2001), particularly for the fabrication of active waveguide structures.

In the compound semiconductor area, epitaxial growth techniques allow the formation of complex layer structures using materials for which variations in composition give rise to a suitable range of refractive indices. As we saw in Section 3.5.3, $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is a well-established III-V material whose composition may be varied from GaAs to AlAs. An $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ waveguiding layer sandwiched into $\text{Al}_{0.8}\text{Ga}_{0.2}\text{As}$ cladding layers is a popular configuration, both for passive waveguides and as the basis for semiconductor lasers. Due to their use in optoelectronic integrated circuits, III-V materials (especially the GaAs and InP systems) represent the largest fraction of the integrated optics market.

Polymers

Polymers have been extensively developed for integrated optics, and a dizzying array of materials has been used for waveguide and guided-wave device fabrication (Eldada and Shacklette, 2000; Kaino, 2000; Ma et al., 2002); thermoplastics and photopolymers are particularly popular, including polyimides, polycarbonate and acrylates such as PMMA. Some polymer materials are also attractive due to their tunable optical and electro-optic properties, so that they may be employed not only for passive waveguides but also for a wide spectrum of active optical devices, including lasers, modulators, amplifiers and detectors (Chen, 1993; Lee et al., 1997; Steier et al., 1999; Oh et al., 1999).

Polymers may be deposited using spin or dip coating on a variety of substrates, generally cross-linked using UV exposure, and then patterned using laser direct writing, standard photolithography or molding. The increasing electro-optic functionality of polymers, coupled with their ease of fabrication, implies that these materials will likely play an increasingly important role in guided-wave optics.

Glass and LiNbO₃

The use of glass substrates for integrated optics is well established (Najafi, 1992), and has become an industrial standard, particularly for passive waveguide components such as couplers and splitters. Typically using ion-exchange waveguide fabrication, which we will see below, standard optical-quality substrates, such as BK7, may be used and the waveguides patterned directly. No layer deposition or subsequent patterning is usually required.

Lithium niobate, LiNbO₃, represents a significant share of the integrated optics market, due primarily to the use of this electro-active substrate for waveguide-based modulators and switches; as we saw in Section 3.6.4, the optical, electro-optic, piezoelectric, and photorefractive properties of LiNbO₃ make it a very attractive material for integrated optics (Syms, 1988; Lawrence, 1993). Waveguides are usually formed by titanium diffusion or Li⁺/H⁺ ion exchange. Due to its excellent electro-optic characteristics, this material is very useful for the fabrication of waveguide-based modulators, and at one time represented the largest market for integrated optical components.

Dimensions

Layer thicknesses depend on the materials and the required optical characteristics. Waveguide core thicknesses vary from several hundred nanometers to micrometers; single-mode operation requires a thin core, as we saw above. The cladding layers must be sufficiently thick to decouple the guided mode in the core from the external environment, unless some type of sensor application is desired. Finally, the spacing between the core and an absorbing substrate is often critical, since a large evanescent field extending into an absorbing layer will lead to mode attenuation.

9.3.2 Waveguide stripes

Figure 9.21 summarizes few variations of channel waveguide structures, using different techniques to achieve lateral guiding. The conceptually simplest approach for fabricating a two-dimensional channel waveguide from the one-dimensional slab waveguide is to etch away the waveguiding layers outside the desired channel, yielding the ridge waveguide of Figure 9.21(a). The field is confined laterally³ by the core/air interface, but the etched ridge may also be buried in a further material, so long as its refractive index is lower than that of the core and it is not too absorbing.

Alternatively, the waveguide core layer may be partially, but not completely, etched, yielding the rib waveguide of Figure 9.21(b). Since the core is thinner outside the channel guiding region, either no modes are supported outside of the channel, or the effective index is significantly different from that of the channel so that total internal reflection takes place at the lateral boundaries. A deeper etch leads to a larger difference in effec-

³ We employ the convention that “lateral” is the horizontal, or x , direction, and “transverse” the vertical, or y , direction.

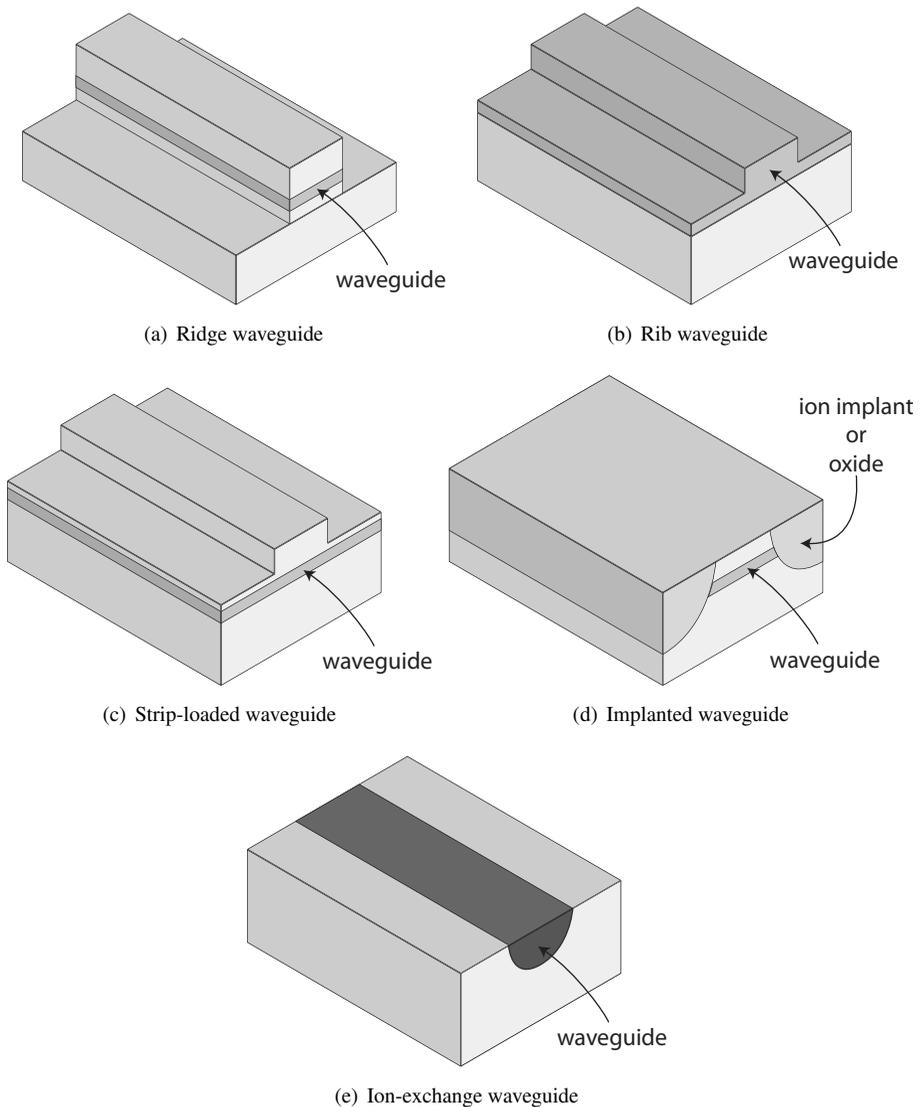


Figure 9.21 Various approaches for the fabrication of channel waveguide structures.

tive index and thus stronger guiding, the limit being that of complete removal, yielding the ridge structure again.

The strip-loaded waveguide of Figure 9.21(c) is based on the same principle, but structures the upper cladding layer, leaving the core intact. Even though the slab waveguide exists everywhere, the strip-loaded waveguide achieves lateral guiding again by a difference in the effective index of the regions below and those to the side of the waveguide stripe. The strip-loaded structure has the advantage that the core remains buried under protective layers, such that the effects of the external environment or surface

roughness and damage due to the etch process are reduced and thereby losses minimized; we will examine this aspect in greater detail in Section 9.4.4.

The above three structures are fabricated by etching and thus have a non-planar surface profile; such surface features are occasionally impractical if the waveguide substrate needs to be further processed or packaged. An alternative means to generate a stripe waveguide is by implantation in the regions outside the channel, using a material which disrupts the optical properties of the core region, as is shown in Figure 9.21(d). Different approaches have been used, such as implantation of H^+ nuclei to disrupt the core and increase the refractive index, or O_2 to generate a buried oxide layer, when using semiconductors.

A final stripe waveguide structure, popular in glass and $LiNbO_3$ integrated optics, is the ion-exchange waveguide of Figure 9.21(e). In this case, a slab waveguide is not required; using a diffusion process, that we will outline in Section 11.5.2, the channel waveguide is defined in a single step as a region with augmented refractive index. The refractive index distribution is that of a typical diffusion profile, and thus does not have a sharp boundary, but forms an effective waveguide with a large mode diameter.

9.3.3 Typical dimensions

Slab waveguide thicknesses, as we saw above, are typically in the range of several hundred nanometers to micrometers. The stripe width is generally a good deal wider, with typical dimensions of several micrometers. Due to the dimensional differences, as well as the poorer confinement in the lateral than in the transverse directions, the mode shapes of many waveguide structures tend to be strongly asymmetric. Waveguide length is arbitrary, and typically limited by the optical losses in the structure. Integrated optical circuits can have waveguides from several micrometers to tens of millimeters in length and optical fibers, a specialized form of waveguide and the subject of Section 9.6, can attain lengths of hundreds of kilometers.

9.3.4 Other waveguide types

There is a considerable variety of other waveguide structures which use a spectrum of physical effects to achieve guiding. Locally-induced stress, for example, may be used to define waveguides using the photoelastic effect (Liu et al., 1996). Photonic crystal waveguides (Mekis et al., 1998; Chow et al., 2001; Krauss, 2003; Bogaerts et al., 2004), based on structures which we will examine in Chapter 14, have allowed the fabrication of waveguides with small mode diameter and the ability to curve these with extremely small radii, even in sharp 90° bends. Slot waveguides (Almeida et al., 2004; Koos et al., 2009) use two closely spaced structures with an etched slot, with dimensions of several tens of nanometers, resulting in an extremely well-confined mode in the slot region, allowing the fabrication, for example, of very compact couplers (Katigbak et al., 2009); such waveguides also show strong non-linear optical effects and are promising for active applications in silicon photonics.

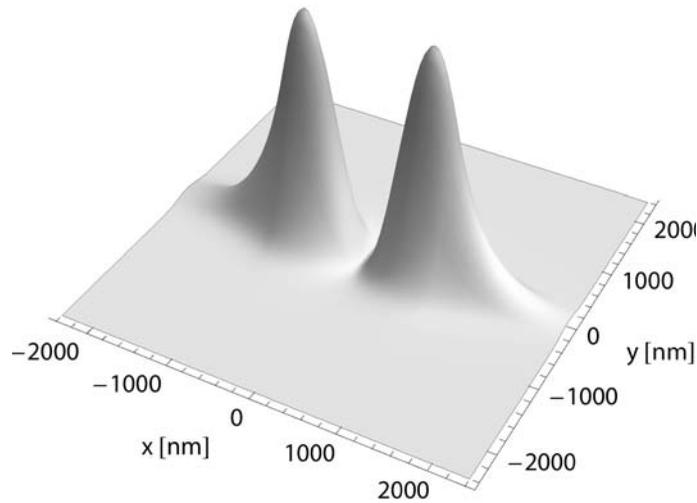


Figure 9.22 Plot of the calculated mode intensity profile for a $0.5 \times 2.0 \mu\text{m}$ waveguide, centered in the figure, with $\lambda = 633 \text{ nm}$; the second order lateral ($m_x = 1$) and first order transverse mode ($m_y = 0$) is shown. The large lateral evanescent fields are clearly seen.

9.4 Waveguide characterization

Channel or stripe waveguides are the workhorses of integrated optics and require analysis in their design and characterization in operation. We provide a brief overview of the measurement techniques and the most relevant measurement parameters.

9.4.1 Numerical analysis

As was the case for the slab waveguide, it is frequently of interest to analyze the channel waveguide mathematically, namely to determine the number of modes and their intensity distribution. The two-dimensional problem which results is generally not solvable analytically, such that numerical approaches are usually applied. A typical mode profile is shown in Figure 9.22, determined for a $0.5 \times 2.0 \mu\text{m}$ waveguide with $\lambda = 633 \text{ nm}$. In this example, two lateral and one transverse mode are supported, and the large extent of the lateral evanescent field is clearly seen.

Numerous approximative approaches have been developed to allow a quick analysis of a channel waveguide by hand, including the effective index approach (Buus, 1982) and the method of field shadows (Marcatili, 1969), and many models have been optimized for particular applications, such as sensors (Stewart and Culshaw, 1994). These approaches were developed when computing power was at a premium: today's students will typically employ commercial software to perform a two-dimensional numerical analysis, generally using finite-element-type simulations (Obayya et al., 2000) or techniques such as the method of lines (Pregla, 2004). Using these tools, the number and types of modes as well as their electric field distribution can be determined for arbitrary refractive index distributions and structural configurations (März, 2004).

More important in many applications is the simulation of the propagation in a waveguide, along the z axis. Using beam propagation (Franco et al., 1999) or vectorial (Hadley and Smith, 1995) models, the development of the field along the waveguide axis can be calculated. Such calculations are of value in determining, for example, the coupling of a mode from one waveguide laterally to another, the losses in Y-couplers or curves and the field distribution in slab-waveguide based structures such as arrayed waveguide gratings; these structures are discussed in Section 9.5 below. Simulations of this nature are difficult to perform by hand.

9.4.2 Input/output

Experimental characterization of a waveguide typically requires that light be coupled into and out of it. Numerous approaches are available to couple an optical signal into a waveguide; butt coupling or end-fire coupling are likely the most direct and most practical, depending on the situation. One key parameter in evaluating coupling techniques is coupling efficiency, the fraction of available light that is coupled into the waveguide. High efficiency is usually only reached if the mode size of the light field to be coupled into a waveguide is approximately the same size as the (two-dimensional) waveguide core; the expected efficiency due just to this parameter may be determined by calculating an overlap integral between the incident mode and the mode supported in the waveguide (Stegeman and Stolen, 1989). Finally, anti-reflection coatings reduce losses due to Fresnel reflection at the facet surfaces, a further potential source of reduced coupling efficiency.

Butt coupling

Butt coupling results when the ends of two waveguiding structures are placed in close proximity, as shown in Figure 9.23. The source can be a passive waveguide, a waveguide-based semiconductor laser or an optical fiber; the latter case is shown in the example in the figure. The emission from the source typically has a Gaussian intensity distribution, to which we were introduced in Section 2.6, and which, as we will see in Section 9.4.3 below, is often asymmetric. However, if the gap between the two elements, d , is sufficiently small, typically less than $1\ \mu\text{m}$, the light emitted by a source is coupled into the destination waveguide.

This approach for coupling light into a waveguide is attractive due to its simplicity, since no further components are required. However, the approach does require high positioning accuracy, since the waveguide dimensions, as we have seen, are typically in the micrometer size regime. The relative positions of the two waveguides typically need to be determined with an accuracy better than several 100 nanometers, since coupling efficiency degrades rapidly as the waveguide cores are offset from one another. As mentioned above, the mode sizes and shapes should be as close to the waveguide dimensions as possible to achieve high coupling efficiency, so that, for example, coupling from a circularly-symmetric fiber into a rectangular dielectric waveguide will only have limited efficiency.

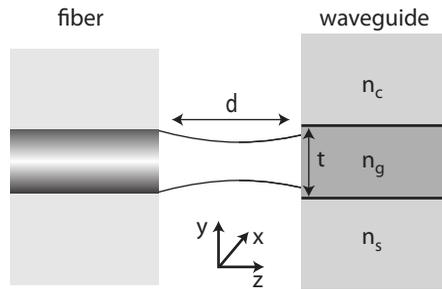


Figure 9.23 Butt coupling of two waveguiding structures; here light is coupled from a fiber *left* into a waveguide *right*. The spacing d should be as small as possible.

Fresnel reflections from the waveguide or fiber facets will not only lead to a reduction in coupling efficiency, but may also, if the two surfaces are parallel to one another, result in a parasitic Fabry-Perot cavity. In that case, small axial (in the z direction) movements can result in the coupling intensity varying periodically, as the resonance and anti-resonance conditions are traversed; this effect can lead to variations in coupling efficiency with time, temperature or vibration.

Nevertheless, butt coupling is often the preferred approach for permanently assembled systems, in which the waveguide structures are optimally positioned and then fixed using gluing or bonding techniques to yield a stable configuration. The types of alignment processes typically employed will be discussed in Chapter 11.

End-fire coupling

An alternative approach for coupling light into a waveguide, more suitable for rapid laboratory characterization, is end-fire coupling, shown schematically in Figure 9.24. Recalling that a Gaussian beam may be focussed to a point with small beam waist, as we saw in Section 7.3.3, a lens or lens system may be used to focus an external optical field onto the facet of a waveguide or fiber. The beam waist width, $2W_0$, should be as closely matched to the waveguide cross-sectional area as possible, and the axial position of the waveguide should be at f , or at least within the depth of focus, $2z_0$.

Facet reflection also plays a role in defining coupling efficiency for this configuration, as it did for butt coupling; anti-reflection coatings will help matters. Due to the extra components and alignment equipment required for this coupling technique, it is more useful in the laboratory and is not usually used for permanently assembled systems. Demands on alignment are slightly more relaxed than for butt coupling, and a suitable optical arrangement can allow imaging of the facet through the lens, allowing accurate positioning of the focus spot on the waveguide core.

Example 9.8: Can light be coupled into a waveguide using butt or end-fire coupling at any input angle?

Consider the coupling of a beam at an angle θ_i from the facet normal into a dielectric waveguide as shown in Figure 9.25. Since a guided mode must satisfy TIR at the

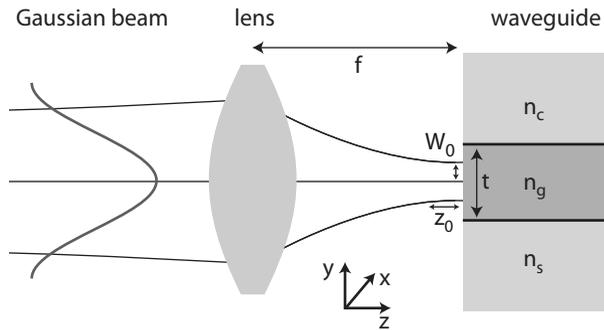


Figure 9.24 End-fire coupling into a waveguide. An external light source, represented by a Gaussian beam, is focused onto the end facet of a waveguide. The beam waist width $2W_0$ should be as close to the waveguide core thickness t as possible, and the waveguide facet should be placed within the depth of focus, $2z_0$.

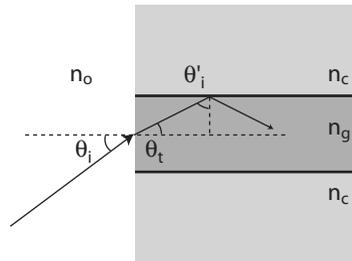


Figure 9.25 Coupling into the facet of a waveguide at an angle θ_i ; since TIR must be satisfied at the core/cladding boundary, there is a maximum value for θ_i from which the NA of a waveguide or fiber can be determined.

core/cladding boundary, there is a minimum allowed value for θ'_i , namely the critical angle $\arcsin(n_c/n_g)$. As a result, there is a maximum value of θ_t and thus θ_i : for beams coupled in at $\theta > \theta_i$, total internal reflection does not take place inside the waveguide and the mode is not guided. The answer to the question above is then: no.

It can be shown (see Problems) that a maximum allowed value for θ_i leads to a numerical aperture, NA_{wg} , given by

$$NA_{wg} = n_0 \sin \theta_i = \sqrt{n_g^2 - n_c^2}$$

for coupling into a waveguide from a material with refractive index n_0 , air in this case. We see that NA_{wg} , which defines the largest angle from the facet normal at which light can be coupled into a waveguide, is only a function of the relative refractive indices of the core and cladding, but not the waveguide thickness, for example.

Looking at typical values for a PVC waveguide core ($n_g = 1.540$) deposited on a BK7 substrate ($n_c = 1.517$), we find that $\theta_i \approx 15^\circ$ and $NA \approx 0.27$. If we instead use two glasses, such as SF6 ($n_g = 1.80518$) on BK7, these values improve to $\theta_i \approx 78^\circ$ and $NA \approx 0.98$, showing that, as expected, a larger index step between core and cladding results in a smaller critical angle, and thus a larger maximum coupling angle and NA.

Grating coupling

Coupling light into (and out of) a waveguide using one of the two techniques above requires that the facet (normal to the propagation direction of the waveguide) is accessible, and is of sufficient quality such that scattering and other losses due to a rough or non-planar facet are negligible. Fabricating a smooth, planar facet is relatively easy in many semiconductors, where cleaving may be employed, but is more difficult with non-crystalline materials, including dielectrics and polymers.

An alternative approach for coupling light in or out is to use a grating etched onto the channel waveguide surface. We recall from Section 8.3 that an optical field incident onto a periodic corrugation will be diffracted into various diffraction orders. Such a diffraction order can be used to couple light into a waveguide from the surface, if the wavelength λ_0/n_g and the grating period Λ are closely matched.

In particular, for a waveguide which has a supported mode with propagation constant β_0 , a grating on the surface will augment the propagation constant to a value β_q given by

$$\beta_q = \beta_0 + q \frac{2\pi}{\Lambda} \quad (9.60)$$

for integer $q = \pm 1, \pm 2, \pm 3, \dots$. If $|\beta_q| < n_c k_0$, for cap material n_c (typically air, so $n_c = n_0$) and $k_0 = 2\pi/\lambda$, then the harmonic q will radiate from the waveguide surface. This radiation will emerge at an angle θ_q found from

$$\theta_q = \arcsin \left(\frac{\beta_q}{n_0 k_0} \right). \quad (9.61)$$

Since optical systems are typically⁴ reciprocal, meaning that they have the same propagation conditions regardless of whether light moves forward or backward through the system, the result above implies that an optical field incident at angle θ_q onto the grating will be coupled into the waveguide. The grating coupler thus requires that a collimated optical field is incident onto the corrugation at the correct angle, and can give rise to high coupling efficiencies without the need for a waveguide facet.

Other approaches

Numerous other approaches for coupling light into waveguides have been traditionally used. Prism coupling is prominent among these as a laboratory technique. A prism, made of a material with a refractive index higher than that of the waveguide, is placed on top of the waveguide substrate. Light is directed into the prism at such an angle that the evanescent field couples across the prism / waveguide interface. If the appropriate prism materials can be obtained, a difficult undertaking for most semiconductors, this technique allows a simple setup for coupling from the waveguide surface.

⁴ But not always. Non-reciprocal optical systems, such as magneto-optical crystals, which we saw in Section 3.6.4 and will revisit in Chapter 10.6.1, behave differently when traversed forward or backward.

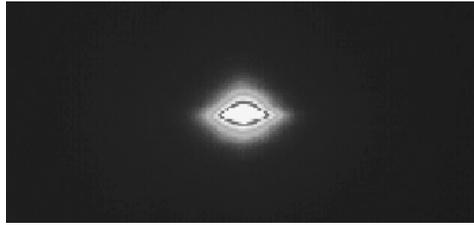


Figure 9.26 Near-field image of the asymmetric intensity profile of the emission from a typical semiconductor strip-loaded waveguide, with dimensions $3\ \mu\text{m} \times 190\ \text{nm}$ in the x and y directions, respectively. The corresponding far-field distribution is given in Figure 9.28.

9.4.3 Near- and far-field measurements

Once light has been coupled into a waveguide, it is of interest to determine the actually supported modes, their number, shape and distribution. In addition, many applications require that the light is coupled back out of the waveguide into another component, in the manner we just saw in the previous section.

Near-field

The modal intensity distribution is most practically determined by a near-field measurement. Typically using a CCD camera and an appropriate objective lens, the optical intensity on the facet is imaged directly, yielding a profile such as that seen in Figure 9.26. It is clear from this measurement of a strip-loaded waveguide (Figure 9.21(c)) that the profile is distinctly asymmetric: it is wider in the lateral direction than in the transverse direction, and transversely the intensity extends further upwards than downwards.

A more detailed look at the supported modes in a waveguide is possible by reducing the intensity (either by diminishing the source intensity or using neutral density filters at the output) and imaging the facet again. The types of intensity distributions which may result are shown in Figure 9.27 for a series of eight waveguides of varying dimension. These III-V-based waveguides all have the same $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ core material and thickness, with waveguide widths varying from 1.5 to $3.0\ \mu\text{m}$. The two rows of images differ by the material of the cladding layers: the top row, using a $\text{Al}_{0.8}\text{Ga}_{0.2}\text{As}$ cladding, has a larger index step between core and cladding than the lower row, which employs a $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$ cladding.

We see from the near-field characteristics that the supported mode number is not only a function of width, but also index step between core and cladding. As the width increases, for both material systems, the $m = 1$ mode becomes more prominent and the separation between the peaks more distinct. Comparing top and bottom characteristics, we see that the lowest order ($m = 0$) mode is only clearly visible for the $1.5\ \mu\text{m}$ waveguide in the lower row. The dimensional and material dependencies seen in this series of measurement neatly confirms the theoretical predictions of the previous sections.

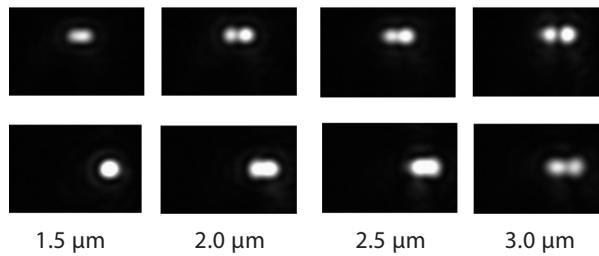


Figure 9.27 Images of the supported modes in a series of eight III-V waveguides of varying composition and dimension. All have a $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ core; the upper row has a $\text{Al}_{0.8}\text{Ga}_{0.2}\text{As}$ cladding and the lower $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$ cladding. The lateral waveguide width is given for each column. Higher-order modes are more prominent for wide waveguides and strong confinement.

Far-field

The near-field image on the waveguide facet corresponds to the intensity distribution of the guided modes in the core. When coupling light out of the waveguide into another optical component, the intensity distribution at some distance away from the facet becomes the relevant parameter, corresponding to the far-field characteristic.

The end of a waveguide (or fiber) is an optical aperture, exactly in the sense that we considered this in Section 8.2. Recall that diffraction effects strongly affect the far-field of the emission pattern from an optical aperture, and in particular that, the narrower an aperture becomes, the wider its diffraction intensity distribution normal to the aperture dimension becomes. Since a typical dielectric or semiconductor waveguide has a relatively thin core (t typically hundreds of nanometers) and a relatively wide width (w typically several micrometers), we expect that diffraction effects may give rise to an asymmetric far-field distribution.

Measuring the far field distribution, most practically with a photo-goniometer (Zappe, 2004, Chap 5.6.2), we indeed see a strong asymmetry, as seen in Figure 9.28. This type of measurement typically plots intensity as a function of divergence angle, and it is clearly seen that the transverse far-field is considerably broader than the lateral distribution, not unexpected since the waveguide is much wider than it is thick. Such an elliptical intensity distribution is also characteristic of edge-emitting lasers, and is frequently termed astigmatic (since the effective focal points of the lateral and transverse fields inside the waveguide are at different positions). Anamorphic lenses are generally required to correct such highly-astigmatic emission before coupling into circularly-symmetric optics.

9.4.4 Losses

In our analysis of waveguides to this point, we have tacitly assumed that all the optical energy in a supported mode is guided in the core, without any loss of intensity as it propagates axially. Waveguides are of course fabricated using real optical materials, and we saw in Section 3.3 that these give rise to attenuation due to absorption, scattering or other processes. As a result, the optical field is attenuated during propagation,

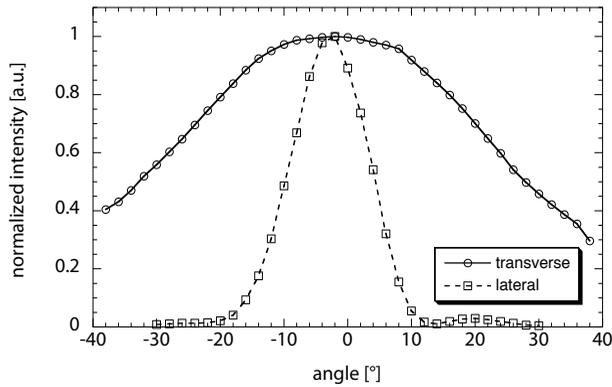


Figure 9.28 Measurement of the far-field divergence of a III-V-based strip-loaded waveguide, of the type defining a semiconductor laser. The lateral (x) dimension, given by a $3\ \mu\text{m}$ ridge width, has a far-field angle with a FWHM of 15° ; the transverse (y) direction is defined by a $190\ \text{nm}$ wide core region and has a far-field angle with FWHM of 51° . The near-field distribution of this waveguide was seen in Figure 9.26.

and characterization of waveguide loss is important to allow an estimate of the optical energy which emerges from it.

The decrease in intensity with propagation of a mode in a waveguide is given by Beer's law, to which were introduced in Section 3.3.1. The z -dependent intensity in the waveguide, $I_{wg}(z)$, thus decreases with propagation distance as

$$I_{wg}(z) = I_{wg0} e^{-\alpha_{wg} z} \quad (9.62)$$

where α_{wg} is the absorption coefficient in the waveguide and I_{wg0} the initial intensity; as in Chapter 3, units of decibels or m^{-1} are used for α_{wg} in waveguides.

Loss mechanisms

The magnitude of α_{wg} depends not only on material absorption but also on structural parameters of the waveguide. Losses due to absorption, as discussed for bulk substrates in Section 3.3, also apply in waveguides, where band-edge and free-carrier absorption may play a dominant role in semiconductor structures, particularly if these are doped. These mechanisms often have a strong wavelength dependence: for wavelengths close to the bandgap for a semiconductor, band or Urbach tails lead to a rapid increase in absorption as wavelength decreases incrementally, even below the bandgap energy. Free-carrier absorption, on the other hand, due to electrons in the conduction band, increases as λ^2 or λ^3 , depending on the material (Fan, 1967).

Scattering effects in waveguides may be due to inhomogeneities in the medium, as for bulk materials, but in waveguides, scattering from structural features may dominate. Roughness at the core/cladding interface, or at the stripe edge (particularly for the deeply etched rib or ridge structures shown in Figures 9.21(a) and 9.21(b), respectively) leads to scattering losses which generally increase as wavelengths decrease

(Tien, 1971). Very smooth, epitaxially grown layers combined with smoothly etched sidewalls or buried structures are effective in reducing scattering losses.

Waveguides are also, finally, prone to radiation losses, particularly in curves. As a waveguide changes axial direction, the mode is pulled to the outside of the curve, giving rise to losses which increase exponentially with decreasing curve radius (Marcatili and Miller, 1969), an effect which becomes worse as the mode is more poorly guided. The total waveguide loss is then the sum of all of these effects, such that determination of α_{wg} theoretically often leads to very approximate and generally unreliable values.

Loss measurement

As a result, measurement of waveguide loss is typically required in the evaluation of an integrated optical design. A number of measurement approaches is useful for the integrated optical engineer. The cut-back technique is perhaps the most easily understandable and can be employed without the need for advanced equipment. Light is coupled into a waveguide of length L_1 and the emitted intensity I_1 measured. The waveguide is then cut into two pieces, preferably of uneven length, and at least one more intensity measurement is made for a length L_2 , yielding I_2 . The optical loss of the waveguide may then be found by applying Beer's law to yield

$$\alpha_{wg} = \frac{1}{L_1 - L_2} \ln \left(\frac{I_2}{I_1} \right). \quad (9.63)$$

A more accurate determination can be made if a series of lengths is measured and the logarithm of intensity is plotted as a function of length, the slope giving the loss value. Although the method is simple, it relies on repeatable coupling of the input field into each of the measured waveguides and very low-loss waveguides may require unmanageable differences in length for a reliable calculation.

An alternative approach, which actually becomes more accurate for lower waveguide losses, makes use of the Fabry-Perot resonances in the resonant cavity formed by the two end facets of a waveguide; recall that we examined the Fabry-Perot etalon in Section 5.2.5. A usable resonant cavity is only usually found for semiconductor-based waveguides, particularly III-Vs, since cleaving forms high-quality parallel facets. In this approach, the wavelength of the light coupled into the waveguide is varied slightly and the emitted intensity monitored. Since a wavelength scan results in periodic resonance and anti-resonance, the contrast ratio of this interference pattern, $C = I_{max}/I_{min}$, can be measured, and the loss may then be found from (Walker, 1985)

$$\alpha_{wg} = \frac{1}{L} \ln \left[\frac{1}{R} \frac{\sqrt{C} - 1}{\sqrt{C} + 1} \right] \quad (9.64)$$

for waveguide length L and facet reflectance R . Either a tunable laser, or a diode laser slightly tunable by temperature, is required for this measurement, although only a few nanometers of tuning range is required. For dielectric or polymer waveguides with relatively poor facet quality, however, this approach is unsuitable and the cut-back technique is generally used.

Typical values

Due to the wide variety of materials and structures used for waveguides, ascertaining “typical” values for waveguide loss is difficult; in addition, we recall that most loss mechanisms are strongly wavelength-dependent. III-V semiconductor structures designed for low loss have been characterized with less than 0.1 dB/cm (Deri and Kapon, 1991). Silicon-based waveguides, using Si on SiO₂ or Si/Ge have been fabricated with $\alpha_{wg} < 0.5$ dB/cm and silicon-based dielectric waveguides are generally in the range of 0.3 dB/cm. Polymer or LiNbO₃ waveguides can be very low loss, with state-of-the-art around 0.01 dB/cm at visible wavelengths (Eldada and Shacklette, 2000). Losses for photonic crystal waveguide tend to be higher, in the range of 20 dB/cm (Kotlyar et al., 2004). As a point of reference, optical fibers, to which we will turn presently, generally have $\alpha_{wg} \leq 10^{-5}$ dB/cm (≤ 1 dB/km), another league altogether.

9.5 Waveguide components

Now that the characteristics of stripe waveguides are clear, we consider a few *passive* waveguide-based components which form essential parts of integrated optical systems. Waveguides also form the structural basis for a wide variety of *active* components, including semiconductor diode lasers, modulators and detectors; we will defer consideration of these to Chapter 10.

9.5.1 Curves

For the channel waveguides presented and analyzed above, we tacitly assumed that propagation in the axial (z) direction was in a straight line. It is of course practical to bend a waveguide in the lateral (x) direction, allowing it to curve on the substrate, and thereby generate flexibility in the layout of integrated optical circuits.

Waveguide curves generally take on one of a number of shapes, including circular, for which

$$x = \sqrt{r^2 - z^2} \quad (9.65)$$

where R represents the curve radius. Alternatively, the raised-cosine form, given by

$$x = L_x \cos\left(\frac{\pi z}{2L_z}\right), \quad (9.66)$$

may be employed; in this last expression, L_x represents the lateral distance the curve traverses, and L_z the axial distance required for it to do so, as is shown in Figure 9.29.

We mentioned in Section 9.4.4 that the radiation losses in a waveguide curve increase exponentially with decreasing curve radius. As a result, for both of these curve forms, the effective radius of curvature, and therefore the transition distance $2L_z$ for a given required translation, is kept relatively large to maintain low losses. A tight curve results in considerable excess loss, and the typical radii employed for dielectric or semiconduc-

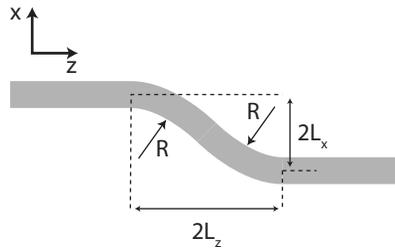


Figure 9.29 Two waveguide curves of radius R resulting in a lateral translation of $2L_x$ over a transition distance of L_z . Not to scale.

tor waveguides of the type shown in Figure 9.21 are in the range of several millimeters. Curve loss considerations imply that many integrated optical circuits require considerable substrate surface area. Due to the completely different nature of guiding in a photonic crystal waveguide, on the other hand, sharp 90° bends are possible in these structures.

9.5.2 Y-junctions

It is frequently of interest to divide the optical intensity in a waveguide into two or more output waveguides, the function of a waveguide splitter; alternatively, the fields from two waveguides sometimes need to be combined into a single output waveguide. The Y-junction shown at the top of Figure 9.30, or in Figure 9.31(a), is the simplest means to accomplish these functions. For use as a splitter, the input waveguide slowly widens, until two modes are supported in a transition region, which then splits into two output waveguides; combining two optical fields into a single waveguide is accomplished by operating it reverse. Combining two input fields into a single output waveguide is accompanied by a loss of 3 dB (i.e., half the input intensity) since higher-order modes in the taper radiate into the substrate and are not coupled into the (typically single mode) output waveguides.

To reduce radiation losses and assure that further higher-order modes are not generated, the taper of the transition region has a very small angle and the curves of the output waveguides are designed with a large radius of curvature. It has been shown that maintaining a taper angle θ_{taper} with a value below (Kuznetsov, 1985)

$$\theta_{taper} < \cos\left(\frac{n_c}{n_g}\right) - \arctan\left(\frac{k_{xg}}{\beta}\right) \quad (9.67)$$

will significantly reduce excess radiation losses in the Y-junction. In the previous expression, n_c represents the refractive index on the *lateral* side of the waveguide core and k_{xg} the x-directed propagation constant, also in the lateral direction. We see that a larger index step between core and cladding (implying better confinement) allows a larger taper angle; typical values for semiconductor waveguide Y-couplers are on the order of 4° .

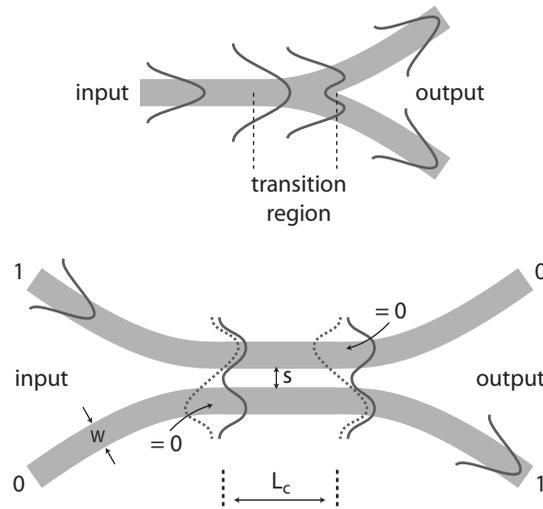


Figure 9.30 Two types of waveguide couplers. The Y-coupler *top* splits the field in one input waveguide into two output waveguides. The proximity coupler *bottom* employs two waveguides closely spaced at s from one another, overlapping over a distance L_c , to couple a field from one guide to the other. Both are not to scale.

The Y-coupler can be extended to split an incoming signal into a larger number of output waveguides, using a succession of branches in series. 1×4 , 1×8 , and 1×16 couplers are popular for optical signal distribution. Due to the small taper angles and large curve radii, these correspond to relatively large integrated optical chips.

9.5.3 Proximity couplers

Two waveguides in close lateral proximity, as shown at the bottom of Figure 9.30 and in Figure 9.31(b), may have an overlap in their evanescent fields. As a result, the optical field from one guide may excite a field in the other and, under appropriate conditions, a portion or all of the energy may be coupled from one waveguide to another.

Analysis

Given two identical waveguides separated by a spacing s , as shown in Figure 9.30, with propagating waveguide fields both moving in the same direction of the form

$$E_1 = E_{10}(x, y)e^{-j\beta z}e^{j\omega t} \quad (9.68)$$

$$E_2 = E_{20}(x, y)e^{-j\beta z}e^{j\omega t}, \quad (9.69)$$

these two fields may be related by the coupling equations, given by (Yariv, 1973)

$$\frac{dE_{10}}{dz} = \kappa E_{20} \quad (9.70)$$

$$\frac{dE_{20}}{dz} = -\kappa E_{10}. \quad (9.71)$$

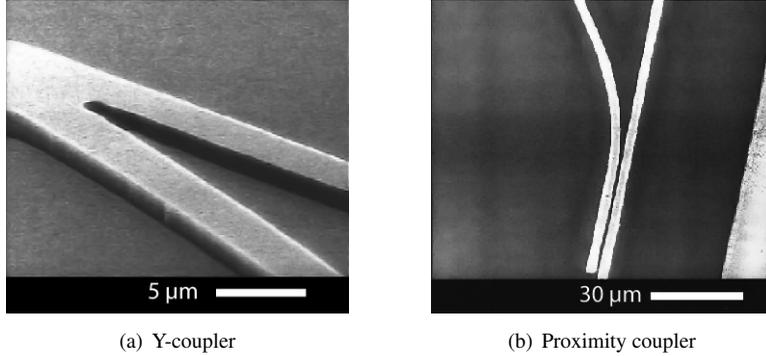


Figure 9.31 SEM photographs of etched III-V semiconductor-based waveguide couplers.

The coupling coefficient κ [m^{-1}], assumed to be real in the absence of waveguide losses, may be expressed as (Somekh et al., 1973)

$$\kappa = \frac{2k_{xg}^2 k_{xL} e^{-k_{xL}s}}{\beta W (k_{xL}^2 + k_{xg}^2)} \quad (9.72)$$

for x-directed propagation constant in the waveguide k_{xg} and lateral decay constant (between the waveguides) k_{xL} ; W is the waveguide width. From the previous expression, we see that the coupling coefficient varies exponentially with waveguide spacing, so that close and accurately spaced waveguides must be part of the design and fabrication of a proximity coupler.

To illustrate the mechanics of coupling, consider the case where we have one non-zero input field, E_2 , and $E_1 = 0$, at the beginning of the overlap region, with $z = 0$. The solutions to the coupling equations then become

$$E_{10}(z, t) = E_2 \sin \kappa z \quad (9.73)$$

$$E_{20}(z, t) = E_2 \cos \kappa z \quad (9.74)$$

representing the amplitudes of two propagating waves, with a relative phase shift of π . Examining these solutions, we see that they imply a periodic transfer of energy from one waveguide to the other: for $\kappa z = 0$, all the energy is in E_{20} and for $\kappa z = \pi/2$, all of it is in E_{10} . The maxima and minima vary periodically with distance so long as the waveguides are in close proximity, with a period which may be expressed as a coupling length L_π given as

$$L_\pi = \frac{m\pi}{2\kappa} \quad (9.75)$$

for integer m . For every length of L_π along the overlap region between the waveguides, all the optical energy is transferred from one waveguide to the other and back again.

As a result, proper design of the overlap length, L_C in Figure 9.30, allows the coupling of a desired fraction of the light between the waveguides: for $L_C = 0.5L_\pi$, half the energy will be transferred (corresponding to a 3 dB coupler) and for $L_C = L_\pi$, all

the energy will be transferred. As mentioned above, accurate design and high-resolution fabrication is required to achieve the desired functionality for this type of coupler.

Example 9.9: Given a 1 μm wide etched ridge waveguide structure, we fabricate a proximity coupler with a spacing of $s = 0.8 \mu\text{m}$. We know from a previous analysis (Zappe, 1995, Chapter 8.6) that the structure has the following propagation constants: $k_{xg} = 2.89 \mu\text{m}^{-1}$; $k_{xL} = 2.49 \mu\text{m}^{-1}$; and $\beta = 24.13 \mu\text{m}^{-1}$. What overlap length L_C is required for a transfer of 3 dB from one waveguide to the other, i.e., a 50:50 power splitter? What happens if our etch process yields a spacing of $s = 0.9 \mu\text{m}$ instead?

From Equation 9.72, we can easily determine the coupling coefficient to be $\kappa = 0.016 \mu\text{m}^{-1}$, which we then plug into Equation 9.75 to yield L_π . Since we require a 3 dB coupler, the overlap length is half the coupling length, so that we obtain $L_C = 0.5L_\pi = 48.6 \mu\text{m}$.

If an overetch leads to a spacing of 100 nm more, so that $s = 0.9 \mu\text{m}$, a repeat of that calculation leads to a significantly different value, namely $L_C = 62.3 \mu\text{m}$. Since the overlap length is fixed by the layout of the waveguide design, this small change in spacing would result in a significant departure from the desired 50:50 ratio of output intensities. We can estimate the intensity ratio which results if we use the 0.8 μm design ($L_C = 48.6 \mu\text{m}$) with a spacing of $s = 0.9 \mu\text{m}$ by taking the square of Equations 9.73 and 9.74 (since we are interested in intensities) with the argument $\kappa z = (0.013 \mu\text{m}^{-1} \cdot 48.6 \mu\text{m})$, where the value for κ results from application of Equation 9.72 with $s = 0.9 \mu\text{m}$. We find that the splitting in this case is almost exactly 67:33, where most of the intensity is in the waveguide in which the original field was found.

9.5.4 Other passive components

There is a variety of other passive waveguide components useful for the conception of integrated optical circuits and guided-wave functionality. For many of these, as for the passive couplers we just discussed, adding electro-optic functionality through the incorporation of electrical contacts allows the fabrication of waveguide-based switches or modulators, active optical devices which we discuss in Chapter 10.

Of particular interest in passive waveguide devices are multimode interference (MMI) couplers (Soldano and Pennings, 1995) which are based on accurately designed two-dimensional slab waveguides in which standing waves generate two-dimensional interference patterns. By placing output waveguides at the proper positions, wavelength-dependent coupling can be achieved at the positions of maximum constructive interference. The MMI coupler is well suited for high-fanout applications, where a single input signal is to be distributed to a large number of output waveguides and the losses associated with, say, cascaded Y-couplers are excessive; we will see it again in Section 9.7.1 below.

The arrayed waveguide grating (AWG) is formed by a large array of waveguides of incrementally differing length (see Figure 9.41 below); using a MMI coupler to couple a broad-band signal into all the waveguides and a second MMI coupler at the output, the spectrum is spread into an array of output waveguides, effectively simulating a high-order grating (Takahashi et al., 1990; Fukazawa et al., 2004). Using silicon or glass substrates, arrayed waveguide gratings with up to 1337 waveguides have been demonstrated, with a wavelength resolution down to 0.4 nm and channel crosstalk below –30 dB. These structures are popular for de-multiplexing optical telecommunications signals (where up to 512 signals at closely spaced wavelengths are transmitted down a single fiber) in the 1.55 μm wavelength range and play a major role in optical telecommunications systems.

A final useful passive waveguide device with a spectrum of applications is the resonant ring filter or ring resonator (Little et al., 1997, 2004). Consisting of a waveguide ring closely spaced to a linear waveguide (forming a proximity coupler at a single point), the device acts as a highly-efficient wavelength filter. The rings, with radii of several micrometers, form high-Q resonators, with values up to 10^8 demonstrated for this type of structure (Vahala, 2003); Q [] represents the quality factor of a resonator. At a certain resonant wavelength, energy is coupled from the straight waveguide into the ring, and a higher Q implies a sharper absorption peak, in wavelength. As a result, the resonant ring structure can be employed as a high-resolution waveguide filter and has also been used as a high-sensitivity sensor structure, relying on the fact that absorption of an analyte on the waveguide surface, for example, changes its effective index and thus the resonance wavelength.

9.6 Optical fibers

Optical fibers are perhaps only two-thirds micro-optics: typical core dimensions are in the range of several micrometers, but the axial dimension can extend to over 100 km. Despite their extreme aspect ratio (up to 10^9), optical fibers are useful both in the optical measurement laboratory and as interconnect components in optical microsystems. For this reason, we provide a brief overview of their structure and functionality here, referring the reader to the literature for a more in-depth treatment (Murata, 1996; Derickson, 1998; Bass and van Stryland, 2001; Buck, 2004).

9.6.1 Fiber structure

The optical fiber is a guided-wave structure which perhaps most closely approximates the concept of an “optical wire”. As shown in the schematic sketch of Figure 9.32, the fiber is a cylindrical structure with a circularly-symmetric core region surrounded by a cladding, with a standard thickness of 125 μm . Typical cross-sections of the index profile of the fiber core are shown in Figure 9.33, where the exact structure used depends strongly on the desired operating characteristics. Optical fiber is usually glass-based with doping used to raise the refractive index of the core slightly above that of the

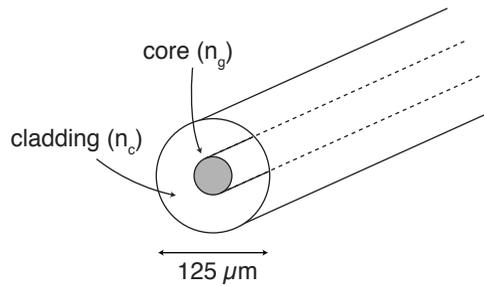


Figure 9.32 Schematic cross-section of an optical fiber, with the circularly-symmetric core surrounded by a cladding with a standard diameter of $125\ \mu\text{m}$. Not to scale.

cladding; typical values for Δn are in the range of 10^{-3} . Polymer-based fibers, of rapidly growing importance for short-range communication, have similar structures and index profiles, as we will see below.

Core variations

Figure 9.33 shows three possible variations of the fiber core structure. The external diameter of the glass cladding is fixed at $125\ \mu\text{m}$, a value which has been established as an industrial standard. A standard fixture in most any laboratory, the multimode fiber, as seen in Figure 9.33(a), has a core diameter of typically 50 or $62.5\ \mu\text{m}$ and a refractive index step of about 0.02 . A “step-index” profile is one for which the core refractive index makes an abrupt step at the core/cladding boundary, as seen in the figure and as was the case for the waveguides we examined in the previous sections. Multimode fiber supports hundreds of modes but is attractive due to the ease with which light can be coupled in and out of the large core.

In contrast, single-mode fiber, shown in Figure 9.33(b), also has a step-index profile, but with a core diameter on the order of 5 to $10\ \mu\text{m}$. Due to the small index step (again on the order of 0.02), this small diameter implies that only a single mode is supported. Single-mode fiber forms the backbone of long-range optical telecommunications, since multimode dispersion effects, which limit modulation bandwidth, are eliminated for single-mode structures. Due to the small core, coupling light into a single-mode fiber can be a headache in the laboratory, but the dimensions involved are not smaller than those for integrated optical waveguides.

A final variation, graded-index fiber, uses a continuously varying refractive index profile, of the same form as the GRIN lens we saw in Section 7.6.2. As shown schematically in Figure 9.33(c), the refractive index varies parabolically from a maximum at the center with a radial variation of the form

$$n(r) = n_m \left[1 - \frac{n_m - n_c}{n_m} \left(\frac{r}{a} \right)^2 \right], \quad (9.76)$$

where n_m represents the maximum refractive index value and n_c that of the cladding; a [m] is the maximum radial extent of the refractive index profile. Due to the radial refractive index profile of graded-index fiber, higher-order modes, with more oblique

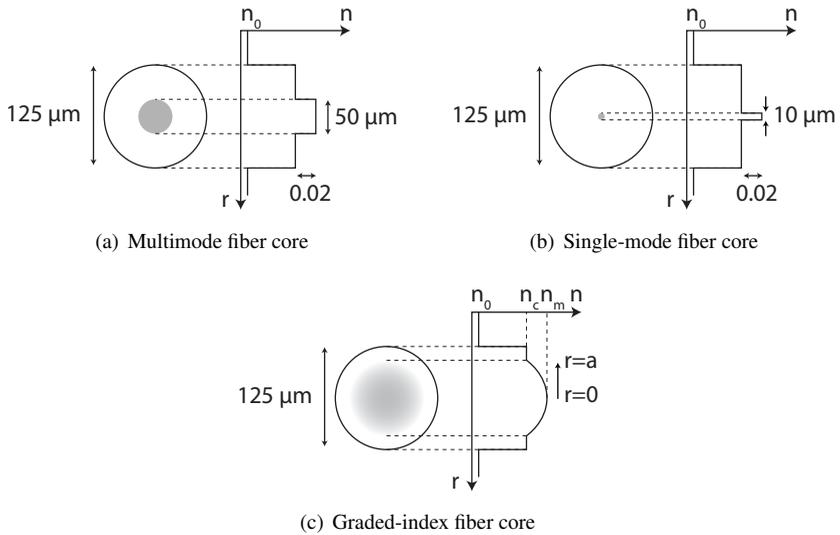


Figure 9.33 Schematic cross-sections of typical optical fiber core refractive index distributions. Not to scale.

trajectories relative to the axial dimension of the fiber, traverse more of the low-index region far from the fiber center, where the velocity is higher. For the same reason that GRIN lenses focus beams incident at a spread of angles to a single point, graded-index fibers result in propagation of all modes at the same speed, so that multi-path dispersion is strongly reduced.

There is a considerable variety of other types of fiber core structures. We will discuss dispersion-shifted and dispersion-flattened fiber in Section 9.6.5 below. Polarization-maintaining fiber, which uses complex asymmetric core structures, is used for applications in which the polarization direction must be determined, since normal fiber tends to rotate the polarization randomly due to the circularly symmetric core. Finally, hollow-core or photonic crystal fiber (Russell, 2006) uses transmission through air or glass in a complex fiber core consisting essentially of a radially-distributed photonic crystal; this structure is fundamentally different from standard optical fiber. Reminiscent of a hollow microwave waveguide, as seen in Figure 14.12, this so-called “holey fiber” can feature strongly reduced dispersion and may also be used for specialized applications such as pulse compression; we will consider this type of fiber in Section 14.2.6.

Cables and connectors

A 125 μm diameter glass fiber is relatively delicate so that for laboratory or external use, optical fiber is usually configured as a cable (Gowar, 1993, Chapter 11.3). A typical cable consists of a 3 mm polyurethane jacket, which protects the fiber at the center and is sufficiently flexible to wind its way around an optical table. Fiber deployed for medium or long-haul applications outdoors, often buried along with power cables or deposited at

the bottom of the ocean, is usually arranged as part of a fiber bundle, placed in a copper tube and then surrounded by a sheath of steel wire; deep sea fiber cables have a diameter of up to 25 mm and, at least in coastal waters, have to resist shark bites (Hecht, 1999, Chapter 15).

Fiber cables are generally configured with connectors at either end, to allow quick connection to the optical system while assuring high coupling efficiency (Shuto et al., 1999). There are a number of connector types which have become established as industrial standards, and these are fabricated with sufficient precision that even single mode fibers, with core diameters in the range of 5 μm , may be simply plugged into a system with high alignment accuracy. Single-fiber connectors are complemented by one-dimensional fiber arrays which connect multi-fiber ribbons to one another, in which the fibers are positioned with a pitch of 250 μm , usually using silicon V-grooves for alignment.

9.6.2 Fiber materials

Most optical fiber is based on glass materials, typically extremely pure SiO_2 . The core region is doped, usually with Ge, P or B, to attain the desired refractive index distribution; up to 13.5% of the core material is replaced by the dopant, but refractive index steps, as we saw above, are small, typically less than 0.02. Typical values for core and cladding index are, for example, $n_g = 1.4515$ and $n_c = 1.4470$, yielding $\Delta n = 0.0045$. Glass fiber materials are of exceptional purity and have lead to overall optical loss figures of less than 0.2 dB/km; we discuss loss further in Section 9.6.4 below.

Glass-based optical fibers are manufactured (Gowar, 1993, Chapter 11) by pulling a strand of glass from the melt of a pre-form, typically a 2 cm diameter glass cylinder which has been manufactured to have the same refractive index profile as the fiber, appropriately scaled in diameter. Heated to temperatures approaching 2000°C in a fiber pulling tower, the fiber is pulled from the melt at speeds of 5 to 10 m/s with a diameter tolerance of < 0.1% and immediately coated with a thin polymer layer, to prevent indiffusion of water. In this manner, 125 μm diameter fiber is fabricated kilometers at a time.

For so-called short-haul optical telecommunications (distances less than about 100 m), the regime of optical LAN (local area networks), glass fiber is often too expensive; plastic optical fiber (POF) has become popular for these applications (Marcou, 1997; Weinert, 1999; van den Boom et al., 2001), and can also be of interest for use in the laboratory or for highly-parallel optical interconnect (Neyer et al., 1999). Fibers made from PMMA have minimum loss around $\lambda = 570$ nm, and α increases rapidly for the red and near-infrared wavelength ranges; minimum losses are in the range of 70 dB/km, about a factor 10^3 higher than glass. Perfluorinated amorphous polymer materials are also used, with a wavelength range extending into the near infrared, with reduced optical attenuation, on the order of 50 dB/km.

Polymer-based fibers feature large core diameters (typically 85 to 1000 μm) and $NA \approx 0.5$, and are thus easy to couple into an optical system; a typical structure is a PMMA core ($n_g = 1.49$) surrounded by a cladding of fluorinated acrylate ($n_g = 1.41$).

Standard fiber diameters (125 μm) are available, although larger diameters, including 1, 2 and 3 mm, are popular; cladding thicknesses are generally only several tens of micrometers. Although they are quite robust, the operating temperature range is usually limited to below 85 – 105°C.

As we saw in Figure 3.9 of Section 3.3.1, the intrinsic absorption of SiO_2 increases rapidly for wavelengths longer than about 1.6 μm . As a result, silica-based glass is unsuitable as a fiber material for mid- to far-infrared wavelengths and alternative glass-like material systems have been developed for use in these wavelength regimes (Sanghera and Aggarwal, 1988). Fluorozirconate glasses, such as ZrF_4 or BaF_2 , are transparent in the wavelength range $400 \text{ nm} \leq \lambda \leq 5 \mu\text{m}$ with losses below 1 dB/km demonstrated in the laboratory. Alternatively, the chalcogenide glasses, the sulphur or selenium-containing materials to which we were introduced in Section 3.4.1 as important for infrared optics, are also used as optical fiber material. With low absorption in the range $3.3 \mu\text{m} \leq \lambda \leq 11 \mu\text{m}$, chalcogenide fibers have been shown in the laboratory to have losses below 0.7 dB/km and are important for long-wavelength applications.

9.6.3 Field distributions

The electric field, and thus intensity, distributions in an optical fiber may be found using the same analytical approach as we employed for the waveguide in Section 9.2. However, due to the cylindrical structure of the fiber, we now need to solve the wave equation in cylindrical coordinates; we thus again take Equation 9.30, but in this case have

$$\nabla^2 = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}. \quad (9.77)$$

The z coordinate corresponds to the axial direction along the fiber (the propagation direction), along with r , the radial, and ϕ , the angular, coordinates.

Field solutions

We take the same approach as for the waveguide: the wave equation is solved in the core and cladding regions, propagating and decaying solutions are determined and thus the field in the fiber may be determined. Unfortunately the ensuing analysis is relatively involved, and has been done comprehensively in, for example, (Iizuka, 2002b, Chapter 11.2); we quote the essential results of this treatment here and refer the interested student to the literature.

Solving the wave equation in the fiber core, we obtain a z -directed propagating solution of the form

$$E_z|_{\text{core}} = E_{g0} J_n(k_g r) \cos(n\phi) e^{j\beta z} \quad (9.78)$$

where E_{g0} is a constant, J_n represents an n^{th} order Bessel function of the first kind, and the parameter k_g [m^{-1}] is defined as

$$k_g = \sqrt{n_g^2 k_0^2 - \beta^2} \quad (9.79)$$

for \mathbf{z} -directed propagation constant β , as we defined for the slab waveguide; n is an integer.

In the cladding, a decaying solution is obtained, for which the \mathbf{z} -directed field takes the form

$$E_z|_{cladding} = E_{c0} K_n(k_c r) \cos(n\phi) e^{j\beta z} \quad (9.80)$$

where K_n is an n^{th} order Bessel function of the second kind and

$$k_c = \sqrt{\beta^2 - n^2 k_0^2}. \quad (9.81)$$

The variation of these field intensities ($\propto E_z^2$) as a function of radius is shown in Figure 9.34(a). The core electric field is oscillatory, with a first minimum at $k_g r = 2.404$. For comparison, a Gaussian decay ($\propto \exp(-r^2/W^2)$) is superimposed, showing that a Gaussian approximation is often adequate for the core electric field distribution, save that the field is slightly overestimated as the first minimum is approached. The field intensity in the cladding, shown by the dashed line, decays very rapidly with radius.

Modes

Using Maxwell's equations, the \mathbf{r} and ϕ directed electric fields, E_r and E_ϕ , respectively, may also be determined. In contrast to waveguides, most fiber modes have field components with all three components (E_r , E_ϕ , and E_z) such that the total field distribution for any given mode can be relatively complicated to visualize. The supported modes in a fiber are again determined using a procedure analogous to that for waveguides: by applying the appropriate boundary conditions at the core/cladding interface, a characteristic equation is derived, and the cutoff conditions for higher order modes determined. The lowest order mode is never cutoff.

We can briefly consider some special cases. For $n = 0$, the axial component of the electric field is zero, leaving only E_r and E_ϕ ; this situation corresponds to TE polarization since the electric field is always transverse to the propagation (\mathbf{z}) direction. On the other hand, for $n \neq 0$, the \mathbf{z} components of the fields are non-zero, and hybrid modes, denoted EH or HE, result. The lowest order mode, never cutoff and the desired mode for a single-mode fiber, is the HE_{11} mode, and is the only mode supported if the condition

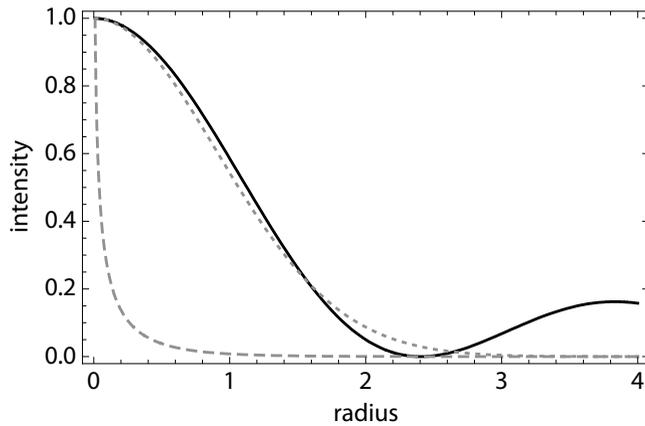
$$k_0 r_g \sqrt{n_g^2 - n_c^2} \leq 2.404, \quad (9.82)$$

for core radius r_g [m], is satisfied. The fields in the fiber core are then given by

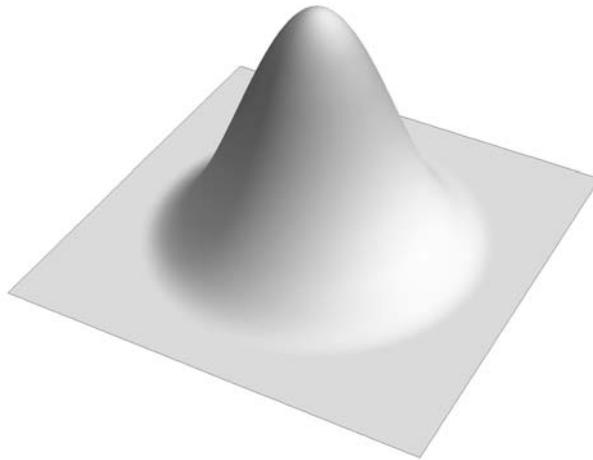
$$E_r = E_{g0} J_0(k_{g11} r) \cos(n\phi) \mathbf{u}_r \quad (9.83)$$

$$E_\phi = -E_{g0} J_0(k_{g11} r) \sin(n\phi) \mathbf{u}_\phi, \quad (9.84)$$

in which the vector directions are given by the unit vectors explicitly and k_{g11} is the propagation constant derived from the characteristic equation for the HE_{11} mode. The corresponding intensity distribution is shown in Figure 9.34(b), showing the typical mono-mode intensity distribution which results if Equation 9.82 is satisfied.



(a) Variation of the fiber field intensities with radius based on the functions $J_\nu^2(r)$ black and $K_\nu^2(r)$ dashed gray; for comparison, the Gaussian distribution is also plotted dotted gray.



(b) Intensity distribution of the lowest order fiber mode

Figure 9.34 Field intensities in a circularly symmetric single-mode fiber

9.6.4 Attenuation

When discussing the various types of fiber materials above, we compared these by the minimum fiber loss attained. Since optical fiber is typically used for transmission of optical signals over longer distances, meters to dozens of kilometers⁵, optical attenuation is an important issue. For comparison, the integrated waveguides we dis-

⁵ The longest single stretch of fiber in an undersea optical telecommunications cable is about 75 km, after which a repeater amplifies the signal and sends it down the next stretch.

cussed in the first part of this chapter have losses down to 0.1 dB/cm, corresponding to 10^4 dB/km; optical fiber, with optimal values of 0.2 dB/km, is about 10^5 times better.

Absorption

The material attenuation mechanisms we discussed in Section 3.3.1 play the dominant role in defining fiber loss; the absorption plot of Figure 3.9 essentially defines the attenuation in SiO₂-based fiber. The minimum of loss is at a wavelength of about $\lambda = 1.55$ μm and increases for shorter wavelengths due to electronic transitions and Rayleigh scattering, and for longer wavelengths due to lattice vibrations and impurities. Of particular relevance for SiO₂ are the silicon-oxygen bonds, with a resonance at $\lambda = 9$ μm , and those due to the dopants of the waveguide core, boron-oxygen at $\lambda = 7.3$ μm and germanium-oxygen at $\lambda = 11$ μm .

The most significant factor in defining fiber attenuation is probably the concentration of OH⁻ molecules, due to water. As we saw in Chapter 3, water has a fundamental absorption peak at $\lambda = 2.73$ μm with higher harmonics leading to absorption peaks at 0.88, 0.95, 1.13, 1.24 and 1.39 μm . Whereas the peak at 880 nm leads to an excess absorption of 0.1 dB/km/ppmOH⁻, the one at 1.39 μm generates 65 dB/km/ppmOH⁻. To achieve losses significantly below 1 dB/km, very low water concentration is thus essential in optical fiber: typically a level below about 10 ppb is required.

Scattering

Despite the fact that fibers are manufactured to very high tolerances and have high uniformity, scattering effects still play a role in defining the total optical attenuation. Scattering from uneven surfaces, such as that which takes place between waveguide layers or at etched channel boundaries, is negligible in fibers, but Rayleigh scattering, due to slight inhomogeneities in the material, is relevant and results in the λ^{-4} increase in absorption for shorter wavelengths. Rayleigh scattering gives rise to the fiber “glow” which may be observed at visible wavelengths and contributes about 1 dB/km of excess loss at $\lambda = 1$ μm .

A final loss source that can be an issue in the laboratory or in a compact optical microsystem using fibers for interconnect is that of bending loss. As with waveguides, the curvature radius of a fiber should be as large as possible, and the mechanical behavior of a fiber (i.e., bend it too much and it will break) usually implies that radii less than several tens of millimeters are unlikely. On the other hand, micro-bends, which result if a fiber segment is pressed against a sharp corner or obstacle, usually unintentionally, can give rise to scattering or radiation from the “kink” which may result. This latter situation can occur if bare fibers are configured in a Gordian knot on a messy optical table.

9.6.5 Dispersion

The effect of dispersion in an optical material was extensively discussed in Section 3.2. We saw that material dispersion results in wavelength-dependent propagation through

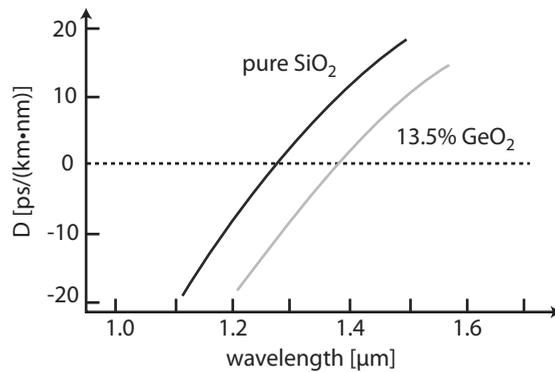


Figure 9.35 Dispersion of pure *black* and Ge-doped *gray* SiO₂ in the NIR wavelength range.

a material, and in particular that wave velocity varies with frequency. We also saw in Section 9.2.6 that modal propagation in a waveguide results in different modes having different propagation velocities, giving rise to waveguide dispersion. These forms of dispersion also play a role in optical propagation through a fiber and, since the overall propagation distances can be extremely long, they often dominate performance limitations, particularly at high bandwidths.

Relevance

The dispersion parameter D [ps/(km · nm)] typically summarizes all dispersion effects in a guided wave structure. The dimensions of D define the difference in propagation time (seconds) resulting from propagation through in a given length of fiber (kilometer) per unit spectral width of the optical signal (nanometers). Thus a signal with a large spectral width (say, an LED) is subject to strong dispersion: the different frequency components move at different speeds, thus pulling the optical pulse apart. Even for narrow band source, such as a DFB laser, if the optical signal is rapidly modulated, say at GHz frequencies, a spectrum of frequencies is generated and these are also subject to dispersion. As a result, over longer propagation distances, short optical pulses are increasingly spread out, resulting in a slow degradation of digital signals and limiting the bandwidth (highest modulation frequency) of the fiber system.

Depending on the fiber structure, D can be positive or negative. For $D > 0$, longer wavelengths travel faster than shorter ones, and for $D < 0$, shorter wavelengths travel faster than longer ones. As we shall see presently, different dispersion components can be combined to reduce D close to zero.

Material dispersion

Material dispersion was extensively considered in Section 3.2 and the relevant behavior of refractive index as a function of wavelength for SiO₂, the most popular long-haul fiber material, is seen in Figure 3.5. From this characteristic, the material dispersion

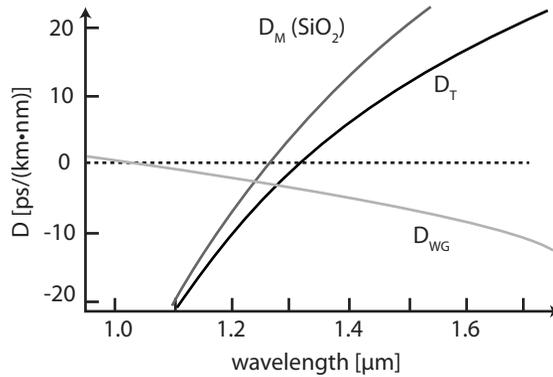


Figure 9.36 Total dispersion of an optical fiber D_T is the sum of the material dispersion D_M and the waveguide dispersion D_{WG} .

parameter D_M may be calculated from

$$D_M = -\frac{\lambda}{c} \left(\frac{d^2 n}{d\lambda^2} \right). \quad (9.85)$$

Applying this expression to the SiO_2 dispersion characteristic, we obtain the dispersion behavior seen in Figure 9.35, for pure and doped silica, the latter typical of the fiber core.

Waveguide dispersion

The total effective dispersion in a fiber, however, is the sum of that due to the material (D_M) and that due to the waveguide itself. We saw that not only do different modes have different velocities, but a single mode also has a wavelength-dependent velocity due to the physics of guided mode propagation. For multimode fibers, each mode has a different path length, so that multi-path dispersion is high; single mode fibers have no multi-path dispersion, but the single mode has a dispersion variation D_{WG} as shown in Figure 9.36. The total dispersion, D_T , is then the sum of these contributions.

As we see from the figure, typical dispersion values in the NIR wavelength range are about -20 to 20 ps/km nm and we also note that the characteristic goes through zero around $\lambda = 1.3 \mu\text{m}$. For this reason, the first long-haul fiber telecommunications systems were designed to operate around this wavelength. As an example for typical industrial dispersion values, the ITU-T G.652 fiber standard⁶ (for single-mode fiber optimized for use at $1.3 \mu\text{m}$) has zero dispersion at 1310 nm, a residual value of $D_T \leq 3.5 \text{ ps/km} \cdot \text{nm}$ in the wavelength range $1285 \text{ nm} \leq \lambda \leq 1330 \text{ nm}$ and a value of $D_T \leq 17 \text{ ps/km} \cdot \text{nm}$ at 1550 nm.

⁶ The ITU-T is the International Telecommunications Union, a body which sets the international standards for telecommunication systems.

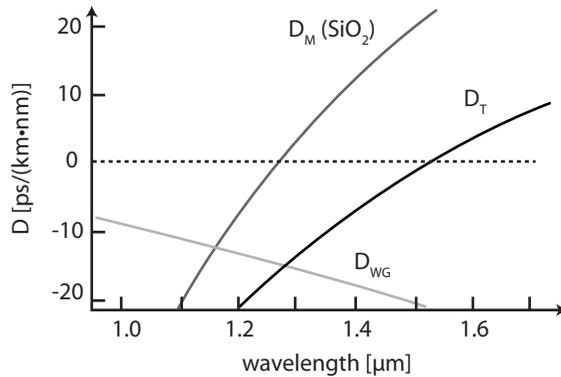
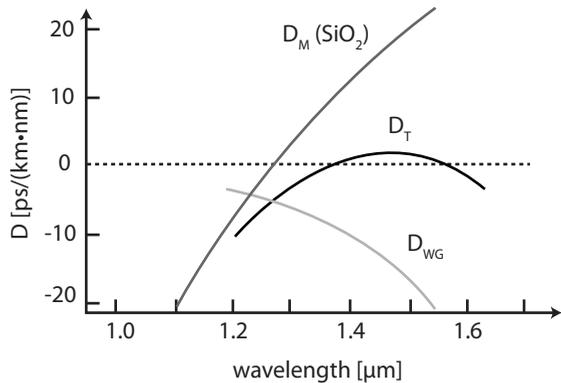
(a) Dispersion-shifted curve, with $D = 0$ at $\lambda = 1.55 \mu\text{m}$.(b) Dispersion-flattened curve with D close to 0 at $\lambda = 1.55 \mu\text{m}$.

Figure 9.37 Dispersion-shifted and dispersion-flattened fiber moves the point of minimum dispersion to $1.55 \mu\text{m}$ and flattens the dispersion curve to be minimal in this wavelength range.

Dispersion shifting

As we recall from the attenuation discussion of Section 9.6.4, the point of minimum absorption for a silica fiber is at about $\lambda = 1.55 \mu\text{m}$, so operation at $1.3 \mu\text{m}$ is advantageous from point of view of low dispersion, but subject to higher losses than for operation at $1.55 \mu\text{m}$. For this reason, novel radial refractive index distributions for the fiber core have been developed which give rise to a large negative waveguide dispersion values, also for longer wavelengths. The result is dispersion-shifted fiber with the characteristic shown in Figure 9.37(a); the point $D = 0$ has been red-shifted to $\lambda = 1.55 \mu\text{m}$. Using dispersion-shifted fiber, modern long-haul fiber systems are thus ultimately limited in transmission by attenuation, not dispersion.

Further improvement in dispersion characteristics can be attained by flattening the dispersion curve through appropriate core design, yielding dispersion-flattened fiber with the characteristic shown in Figure 9.37(b). Particularly for applications in which a spread of wavelengths around $\lambda = 1.55 \mu\text{m}$ is used, such as wavelength-

division multiplexing (WDM), a broad wavelength region with minimum dispersion is desirable. Dispersion-flattened fiber is usually optimized for the wavelength range $1.25 \leq \lambda \leq 1.65 \mu\text{m}$.

9.7 Waveguide micro-optics: case studies

Waveguides form the basis for the field of *integrated optics*. As we mentioned in the introduction to this chapter, integrated optics, analogously to integrated electronics, uses guided-wave structures fabricated on a substrate to allow the design of integrated optical circuits, based on optical functionality. A wide variety of active optical devices, especially lasers, detectors, modulators and switches also employ guided-wave structures, and thus represent integrated optical components, interconnected by waveguides (Murphy, 1999). Integrated optics is usually considered as distinct from micro-optics (narrow definitions of the latter assume free-space propagation) but advanced optical microsystems employ and combine both into complex and highly functional assemblies.

Numerous designations for the resulting concepts have been developed, including *photonics* for laser-based systems (not necessarily integrated); *photonic integrated circuits* (PICs), which integrate waveguide circuits and active optical components on a single substrate (Koch and Koren, 1991) and *opto-electronic integrated circuits* (OEICs) in which guided wave, photonic and electronic components are all integrated onto a single multi-functional chip (Wada et al., 1986). The distinctions between these concepts are sometimes relatively fluid and not terribly relevant.

Two application areas have provided the primary impetus for the development of integrated optics technology: optical telecommunications and optical sensors. We consider a few examples from both of these disciplines to show how the simple waveguide plays a significant role in widely-deployed micro-optical systems.

9.7.1 Optical data transmission

Most telecommunications data is sent optically: from long-haul data streams under the Pacific ocean using fiber networks to optical data networks in automobiles to short-range high-capacity optical links within and between computers, the majority of bits and bytes sent about the world are today optical ones. Integrated optical components and circuits have long played and continue to play an important role in enabling optical communication networks (Alferness, 1981). Whereas the data transmission itself employs glass or plastic optical fiber, and is thus not really a microsystem, generation, detection and switching of digital optical data falls squarely into the most important applications regime for integrated optics.

A rudimentary optical communications system has five components, as shown schematically in Figure 9.38: an optical source, usually a laser or an LED; a high-speed modulator with which the optical field is modulated digitally; one or more switches which guide the signal to the desired path; a detector, which converts the optical signal back into an electrical one; and high-speed electronics to drive all this optics (Razavi,

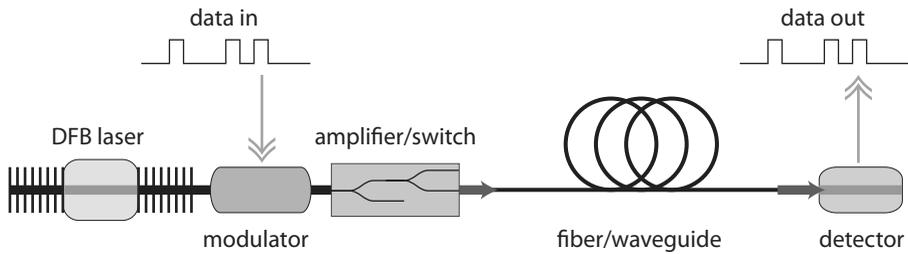


Figure 9.38 Schematic representation of an optical telecommunications system. From left to right: the light from a narrow-bandwidth, often a DFB, laser, is modulated digitally by a high-speed optical modulator; a switch directs the digitally-modulated signal to a transmission waveguide or fiber; and finally a photodetector converts the signal back into an electrical one.

2002). Many of the components in an optical telecommunications system are guided-wave structures, including, of course, the fiber; however, discrete micro-optical components still play an important role in these systems, particularly for switches, filters and attenuators, and polarization-sensitive devices (Pennings et al., 1996). We will look at the structure and functionality of the individual active components such as the laser, modulator and detector in more detail in Chapter 10.

Key to optical data communication is high modulation speed (a high digital bit rate) and a long transmission distance. Optical modulation speeds of 10 Gbit/s correspond to those of most systems currently deployed, as given by the optical carrier standard OC192 (Held, 2001)⁷; by time-multiplexing 64 signals, speeds of up to 640 Gbits/s have been demonstrated in the laboratory (Nakazawa et al., 1998). The current transmission record is held by Alcatel-Lucent Bell Labs⁸, who employed 155 lasers, each operating at 100 Gbit/s, over a fiber stretch of 7 000 km, to reach an aggregate speed of 100 Petabits/s·km in a demonstration in Villarceaux, France⁹.

Wavelength-division multiplexing

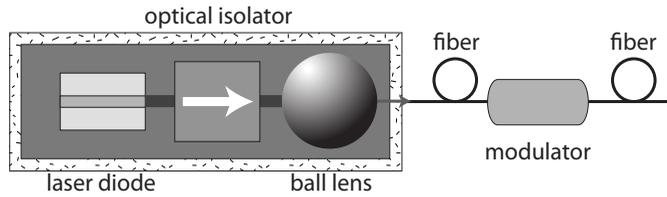
One of the most widely applied means to increase optical channel capacity is through wavelength-division multiplexing (WDM). Using laser sources of different closely-spaced wavelengths in the telecommunications wavelength band (around $\lambda = 1.55 \mu\text{m}$), each wavelength channel is modulated separately and all the signals are multiplexed onto a single fiber. On the receiver end, the different wavelengths are then demultiplexed and each directed onto a separate high-speed photodetector.

WDM systems can significantly increase fiber data capacity; in 2002, for example, the optical network vendor Alcatel demonstrated a transmission of 365 wavelengths, each modulated at 10 GBit/s, over a distance of 6 850 km on a single fiber, resulting in a total fiber bandwidth of 3.65 TBit/s. Many of the integrated optical devices and com-

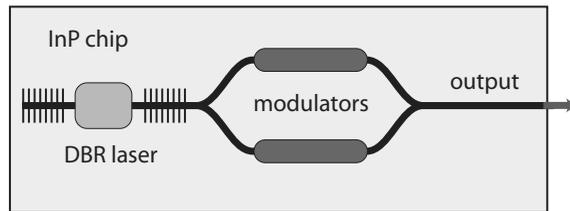
⁷ “OC” stands for “optical carrier”. Long-haul fiber systems are slowly being upgraded to the OC768 standard, 40 GBit/s.

⁸ This data stems from an Alcatel-Lucent Bell Labs press release dated September 28, 2009.

⁹ Is that a lot? To provide a reference for the amount of data being shoveled around the planet, in 2008 Google claimed to process about 20 PBytes (20 000 TBytes) of data per day (Dean and Ghemawat, 2008).



(a) A hybridly-assembled transmitter in which light from the laser is coupled through an optical isolator into a fiber through a ball lens and the signal digitally modulated by an external modulator.



(b) An integrated optical transmitter fabricated on a single InP chip; the emission from the DFB laser is modulated by a waveguide-based Mach-Zehnder interferometer modulator.

Figure 9.39 Schematic representations of two variations of optical transmitters.

ponents used in optical telecommunications are designed to realize WDM functionality in compact, high-performance optical chips.

Transmitters

The heart of a WDM-based optical communications system is the light source, an array of lasers emitting at precisely-defined wavelengths around $1.55\ \mu\text{m}$, whose emission is modulated at 10 to 40 GHz, amplified, multiplexed with emitters at other wavelengths and sent down an optical fiber. Single-wavelength optical transmitters are frequently assembled as shown schematically in Figure 9.39(a): the laser is coupled through an optical isolator into a fiber using, for example, a ball lens, and the signal modulated digitally by a subsequent fiber-based modulator; the different wavelengths from an array of such transmitters is then multiplexed onto a single fiber using waveguide couplers. One of the limitations of this approach, which is nevertheless employed extensively, is the optical loss due to the numerous interfaces between successive components, in addition to the assembly costs and possible reliability limitations.

Integrated optical technology allows integration of many of these functions onto a single chip, as is shown schematically in Figure 9.39(b). If an optically-active substrate material, such as InP or another III-V semiconductor, is used, the light source, typically a DFB or DBR laser, may be directly coupled into a waveguide-based Mach-Zehnder interferometer modulator, both on a single chip. The external digital modulation of the laser emission, achieved in hybridly assembled systems using LiNbO_3 modulators, can be replaced using on-chip Mach-Zehnder interferometer-based mod-

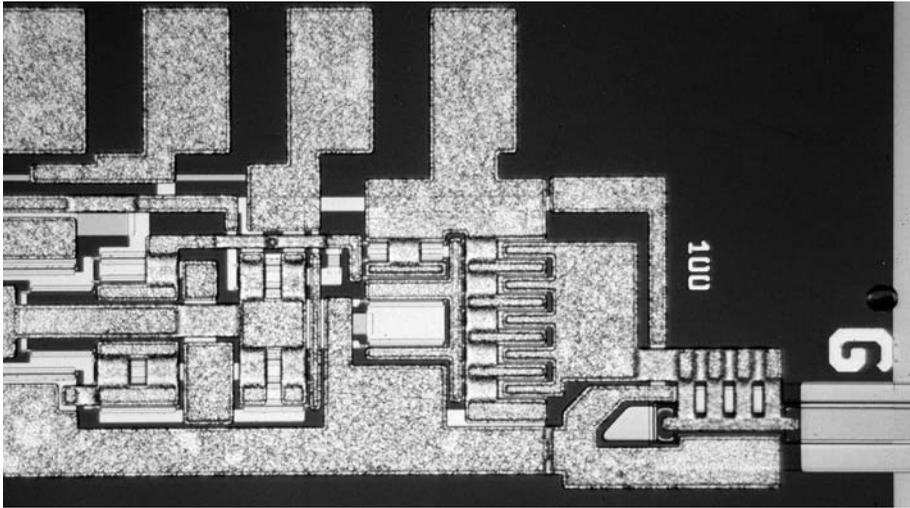


Figure 9.40 Photograph of an optical transmitter chip including a Fabry-Perot laser, at the lower right, and the associated driver circuitry. The Fabry-Perot cavity is formed by one etched and one cleaved facet, the latter formed in the open region below the sideways “G” at the right. The contact pads are 100 μm wide. Photo courtesy of the Fraunhofer Institute for Applied Solid State Physics.

ulators, demonstrated with 18 GHz bandwidth (Barton et al., 2003). An example of a laser integrated with high-speed driver circuitry is shown in Figure 9.40, in which a Fabry-Perot laser, with one etched and one cleaved facet, is directly integrated with the III-V based transmitter circuitry.

As a further example, DBR lasers have been monolithically integrated with semiconductor optical amplifiers (SOA) using an etched buried waveguide on InP for interconnect (Mason et al., 2000). These monolithically integrated waveguide structures require a means to generate transparent waveguides using the same layer structure as the lasers (which emit and absorb at roughly the same wavelength); quantum well intermixing techniques are usually employed to blue-shift the absorption edge of the waveguide layers so that these become transparent at the laser emission wavelength (Hofstetter et al., 1998; Skogen et al., 2005).

Fabrication techniques have been extended further such that single-chip transmitters with up to 40 optical channels, each with its own wavelength and each modulated at 40 GBit/s, on a single InP-based substrate have been demonstrated (Nagarajan et al., 2005; Welch et al., 2006); integrating 240 optical components, the transmitter consists of thermally wavelength-tunable lasers, each modulated using an electro-absorption modulator, the intensity adjusted by a variable optical attenuator (VOA) and monitored by an optical power monitor. The emission from all lasers, with a wavelength spacing of 50 GHz, is combined into a single output using arrayed waveguide grating (AWG) multiplexer (which we will see again below) and this functionality requires only one (albeit complex) optical chip.

Planar lightwave circuits

Monolithic III-V semiconductor-based optical transmitters provide high functionality, albeit at the price of complex technology and the associated high cost. A lucrative alternative is found in the use of silicon technology for those optical functions for which the material is suitable; the result, silicon photonics, at present includes most optical functionality except efficient laser emission (Zimmermann, 2000; Reed and Knights, 2004; Lipson, 2009). Silica-based planar lightwave circuits (PLC) are usually based on silicon-on-insulator (SOI) structures, consisting of an epitaxial silicon layer grown on a buried layer of SiO_2 , which is in turn deposited on a bulk silicon substrate. This layer sequence forms the basis for strongly-confining waveguides (the guide layer, Si, with $n_g = 3.45$ and the cladding, SiO_2 , with $n_c = 1.45$) which have reasonably low loss in the wavelength range $1.1 \leq \lambda \leq 7.0 \mu\text{m}$. It was already realized in the 1980s that SOI could form the basis for an efficient waveguide technology in the 1.3 and 1.55 μm wavelength ranges (Soref and Lorenzo, 1986), using silicon fabrication technology for patterning.

Initially, simple passive waveguide structures, such as Y- and proximity-couplers, were fabricated using this approach; as the technology progressed, Mach-Zehnder interferometers, ring resonators, and $N \times N$ couplers were demonstrated, and now the PLC is the technology of choice for MMIs, AWGs and active integrated optical components (Kawachi, 1996). Relatively complex optical functions, such as wavelength-selective switches, optical equalizers, VOAs, resonator filters and reconfigurable optical add-drop multiplexers have been realized on single silicon chips (Doerr and Okamoto, 2006; Jalali and Fathpour, 2006). Silicon substrates are also employed as the basis for a variety of optical switches, including silicon nanophotonic ring resonators (Vlasov et al., 2008) and SOI-based micromachined waveguide structures, such as free-standing cantilevers, which have been used as MEMS-based optical switches (Ollier, 2002).

Modulation, one of the most important active functions for an integrated optical circuit, is usually accomplished in silicon using the thermo-optic effect (dn/dT); the effect is unfortunately relatively slow (compared to electro-optic effects in III-V semiconductors) and requires high powers. Nevertheless, thermo-optic switches form the basis for interferometer modulators, as well as variable couplers and switches; very compact 1×1 , 1×2 , and 1×4 switches (with sizes less than $190 \times 75 \mu\text{m}^2$) have been fabricated using photonic wire waveguides (Chu et al., 2005). In addition, since birefringence is a function of waveguide width, efficient polarization beam splitters have been demonstrated using SOI waveguides (Liang and Tsang, 2005). Finally, the PLC has also proven to be useful as a platform for hybrid integration of silicon and non-silicon components, an approach referred to as the “silicon optical bench”, which we will examine in more detail in Chapter 11.

Hybrid systems

Silicon photonics has also given rise to advanced systems using silicon processing technology for the fabrication of transceivers in which the lasers are hybridly attached to the chip using flip-chip bonding or other techniques. The silicon is in this case not only used for passive waveguide coupling structures but also for the fabrication of active

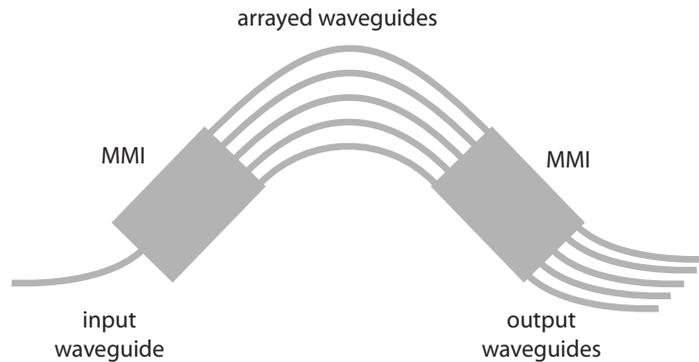


Figure 9.41 Schematic representation of an arrayed waveguide grating (AWG) demultiplexer. The multimode interference coupler (MMI) at the input generates a wavelength-dependent input field among the arrayed waveguides resulting in a wavelength-dependent interference pattern in the output MMI and thus one wavelength at each output waveguide. For M input waveguides (only 1 is shown) and N output waveguides, the AWG may be used as an $M \times N$ coupler.

components. As an example, a 16 channel, 160 GBit/s CMOS-based transceiver has been configured as a Si/III-V hybrid; the $3.25 \times 5.25 \text{ mm}^2$ chip uses InP-based laser diodes and a 40 GBit/s silicon Mach-Zehnder modulator (Analui et al., 2006). Similar technology has also been used to include filters, detectors and amplifiers on the transceiver substrate (Jalali and Fathpour, 2006).

De-multiplexers and receivers

Optical telecommunications systems based on WDM not only require an array of laser sources with precisely-defined wavelengths, but also a means to de-multiplex the numerous wavelengths emerging from an optical fiber into separate output channels. Silicon-based photonics also plays a considerable role in this function. The arrayed waveguide grating, to which we were introduced in Section 9.5.4 and which is shown schematically in Figure 9.41, is a combination passive slab and channel waveguide structure with one or more inputs and N outputs for N wavelengths. As we saw above, the multimode interference coupler spreads the input field among the arrayed waveguides as a function of wavelength, and these give rise to an interference pattern in the output MMI which then results in a single wavelength at each output waveguide (Doerr and Okamoto, 2006).

The AWG has been realized using both silicon and III-V waveguide technologies. As an example of the latter, a 64 channel AWG with 232 arrayed waveguides and a spectral resolution of 50 GHz has been realized on a $3.6 \times 7.0 \text{ mm}^2$ InP chip, with less than -20 dB crosstalk between channels (Yoshikuni, 2002). Photodetectors, SOAs and electro-absorption modulators could be integrated onto the same substrate with this technology.

Finally, advanced electronics may be integrated with waveguide-based photodetectors to realize high-bandwidth optical receivers. InGaAs-based waveguide photodiodes, for example, have been flip-chip bonded onto GaAs-based HEMT (high electron mobil-

ity transistor) amplifiers, resulting in hybrid receiver assemblies with a bandwidth of 65 GHz (Leich et al., 2002).

Short-range communication

Of the estimated roughly one billion kilometers of fiber deployed around the world today, an increasing fraction is used in short-range data communications, particularly for distances below 300 m. More and more of this infrastructure is based on plastic optical fiber, primarily for reasons of cost. POF remains flexible even with a diameter of 1 mm or more, sizes at which glass optical fiber would break due to its brittle nature and the large fiber cores available for POF ease the coupling restraints, thereby reducing the assembly costs of plastic-based fiber optical systems. Most POF systems use PMMA-based fiber, with absorption minima around 520 and 650 nm, and loss values of 80 dB/km and 140 dB/km, respectively. Light sources are generally 650 nm LEDs or occasionally green or blue sources, whereby the latter require power supplies with voltages higher than the standard 3.3 V.

Due primarily to the multimode behavior of large-core POF, the modulation bandwidths for these systems are considerably smaller than those found in glass-based long-haul optical networks. Typical short-range bandwidths are on the order of 200 MBit/s for distances up to 50 m; optical Ethernet standards typically require 100 GBit/s. Two rapidly growing application areas are the fiber-to-the-home (FTTH) and automotive sectors. For FTTH, it is expected that data rates of 0.1 to 1.0 GBit/s can be reached, the latter with graded-index POF to reduce multi-path dispersion; very low cost transceiver and optical switch components will be necessary for this market to succeed, such that a variety of polymer and plastics technologies are being employed. For automotive applications, POF has been used since 2001 (when it was featured in a BMW 7-series automobile) for the “infotainment” systems which have grown rapidly in high-end models. Using an optical ring, navigation, entertainment and communications data is transferred about the vehicle; the POF-based network is light-weight and immune to electromagnetic interference, both essential issues in the automotive environment.

9.7.2 Guided-wave sensors

Whereas telecommunications has likely played the leading role in stimulating the development and application of guided-wave technologies, an area which has benefitted greatly from these advances is that of integrated optical sensors. Waveguide-based optical sensors are attractive for a wide spectrum of applications since they are compact, require little or no alignment, may often be extended into multi-sensor arrays and, since they are optical, they neither suffer from electromagnetic interference nor do they generate any.

The sensor field in most any technology (electrical, mechanical, or optical) is generally very fragmented, since virtually every sensor application has its unique demands and, usually, unique solution. As a result, a considerable variety of concepts for integrated optical sensors has been proposed and demonstrated (Zappe, 1999, and references therein). We will look at a small sampler here.

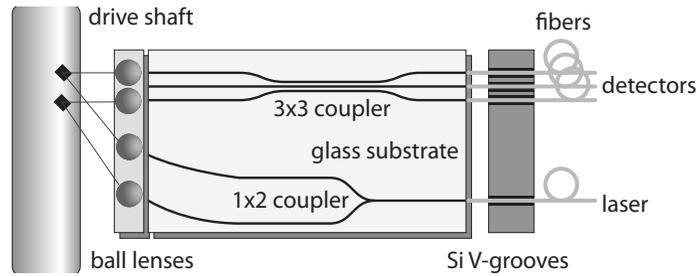


Figure 9.42 An integrated-optical non-contact torque sensor. Light from an external laser is split into two beams by the 1×2 coupler on the glass waveguide chip, which are reflected from two reflectors embedded on the rotating shaft; these move in relative position as torque is applied, resulting in a phase shift between the three output beams.

Materials and technologies

Integrated optical sensors occasionally use the same fabrication technologies as telecommunications implementations, but the emphasis on ultra-low cost for the former implies that economically viable technologies are usually preferred. Silicon (Valette, 1988), III-V semiconductor (Zappe et al., 1994a,b), polymer (Gale et al., 1995) and glass technologies have been used. Active semiconductors and certain polymers allow the integration of active optical functions, such as light emission or modulation, whereas glass, for example, is for purely passive applications.

The structure of optical sensors can range from simple waveguides whose transmission properties vary with an analyte of interest to complex interferometers with extensive signal processing on-chip. The most popular optical structures tend to be interferometers, integrated optical implementations of the structures to which we were introduced in Section 5.2. Mach-Zehnder, Michelson and Young interferometers are often used, as are resonator structures such as the Fabry-Perot cavity or the ring resonator structure we saw briefly in Section 9.5.4. The ease with which these may be fabricated at small size using waveguide technology implies that many optical sensing techniques rely on sensor quantities which may be translated into an optical phase shift, and are thus measurable by an interferometer.

Mechanical sensors

Optical measurement of movement, vibration or deformation is well-established, and integrated optical measurement techniques for mechanical sensing have been demonstrated with high sensitivity. A simple example is shown in Figure 9.42, in which a glass-based integrated optical chip is used to measure torque on a rotating drive shaft (Ebi et al., 1999). Light from a 780 nm semiconductor laser is split into two beams by the 1×2 coupler and focused obliquely onto two retro-reflectors embedded in the drive shaft using ball lenses. The reflected light is coupled into the 3×3 coupler, resulting in three output signals, phase shifted by 120° with respect to one another. The entire system was hybridly assembled using gluing techniques.

When torque is applied to the drive shaft, the relative positions of the reflectors change as the shaft deforms, giving rise to a change in the phase shifts between the three output signals, from which the torsional strain can be determined. The signal output allowed a torque resolution of 2 Nm for a 50 mm diameter shaft, using a contact-free measurement. This example shows how a passive waveguide-based system can employ sensitive interference effects in a setup which is compact and robust.

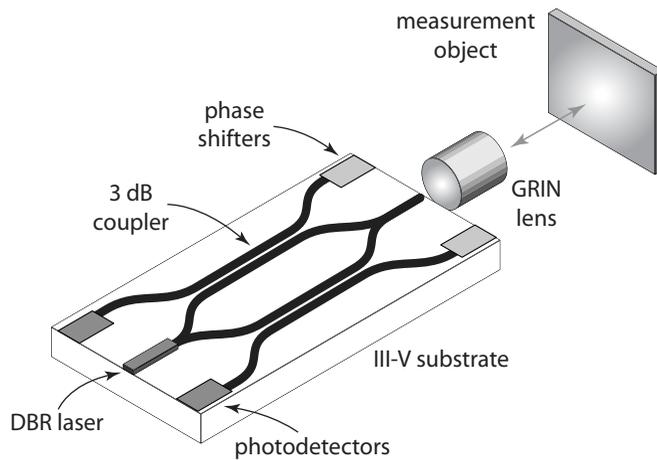
Numerous other physical parameters have been measured using waveguide optics, including vibration (Ollier et al., 1999), pressure (Pelletier et al., 2007), acceleration (Wu and Frankena, 1993; Plaza et al., 2004; Wu et al., 2004), and rotation (Suzuki et al., 2000). The latter, for example, uses the same Sagnac interferometer as the FOG, and thus represents an integrated optical gyroscope. Based on a silicon PLC, the structure is compact but performance does not approach the fiber (or micro-mechanical) alternatives.

Non-contact optical displacement sensing is a standard, high-precision measurement technique with a wide variety of free-space optical implementations and has been demonstrated integrated optically using various technologies (Valette, 1988; Poffo et al., 2005). An integrated optical approach using III-V semiconductors is shown in Figure 9.43 (Hofstetter et al., 1995, 1996, 1997). Based on GaAs, the pair of Michelson interferometers includes a single DBR laser as a near-IR light source, phase shifters on two reference reflectors and on-chip photodetectors, as seen in Figure 9.43(a). Light is coupled off the chip and focussed by a GRIN lens onto a moving object, such as a mirror; the reflected light then returns to the interferometer and gives rise to two phase-shifted output signals at the photodetectors.

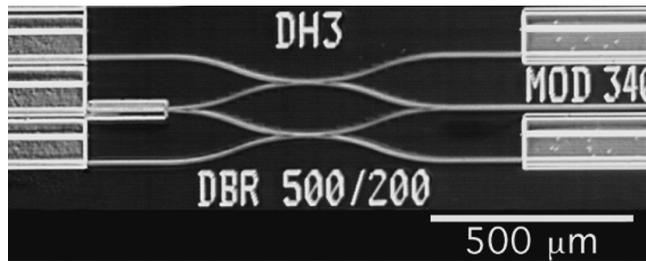
The monolithically-integrated microsystem was fabricated using a GaAs substrate and the quantum well intermixing technology mentioned above for achieving transparency in the waveguide regions; as seen in Figure 9.43(b), the entire dual-Michelson system has dimensions of $3.5 \times 0.25 \text{ mm}^2$. In operation, the movement of the external object gives rise to two detector signals, each varying sinusoidally with object displacement distance d as $\cos(4\pi d/\lambda)$, for laser wavelength $\lambda = 822 \text{ nm}$ (see Section 5.2.2). The phase shifters are configured so that the output signals are phase shifted by 90° , such that a change in movement direction is unambiguously detected, even if it occurs at the maximum or minimum of the interference signal. A measurement resolution of 20 nm at a maximum chip/object distance of 0.45 m was demonstrated using this compact, integrated configuration.

Bio/chemical sensors

The largest variety of integrated optical sensors is likely to be found in the area of biological and chemical detection (Lambeck, 2006). These sensors are usually based on a shift in the effective index of a waveguide in response to a biological or chemical analyte, often due to the adsorption of a compound onto the waveguide surface, thereby affecting the propagation of the evanescent field, or by absorption in the waveguide itself (Klein and Voges, 1994). As an alternative, fluorescence may be stimulated using appropriately sensitized fluorophores on a waveguide surface (Chang-Yen and Gale, 2003). Also of interest are surface plasmon resonance (SPR) sensors, which are often free-



(a) Schematic representation of the dual Michelson interferometer: light from the DBR laser is split into two interferometers, defined by the proximity couplers, each of which has a unique phase shifter on the reference arm. Light is coupled off-chip onto the moving measurement object, directed back through the interferometers and is converted to an electrical signal in the two photodetectors.

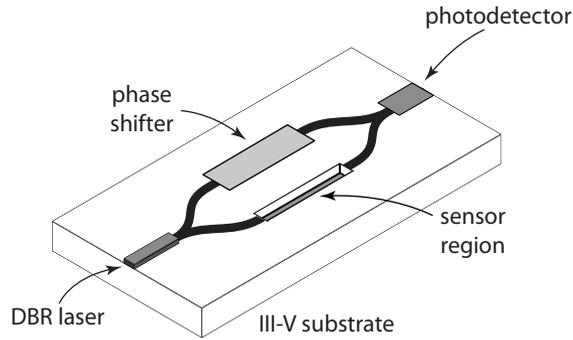


(b) The GaAs-based interferometer chip, with dimensions $3.5 \times 0.25 \text{ mm}^2$. The DBR laser is at the left, the two phase shifters are at the right edge of the chip and the waveguide for optical emission off-chip is below the label "MOD 340".

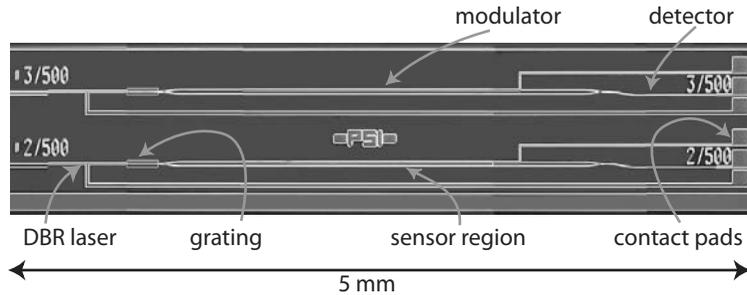
Figure 9.43 A monolithically-integrated Michelson interferometer for optical displacement measurement.

space optical systems using grating or prism coupling, but have also been demonstrated to have suitably high sensitivity in waveguide configurations (Homola et al., 1999); we will examine SPR systems in more detail in Chapter 14.

Evanescent wave sensors are generally the most popular integrated optical configurations for chemical sensing, since a change in the refractive index of the cladding layer, or deposition of molecules on the core itself, gives rise to a change in the effective waveguide index. As a result, the propagating field is subject to a phase shift with respect to an identical waveguide unaffected by the analyte to be sensed, and phase shifts are measured with high sensitivity using interferometers. Since entire waveguide-based



(a) Refractive index or biological sensing using a monolithically-integrated Mach-Zehnder uses an on-chip DBR laser with integrated detector and modulator; in the sensor region, the waveguide core is exposed to the environment.



(b) Photo of two Mach-Zehnder interferometers on a GaAs chip, each with an area of $0.35 \times 5.0 \text{ mm}^2$; the spacing between the two arms of the interferometers is not discernible at this resolution.

Figure 9.44 A monolithically-integrated Mach-Zehnder interferometer for optical refractive index and biological sensing.

interferometers can be fabricated as easily as individual waveguides on an integrated optical chip, effective-index bio/chemical sensing has been extensively developed.

Numerous optical evaluation schemes have been demonstrated, including silicon-based ring resonators, used for protein sensing with a sensitivity of 0.1 nM (Ksendzov and Lin, 2005), or hybridly-assembled Young interferometers (Schmitt et al., 2007) using Ta_2O_5 on glass with a refractive index resolution of $4 \cdot 10^{-8}$ for a 5 mm long sensing region.

The Mach-Zehnder interferometer is also a popular configuration; one arm of the interferometer is exposed to the analyte of interest and the other isolated, so the phase difference between measurement and reference paths may be easily detected when the two interfere. A silicon-based implementation, using an external HeNe laser as a source, has been shown to have sensitivity of $\Delta n = 7 \cdot 10^{-6}$ for a 15 mm long sensor waveguide (Prieto et al., 2003). As an alternative, a monolithically integrated GaAs-based version (Maisenhölder et al., 1997) is presented in Figure 9.44. Using the same tech-

nology as the Michelson interferometer of Figure 9.43, this structure also includes an integrated DBR laser, detector and phase modulator in the reference arm of the interferometer. The sensor region has the core exposed to the environment, such that the evanescent field detects the appropriate chemical changes.

This arrangement has been used for refractive index and protein concentration measurements. Changes in refractive index of a fluid covering the 2 mm long sensor region could be measured with a resolution of $\Delta n = 1.3 \cdot 10^{-5}$ and adhesion of molecules onto a surface-immobilized protein layer could be determined down to levels of 3 pg/cm^2 . The phase modulator in the reference arm is used to set the operating region of the interferometer to the most sensitive regime, which is where the slope of the \cos^2 -type interference characteristic is maximum, far from the interference maxima or minima.

Fiber sensors

The widest deployment of guided-wave sensors is probably not of micro- or integrated-optical implementations but rather fiber-based systems. Fiber sensors have a long history and have been extensively developed, particularly for remote sensing applications where the long transmission distances through fiber are a distinct advantage (Culshaw and Dakin, 1997; López-Higuera, 2002). Whereas a variety of chemical and biological fiber-based sensors have been proposed, the most widespread application is through the use of fiber Bragg gratings (FBG) for distributed strain, temperature and pressure sensing (Rao, 1997). FBGs employ a UV-written Bragg grating formed in Ge-doped fiber during fabrication; these typically 20 cm long regions reflect at a wavelength, λ_{Bragg} , defined by their period Λ , $\lambda_{Bragg} = 2n_g\Lambda$, and Λ changes with applied strain, for example. Using a broad-band light source, a wavelength-selective detector and a FBG with various periods along the fiber length, these structures are used for distributed sensing and are now standard in monitoring the structural integrity of bridges, buildings and other feats of civil engineering.

Problems

1. For an 800 nm thick symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$, where we recall that SiO_2 has $n = 1.46$ and Si_3N_4 has $n = 2.05$, solve the characteristic equation graphically for the $m = 1$ TE mode at $\lambda_0 = 633 \text{ nm}$.
2. Using the same waveguide as in the previous example, perform a numerical solution, again for the $m = 1$ TE mode at $\lambda_0 = 633 \text{ nm}$.
3. Design a single-mode III-V semiconductor waveguide using GaAs ($n = 3.06$) and $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ ($n = 3.23$) for use at telecommunications wavelengths, $\lambda = 1550 \text{ nm}$. Sketch the structure and give all relevant dimensions. Determine reflection angle, N and β for operation in TE polarization.
4. Given a $1.5 \text{ }\mu\text{m}$ thick waveguide core made in a silicon oxynitride system, for which $n_g = 1.90$, and a symmetric oxynitride cladding with $n_c = 1.55$, how many modes are supported for $\lambda = 543 \text{ nm}$? Solve the waveguide for TM polarization and compare the values for θ , N and β for the $m = 0$ and highest order modes.

5. Plot the allowed number of modes as a function of waveguide core thickness in the range $0 \leq t \leq 2 \mu\text{m}$ for a symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide operated at $\lambda = 0.85 \mu\text{m}$.
6. Show that for a symmetric waveguide with weak confinement (i.e., $n_g \approx n_c$), the cutoff wavelength for TE modes with index $m > 1$ is given approximately as

$$\lambda_{0|cutoff} \approx \sqrt{\frac{8n_g t^2 \Delta n}{m^2}}$$

for $\Delta n = n_g - n_c$.

7. For an asymmetric waveguide with $n_s > n_c$, which interface defines the cutoff criterion?
8. Solve an asymmetric waveguide consisting of $0.9 \mu\text{m}$ polycarbonate on BK7 glass with air as the cap material, for operation with a green HeNe laser ($\lambda = 543 \mu\text{m}$) in the TE polarization. How many modes are supported, and what are θ , N and β for $m = 1$?
9. Calculate the penetration depths into the glass substrate and into the air for the waveguide and operating conditions of Problem 8.
10. Plot the electric field intensity for the $m = 1$ mode for the asymmetric waveguide and operating conditions of Problem 8 and verify that the penetration depths are as expected from the previous problem.
11. Calculate the confinement factor for the $m = 1$ mode for the waveguide and operating conditions of Problem 8.
12. Determine the Goos-Hänchen shift on the substrate and surface sides for the $m = 1$ mode for the waveguide and operating conditions of Problem 8.
13. Show that Equations 9.15, 9.54 and 9.55 are all identical.
14. Explain how a rib waveguide works: why is there lateral guiding? Tip: have a look at the values for k_{yg} which are found under the strip and those found at the sides of the strip.
15. Now explain how a strip-loaded waveguide works.
16. Derive the expression for the numerical aperture of a waveguide as presented in Example 9.8.
17. What is the maximum coupling angle relative to the facet normal for the symmetric $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguide in air?
18. Calculate and plot the variation in butt-coupling efficiency between a fiber ($n_g = 1.4515$) and a GaAs waveguide ($n_g = 3.73$) for an operating wavelength of 780 nm as a function of axial spacing between the two components in the range $0 \leq d \leq 2 \mu\text{m}$. Recall that a parasitic Fabry-Perot cavity is established between the surfaces and that Fresnel reflection exists at both boundaries.
19. Calculate NA and the maximum coupling angle (relative to the surface normal) of a cleaved fiber, with $n_g = 1.4515$ and $n_c = 1.4470$, in air and in water. For diffuse radiation, what portion of optical energy is coupled into the fiber core?
20. Given a 75 km stretch of single-mode fiber with $\alpha = 0.2 \text{ dB/km}$. What fraction of the optical power input at one end emerges at the other? How long would a typical III-V-based waveguide giving rise to the same decrease in signal strength be?

21. Estimate how many of times the total length of fiber installed world-wide could be wrapped around the equator. Estimate the volume of *core material* installed in the world-wide fiber network and the approximate dimensions of a cube which would contain all of it.
22. Design a waveguide sensor for use at 780 nm with high sensitivity to external changes in n_c , where the cap material is water. What are the optimal field conditions for such a waveguide sensor, and how might we realize it using typical materials? For the energetic student: determine the optimal waveguide core thickness for high sensitivity.

10 Active micro-optics

In the beginning there was nothing. God said, 'Let there be light!' And there was light. There was still nothing, but you could see it a whole lot better.¹

Ellen Degeneres

The micro-optical devices we considered in Chapters 6 through 9 were *passive* optical components: mirrors, lenses, diffractive structures and waveguides which interacted with light but, with the possible exception of actuated micromirrors, did not require any input of electrical power. *Active* optical components, in contrast, are those which rely on an interaction with (typically) electrical and optical fields to generate, absorb and modulate light. They form an important category of micro-optical structures, since most optical systems require at least a light source and a means for optical detection.

As with most of the other topics of this text, active optics is a field onto itself, or better yet, a union of numerous very extensively developed fields. We provide in this chapter an overview of the active optical components of highest relevance for micro-optics and optical microsystems, including laser diodes and LEDs, detectors, modulators, optical isolators and liquid crystal devices. The literature is of course extensive (Fukuda, 1999; Iizuka, 2002a,b; Liu, 2005; Solgaard, 2009) and we will provide further references to texts which go into greater depth in each of the following sections.

10.1 Physics of light emission

Optics without light is pretty boring. All optical systems require a source of photons to realize their functionality, and micro-optical systems have an array of suitably sized light sources at their disposal. These convert electrical into optical energy and represent an important class of active optical devices. The most important light emitters for micro-optics are the light-emitting diode (LED) and laser diode, since the small size and high efficiency of these components implies that they may easily be integrated into optical microsystems.

The physics of light-emitting diodes and lasers is one of the truly fascinating branches of theoretical optics, in the widest sense, and we provide a brief summary of the physical basics here to allow the student to understand the functionality of these devices. An in-

¹ Ellen Degeneres (b. 1958), American stand-up comedienne.

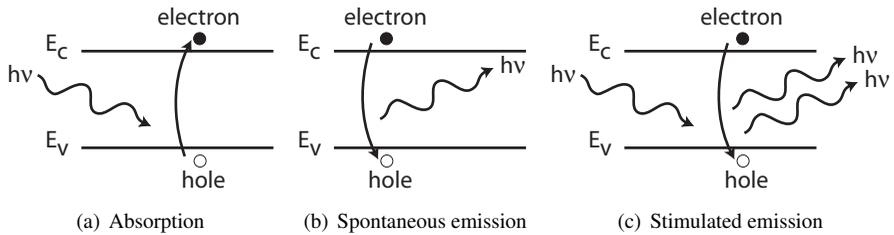


Figure 10.1 Possible electronic transitions in a two-level system, using the example of the conduction band (E_c) and valence band (E_v) of a semiconductor.

depth treatment can be found in one of numerous classic and didactically excellent texts (Siegman, 1986; Silfvast, 1996; Chow and Koch, 1999) which consider the subject in greater depth. We will first make a quick diversion into the physics of light emission and absorption, which forms the basis for understanding lasers and also detectors.

10.1.1 Basic optical transitions

Light emission and absorption generally rely on the transitions of electrons between different energy levels in an appropriate optical medium. Most LEDs and laser diodes are based on semiconductors, and we will implicitly restrict ourselves to this family of optical materials in the following discussion, while remembering that other types of crystals, gasses or liquids are widely employed in macroscopic laser configurations.

Multi-level systems

As we saw in Section 3.5, the band structure of semiconductors allows us to consider them as electronic systems with two energy levels, as shown in Figure 10.1. Electronic transitions between the conduction band (E_c) and valence band (E_v) define the nature of photon emission or absorption.

In thermal equilibrium, the energy distribution of electrons in such a two-level system may be described by the Maxwell-Boltzmann² distribution, namely

$$\frac{n_2}{n_1} = \exp\left(-\frac{E_2 - E_1}{k_B T}\right) \quad (10.1)$$

where n_2 [m^{-3}] represents the density of electrons in the higher energy level, E_2 , and n_1 [m^{-3}] the density in the lower energy level, E_1 ; k_B is the Boltzmann constant and T the temperature. Since electrons are fermions, the astute student will protest that,

² Ludwig Boltzmann (1844–1906), the great Austrian physicist, is considered to have been the founder of the fields of statistical mechanics and thermodynamics, and spent most of his peripatetic career at the University of Vienna. A manic-depressive, he hung himself near Trieste and is immortalized by the epitaph on his grave in Vienna, reading “ $S = k \log W$ ”.

strictly, their energy distribution is described by the Fermi-Dirac distribution³, which is, however, well approximated by the Maxwell-Boltzmann distribution for the case $E_2 - E_1 \gg k_B T$. In the semiconductor bands shown in Figure 10.1, we have $E_2 = E_c$ and $E_1 = E_v$, and the density of electrons in the conduction or valence bands is given by n_c [m^{-3}] and n_v [m^{-3}], respectively.

Transitions

Based on this two-level system, we can define three basic types of electronic transitions which result in the generation or absorption of photons:

Absorption (Figure 10.1(a)) takes place when a photon with energy $E = h\nu > E_c - E_v = E_g$ causes an electron to be raised from the valence to conduction bands. The photon is absorbed and its energy transferred to the electron. The level of absorption is proportional to the intensity of the optical field and the number of electrons in the valence band, n_v .

Spontaneous emission (Figure 10.1(b)) is the opposite process: an electron falls from the conduction to valence bands, emitting a photon with energy $h\nu = E_c - E_v = E_g$. This process takes place spontaneously, with no external stimulus, and its rate is proportional to the number of electrons in the conduction band, n_c .

Stimulated emission (Figure 10.1(c)), in contrast, occurs when an incident photon stimulates the transition of an electron from conduction to valence bands, such that two photons, with energy $h\nu = E_g$ and with the same phase, are emitted. The rate of stimulated emission is proportional to n_c and the intensity of the incident optical field.

These three processes form the basis for light emitters and absorbers, as we will see repeatedly below. LEDs generally rely on spontaneous emission, lasers require stimulated emission and the functionality of most photodetectors is based on absorption.

Processes in thermal equilibrium

Equation 10.1 tells us that the probability of finding an electron at higher energies decreases exponentially with increasing energy, so that, at thermal equilibrium, there will always be considerably more electrons in the lower energy levels and the upper levels will be relatively empty. Therefore, a direct-bandgap semiconductor in thermal equilibrium has very few electrons available for spontaneous emission, and the material does not spontaneously glow in the dark.

³ Enrico Fermi (1921–1954), an Italian physicist, was one of the truly brilliant theoreticians and experimentalists of the twentieth century. He developed statistical laws describing fermions, worked extensively on nuclear transformation due to neutron bombardment, and presided over the first controlled nuclear chain reaction. He was awarded the Nobel Prize for physics in 1938, whereupon he fled Italy for the United States. The English physicist and mathematician Paul Dirac (1902–1984), of apparently almost surreal intellectual prowess, made wide-ranging contributions to the foundations of quantum mechanics. An exceedingly taciturn personality, upon receiving the physics Nobel Prize in 1933 he was described by an English newspaper to be “as shy as a gazelle and as modest as a Victorian maid” (Segré, 2007, p. 83).

An optical field incident on a semiconductor will stimulate transitions from E_v to E_c , resulting in photon absorption, Figure 10.1(a). Since the electron population of the conduction band is exponentially smaller than that of the lower band, stimulated emission processes are possible but exponentially less likely than absorption, and thus materials in thermal equilibrium primarily absorb light. The optical intensity I in such a material decreases with propagation distance in z as

$$I = I_0 \exp(-\alpha z) \quad (10.2)$$

for absorption coefficient α ; this last expression is of course Beer's law, to which we were introduced in Section 3.3.1. Optical absorption forms the basis for semiconductor detectors.

Processes in non-equilibrium

By adding energy to a multi-level optical system, the population of the higher energy levels can be increased, such that the system is no longer in thermal equilibrium and the Maxwell-Boltzmann distribution of Equation 10.1 no longer holds. In semiconductors, this non-equilibrium situation is generally achieved by electrical current injection, a process known as pumping, thereby populating the conduction band so that n_c becomes non-negligible.

When electron density in the conduction band is high, the probability of spontaneous emission, Figure 10.1(b), increases and a direct bandgap semiconductor will emit photons in direct proportion to n_c . As a result, efficient optical emission results and this process is that on which the light emission from an LED directly relies. In addition, with a populated conduction band, the presence of an optical field can lead to stimulated emission, whereas absorption also continues to take place so long as electrons are available in the valence band.

10.1.2 Optical amplification

In non-equilibrium, then, all three optical transitions become likely. Under the appropriate circumstances, namely strong pumping and the existence of a strong optical field, stimulated emission can dominate over optical absorption. This condition is that required by a laser.

Optical field

A semiconductor with a high excited electron density emits photons, but also continues to absorb these. With sufficiently high electron injection, we can achieve negative absorption, or amplification, in the presence of an intense optical field.

The optical field, or spectral energy density, in the material, $\Phi(\nu)$ [$\text{J}/\text{m}^3\text{Hz}$], is given by

$$\Phi(\nu) = \Phi_\nu g(\nu) \quad (10.3)$$

where Φ_ν [J/m^3] corresponds to an optical energy density per volume and $g(\nu)$ [Hz^{-1}] represents the optical lineshape function, and thus the optical linewidth, as we defined

in Section 5.1.1. The lineshape function typically has the Lorentzian form of Figure 5.1, and is normalized so that

$$\int_{-\infty}^{+\infty} g(\nu) = 1. \quad (10.4)$$

The optical field induces stimulated emission, if there are electrons in the conduction band which can undergo a transition to the valence band and emit a photon, and this leads to optical amplification.

The two necessary conditions for achieving amplification, and thus lasing, in a material are:

1. Stimulated emission is dominant over spontaneous emission; and
2. Stimulated emission is dominant over absorption.

The first condition can be fulfilled if the optical field $\Phi(\nu)$ is sufficiently high, so that downward transitions are predominantly stimulated ones. The second condition is fulfilled if pumping (the injection of electrons) is so strong that $n_c > n_v$, a condition known as population inversion. In this case, there are few electrons in the valence band and absorption is suppressed. Thus a laser requires a strong pump and an intense optical field. We will see presently how these conditions are achieved.

Gain

Under conditions of high optical field and population inversion, a material can exhibit optical amplification, or gain. The gain factor γ [m^{-1}] may be derived as (Zappe, 2004, Chapter 2.4)

$$\gamma = \frac{\lambda^2}{2\pi\tau} (n_c - n_v) g(\nu) \quad (10.5)$$

where n_c and n_v are the electron densities in the upper and lower energy bands, respectively, and τ [s] is the lifetime of electrons in the upper energy level. We see thus that $\gamma > 0$ if $n_c > n_v$, implying that population inversion due to high levels of pumping is essential for gain. Equation 10.5 also shows that gain is a function of wavelength and the lineshape of the optical field, such that a strong frequency dependence is expected. Finally, high gain requires a short electron lifetime τ in the upper energy level.

With a positive gain factor, the intensity of an optical field in the material increases with propagation as

$$I = I_0 \exp(+\gamma z), \quad (10.6)$$

in direct contrast to the case for absorption, as we saw in Equation 10.2.

An example of the variation of semiconductor gain with wavelength and the strength of electrical pumping is shown in Figure 10.2. In this plot, γ has been calculated for GaAs for four different pumping levels, and thus different electron concentrations, in the conduction band. We see a strong wavelength variation, with gain increasing for energies above E_c (or for wavelengths below the gap wavelength, about 869 nm). For sufficiently strong pumping, γ becomes positive for a range of wavelengths.

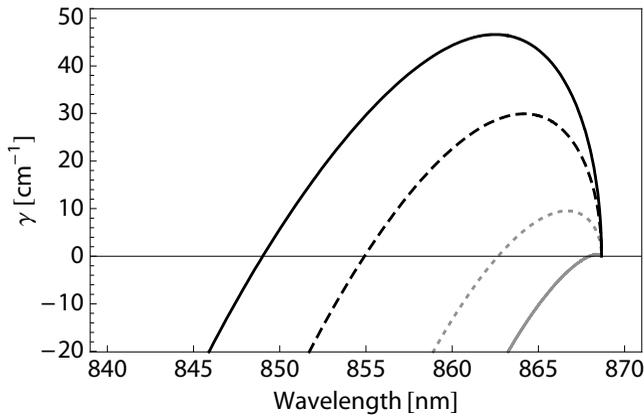


Figure 10.2 Calculation of gain in GaAs for different electron concentrations: $n = 1.2 \cdot 10^{18} \text{ cm}^{-3}$ gray, solid; $n = 1.5 \cdot 10^{18} \text{ cm}^{-3}$ gray, dashed; $n = 2.0 \cdot 10^{18} \text{ cm}^{-3}$ black, dashed; and $n = 2.5 \cdot 10^{18} \text{ cm}^{-3}$ black, solid. The bandgap wavelength of GaAs is about 869 nm. Positive γ implies optical amplification.

Even with positive gain, the semiconductor material still exhibits absorption and other forms of optical loss, due to scattering or transmission outside of the material. If γ is sufficiently high to compensate all of this residual loss, the *net* gain is positive, implying that the optical field is amplified in a certain wavelength range.

10.1.3 Laser resonance

Using an optically-active material which exhibits gain under strong pumping, we have all but one of the necessary components of a laser. To make the final step from light-emitting material to a laser, the pumped laser material is placed into an optical resonator: the resonator allows the generation of the high optical field intensities necessary to obtain positive gain and the resonance frequency defines the lasing wavelength. We see in this section what is necessary to generate laser emission.

Resonator with gain

The conceptually simplest laser resonator, and that probably most widely used, is the Fabry-Perot cavity, to which we were introduced in Section 5.2.5. As seen in the schematic representation of a laser cavity in Figure 10.3, the gain material, with real refractive index n , residual absorption loss α_0 and gain γ , is placed into a Fabry-Perot cavity of length L . Upon pumping, the generated optical field is reflected by the mirrors with reflectance R , leading to resonance for oscillation in the z direction. As a result, the spontaneous and stimulated emission generated by the semiconductor material is reflected by the resonator mirrors and the optical field increases rapidly in intensity at the resonance wavelength.

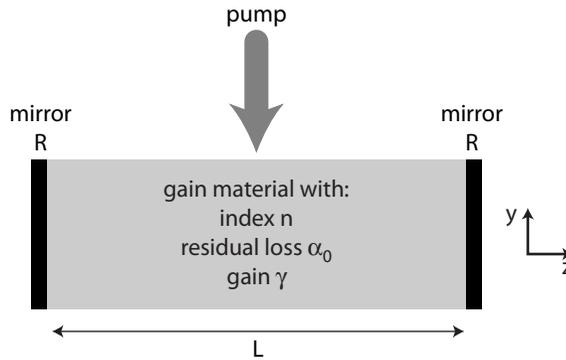


Figure 10.3 The schematic laser: optically-active gain material with real refractive index n , residual background loss α_0 , and gain γ is pumped to generate population inversion and an optical field in the Fabry-Perot cavity of length L . The mirrors with reflectance R lead to resonance for oscillation in the z direction.

Based on our analysis of Section 5.2.5, we can write the propagation intensity in the z direction in the cavity, $I(z)$, as

$$I(z) = I_0 e^{j2kz} = e^{j2k_R z} e^{-2k_I z} = e^{j2\beta z} e^{-\alpha z}, \quad (10.7)$$

recalling that in Equation 3.32 of Section 3.1.1, the complex propagation constant k was separated into its real and imaginary components, k_R and k_I . We have further used the definition

$$\beta = \frac{2\pi n}{\lambda_0} \quad (10.8)$$

for real refractive index n and vacuum wavelength λ_0 , and set $2k_I$ equal to the optical absorption coefficient, α ; refer to Section 3.3 for the general discussion of propagation through absorbing media.

In the absence of externally supplied energy, the laser material has unexcited values of $\beta \rightarrow \beta_0$ and $\alpha \rightarrow \alpha_0$. Under pumping, however, the injection of free carriers (electrons and holes) leads to a change in both the real propagation constant,

$$\beta \rightarrow \beta_0 + \Delta\beta \quad (10.9)$$

and the absorption coefficient,

$$\alpha \rightarrow \alpha_0 - \gamma, \quad (10.10)$$

where we recall that γ is the gain and thus represents “negative absorption”. The intensity propagation through the cavity in its excited state then becomes

$$I(z) = e^{j2(\beta_0 + \Delta\beta)z} e^{-(\alpha_0 - \gamma)z}. \quad (10.11)$$

From Equation 10.10, we see that for $\gamma > \alpha_0$, the total absorption becomes negative and the argument of the second exponential in Equation 10.11 becomes positive. As a result, the optical intensity grows exponentially with distance when propagating through the medium; the point at which this occurs is known as laser threshold.

Laser threshold

We now consider the effect of the resonant Fabry-Perot cavity. Recalling from Equation 5.50 of Section 5.2.5 that the intensity transmission through a Fabry-Perot etalon is given by

$$\frac{I_t}{I_0} \propto \frac{1}{1 - Re^{j\phi}}. \quad (10.12)$$

for phase ϕ and mirror reflectance R , where we recall that in general $R < 1$. Cavity resonance occurs when the denominator goes to zero, or

$$Re^{j\phi} = 1, \quad (10.13)$$

which, when considering a cavity of length L and the total propagation constant used in Equation 10.11, leads to

$$Re^{j2(\beta_0 + \Delta\beta)L} e^{-(\alpha_0 - \gamma)L} = 1. \quad (10.14)$$

The previous expression has two components: a harmonic one (with imaginary exponent) and an exponentially decaying or growing one (with real exponent). These must independently equal unity at resonance, so that we can evaluate them separately. Taking the harmonic solution first, we have

$$Re^{j2(\beta_0 + \Delta\beta)L} = 1 \quad (10.15)$$

implying that

$$2(\beta_0 + \Delta\beta)L = 2i\pi \quad (10.16)$$

for integer i . Since $\beta_0 + \Delta\beta$ represents the value for z -directed propagation constant, we may re-write this as

$$\beta_0 + \Delta\beta = \frac{2\pi n}{\lambda_0} \quad (10.17)$$

where n is the refractive index under conditions of pumping, and may be different than that for unpumped material, primarily due to the temperature changes which occur with high carrier injection. Combining the last two equations, we can then solve for wavelength to yield

$$\lambda_0 = \frac{2nL}{i} \quad (10.18)$$

recalling that $i = 1, 2, 3, \dots$. This last expression defines the resonant lasing wavelengths in the Fabry-Perot cavity, which are a function of cavity length L and its refractive index. These discrete resonant wavelengths represent the longitudinal modes of the laser; we remember that the refractive index of the material varies with pumping.

Now considering the exponential solution of Equation 10.14, we have

$$Re^{-(\alpha_0 - \gamma)L} = 1 \quad (10.19)$$

which we can simply re-write as

$$\gamma_{th} = \alpha_0 + \frac{1}{L} \ln \frac{1}{R} \quad (10.20)$$

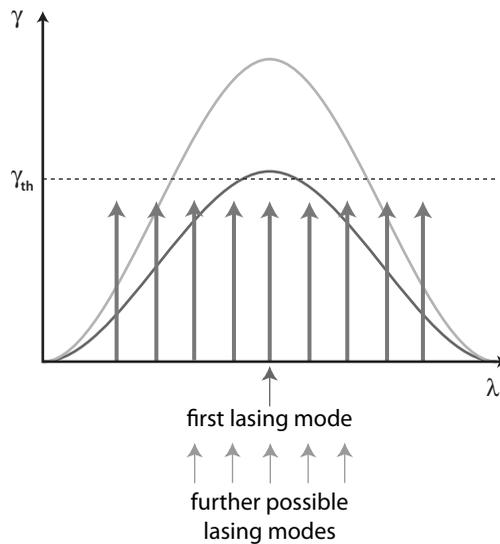


Figure 10.4 The principle behind lasing: two gain curves, for low *dark gray* and high *light gray* pumping are shown along with the cavity resonance wavelengths, indicated by the vertical arrows. The first lasing mode occurs at the resonance wavelength at which the gain γ compensates all the residual losses and thus reaches the threshold gain, γ_{th} . For higher pumping, more lasing modes are possible, but generally the one with the highest gain lases.

where we have defined the threshold gain γ_{th} [m^{-1}]. The parameter γ_{th} represents the gain required to offset all other losses in the system: the residual absorption, α_0 , and the loss through the mirrors, since $R < 1$. When pumping is sufficient to result in net gain $\gamma > \gamma_{th}$, the lasing threshold is reached and the laser lases. The photons “lost” through the mirrors are those emitted as laser radiation.

The complete laser

In Figure 10.4 we have a schematic summary of the effects which lead to lasing. Plotted as a function of wavelength, the gain curves increase in magnitude as pumping increases, until at some point all the residual losses have been compensated and the laser threshold is reached. Concomitantly, the laser cavity has discrete resonance wavelengths, represented by the vertical arrows in the figure, and it is at those wavelengths that the optical field intensity grows strongly.

The wavelength at which a cavity resonance coincides with $\gamma > \gamma_{th}$ is then the wavelength at which the laser begins to lase; this point is indicated by “first lasing mode” in the figure. As pumping increases, γ becomes larger than γ_{th} for an increasing range of wavelengths, as shown by the higher gain curve. Each cavity resonance for which $\gamma > \gamma_{th}$ may then potentially lase, but in practice the mode (or modes) with the highest gain are the ones which ultimately lase.

We may thus summarize the processes required to obtain lasing from an optically active material:

- The material must be pumped, for semiconductors usually electrically or occasionally optically, such that a non-equilibrium condition is achieved and spontaneous emission generates a high optical field; with this condition, we already have an LED.
- Through pumping, the density of electrons in the upper energy level becomes larger than that in lower levels (population inversion) and the dominance of the resultant stimulated emission leads to positive gain and thus amplification of the optical field.
- The resonant cavity in which this material is then placed gives rise to strong growth of the optical field at the resonant wavelengths, eventually leading to lasing at that wavelength for which gain is positive; we now have a laser.

It should be emphasized that *both* an intense optical field and population inversion through pumping are required for lasing. Population inversion cannot be achieved by just increasing the intensity of the optical field itself, since we saw in Equation 10.5 that the factor $(n_c - n_v)$ must be positive for γ to be positive. Optically pumped lasers use a different (shorter) wavelength to achieve population inversion, not the wavelength of the laser itself.

Example 10.1: Let us estimate the threshold gain and resonance wavelength of a 300 μm long GaAs-based laser which has a gain peak at 862 nm and residual optical absorption of 7 cm^{-1} .

From Table 3.2, we find that the refractive index of GaAs is about 3.65 (determined for $\lambda = 862\text{ nm}$). As a result, the reflectance of the mirrors, found from Equation 4.40, may be determined to be $R = 0.32$. From Equation 10.20, we can then directly determine

$$\gamma_{th} = \alpha_0 + \frac{1}{L} \ln \frac{1}{R} \approx 44\text{ cm}^{-1}$$

which is a typical value for gain in a semiconductor, as we saw in the plot of Figure 10.2.

The resonance wavelength can be found with the assistance of Equation 10.18. Using the wavelength of the gain peak, we find that a typical value for i for this cavity length would be $i = 2540$ or $i = 2541$; these longitudinal mode indices represent the integer number of wavelengths which would fit into the 300 μm long cavity. Plugging these two values for i into Equation 10.18, we find resonant wavelengths of 862.21 nm and 861.87 nm, respectively. It is possible that either one or both of these closely spaced modes will first lase, since the gain peak is relatively broad. A caveat concerning this calculation is that it is unlikely that the cavity length (assumed to be $L = 300.00\text{ }\mu\text{m}$) is really known to this level of accuracy.

10.1.4 Materials

From the previous discussion, we see that LEDs and lasers require materials which are optically active, meaning that they efficiently convert electrical energy, in the form of

injected electrons, into optical energy. As we mentioned in the introduction, a wide array of materials (solids, liquids, gasses and even a gelatin dessert (Hänsch et al., 1971)) are suitable for fabricating lasers.

The most relevant materials for active micro-optical components are semiconductors. In addition, optically active polymers are playing an increasingly important role in active micro-optics, and we discuss some of these in Sections 10.2.4 and 10.4.2 below. We saw in Section 3.5 that the optical and electrical properties of semiconductors are defined by their band structure. Of primary interest is the magnitude of the bandgap energy, E_g , since that defines the photon emission wavelength. Recall that many compound semiconductors, such as those listed in Table 3.3 or plotted in Figure 3.14, have an E_g giving rise to transitions with photon energies in the ultraviolet to infrared wavelength ranges; the exact semiconductor employed will depend on the desired emission wavelength, as given in Table 3.3.

In addition to the magnitude of the bandgap, a further important consideration is the nature of the bandgap itself. In Figure 3.12, we sketched the energy bands for a direct-bandgap semiconductor, for which the maximum of E_v (as a function of k) is at the same position on the k axis as the minimum of E_c . In a direct-bandgap material, the transition of electrons between valence and conduction bands occurs with $\Delta k = 0$ and, since photons can have high energies (E) but only negligible momentum (k), this process is a very efficient one. Direct bandgap semiconductors are thus the most suitable materials for optical emission.

Silicon, on the other hand, is an indirect-bandgap semiconductor, such that an electronic transition between E_c and E_v requires the addition of considerable momentum ($\Delta k > 0$) for conservation of E and k . Non-negligible k is usually derived from phonons in the material, such that an electronic transition between conduction and valence bands requires the combined efforts of a photon and one or more phonons, making it a very inefficient and thus unlikely process. For this reason, pure silicon is a poor emitter of light, although some progress has been made in using Si for light emission, as we will see in Section 10.3.6. Silicon is, nevertheless, an adequate absorber of optical radiation, since the absorption of photons with energies above E_g is possible in a direct transition; the steepness of the absorption curve with decreasing wavelength is less than that of a direct-bandgap semiconductor, however.

10.2 Light-emitting diodes

The light-emitting diode uses current injected into an optically active material to generate photons; thus semiconductor LEDs rely on the spontaneous emission process of Figure 10.1(b) for a direct bandgap material. LEDs are relatively simple to fabricate and thus are used as efficient, and usually low-cost, light emitters for a wide variety of optical systems; anyone who has strolled through Ginza 4-chome in Tokyo at night has been blinded by truly macroscopic LED implementations. For micro-opticians, it is the small size, high efficiency and wide range of available wavelengths that makes the LED attractive as a light source for many types of optical microsystems.

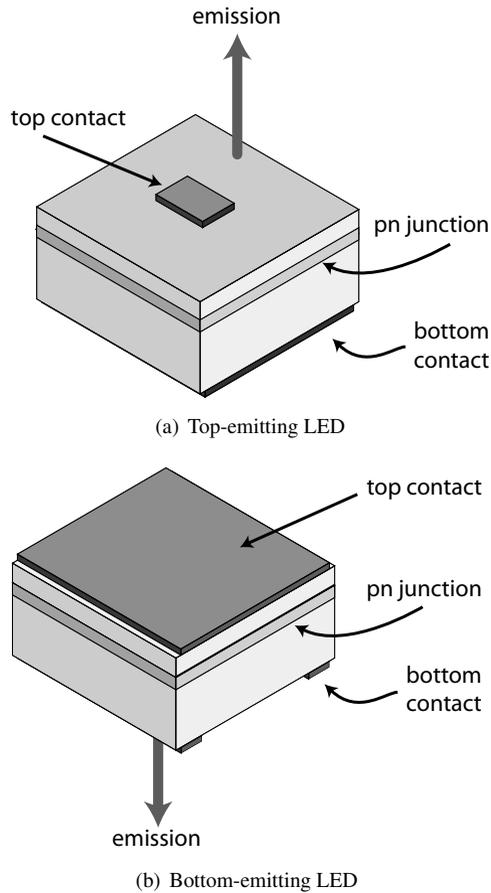


Figure 10.5 Simple semiconductor LED chip structures, designed for emission from the surface (top-emitting) or through the substrate (bottom-emitting).

10.2.1 LED structures

As its name suggests, the LED is electrically a diode. The p and n regions are defined by doping and when the diode is forward biased, a high density of electrons in the conduction band is generated at the pn junction. In direct-bandgap materials, these electrons then recombine spontaneously across the bandgap and thereby emit photons at the bandgap energy, as we just saw. The basic structure of the semiconductor LED is thus quite simple, although considerable care needs to be taken in the design to assure that most of the generated photons are actually emitted from the material; this latter consideration is usually non-trivial.

Examples of two classical LED chip structures are shown in Figure 10.5. A pn diode is typically formed in the entire plane of the semiconductor wafer, close to one surface. Electrical contacts are deposited and patterned on the front and back surfaces. The simplest structure is the top-emitting LED, as sketched in Figure 10.5(a); emission is



Figure 10.6 Photograph of a typical surface-emitting LED chip; the length of one side is $360\ \mu\text{m}$. The X-shaped pattern is the surface electrical contact and the bond wire is seen emerging toward the bottom of the photo. Photograph courtesy of Robert Gehrke.

through the thin (typically $1\ \mu\text{m}$ or less) surface layer above the pn junction and the top electrical contacts are patterned so as not to cover an excessive area of the light emitting active region. Alternatively, the LED can then be designed to be bottom-emitting, as shown in Figure 10.5(b), if the substrate is transparent for the emission wavelength; this condition is typically realized by using a different (lower bandgap) material for the active region around the pn junction than for the bulk substrate.

Materials

Semiconductor-based LEDs use direct-bandgap semiconductors with E_g chosen so that emission is at the desired wavelength; most of these are III-V compound semiconductors. As we summarized in Table 3.3, a variety of materials is commonly used: InGaN for blue LEDs (Nakamura and Chichibu, 2000), AlGaInP for the red wavelengths and GaAs/AlGaAs for the near infrared. For the infrared wavelengths in the telecommunications bands, InGaAsP is most suitable. Of increasing relevance are organic and quantum dot-based light emitters, which we both address below.

Sizes

LEDs are efficient light emitters and can emit considerable optical power from a small active area. Typical chip sizes, as in the example shown in Figure 10.6, are $350 \times 350 \times 100\ \mu\text{m}^3$, with an emitting area in the range of $0.3 - 0.5\ \text{mm}^2$. For use in microsystems, these are available as individual chips, in surface-mount device (SMD) configurations or in relatively macroscopic packages, often including reflectors and integrated lenses for optimizing the out-coupling efficiency of the generated light.

Structural variations

As might be clear from examination of the simple LED structures of Figure 10.5, one of the primary design challenges in developing an efficient light-emitting diode is assuring that most of the photons generated at the pn junction emerge from the semiconductor

chip. Significant optical losses result from Fresnel reflection at the chip surfaces, and total-internal reflection in particular can trap a significant fraction of the generated light. In addition, absorption in the bulk of the semiconductor reduces the emitted intensity, as does reflection back into the substrate from the (typically non-transparent) electrical contacts.

As a result, numerous variations in LED chip design have been designed to optimize the overall optical efficiency. The Burrus LED structure (Burrus and Ulmer, 1971) results if the substrate of the chip is etched away from the back side almost to the active region, so that emission takes place from both top and bottom surfaces, even if the substrate was absorbing. An alternative uses epitaxial lift-off to remove the light-emitting active layers from an absorbing substrate and re-attaches the layers to a transparent carrier substrate; orange-red AlGaInP LEDs, for example, are grown on (absorbing) GaAs and transferred in this manner to a (transparent) GaP substrate.

As we will see in the following example, the extraction of photons from the chip is primarily limited by Fresnel and total internal reflection at substrate boundaries; for this reason, considerable effort has been expended on the development of optical features which enhance photon output coupling. Suitable surface features include etched prisms or lenses, roughened interfaces (Schnitzer et al., 1993; Möller and Forrest, 2002) or photonic crystals (Wierer et al., 2009); in addition, buried low-refractive-index grids have been used with organic LEDs (Sun and Forrest, 2008).

Example 10.2: Let us estimate the photon emission efficiency due to the chip structure of an AlGaInP LED. We consider first a $100\ \mu\text{m}$ thick chip, with refractive index $n = 3.6$ and a material absorption of $\alpha = 20\ \text{cm}^{-1}$; this configuration is a somewhat artificial one, since most LEDs of this type are fabricated by epitaxial growth of AlGaInP on GaAs (Streubel et al., 2002), but we can use it for purposes of illustration. Assume the photons are generated in a pn junction formed $1.5\ \mu\text{m}$ below the surface.

There are three components to photon loss in the chip: absorption in the material; Fresnel reflection at the surface; and total internal reflection. Absorption may be calculated from Beer's law, using the thickness of the material and the known value of α . From this we determine, for the back surface

$$\frac{I_{back}}{I_{source}} = \exp[-\alpha(100\ \mu\text{m} - 1.5\ \mu\text{m})] \approx 0.821$$

and for the front surface

$$\frac{I_{front}}{I_{source}} = \exp[-\alpha(1.5\ \mu\text{m})] \approx 0.997.$$

Fresnel reflectance, as we know from Section 4.3, has a value based on the refractive indices of the material and the external medium, which we can assume to be air; since the emitting pn junction is a large planar area parallel to the top and bottom surfaces of the chip, we can approximate reflectance by the normal incidence condition of

Equation 4.40, namely

$$R_{normal} = \left(\frac{n_{AlGaInP} - n_0}{n_{AlGaInP} + n_0} \right)^2,$$

such that the transmittance of the field outside the chip from the top and bottom surfaces is given by

$$T = 1 - R \approx 0.68.$$

Rays incident at non-normal incidence will have a higher value for reflectance (recall Figure 4.7) and thus lower transmittance; we overestimate the efficiency with the normal-incidence approximation.

Finally, we determine the critical angle of the substrate/air interface,

$$\theta_c = \arcsin \left(\frac{n_0}{n_{AlGaInP}} \right) \approx 16.1^\circ.$$

Realizing that any rays emitted from the planar emitter which impinge on the interface at angles greater than θ_c are subject to total internal reflection, we can determine the fraction which emerge (assuming uniform emission from the junction into an angle of 2π) as

$$\eta_{crit} = \frac{2\theta_c}{2\pi} \approx 0.179.$$

An estimate for the total fraction of the photons emitted from the chip (ignoring emission from the facets normal to the plane of the substrate) is then the product of the efficiencies just calculated, which, taking an average of the values for the upper and lower interfaces, is determined to be

$$\eta_{total} = \frac{1}{2} \left(T \frac{I_{back}}{I_{source}} \eta_{crit} + T \frac{I_{front}}{I_{source}} \eta_{crit} \right) \approx 0.111.$$

Thus only about 11% of the photons generated in the active region manage to emerge from the chip, unless measures are undertaken, for example, to reduce R , as we mentioned above. This example should illustrate the importance of packaging in the overall design of an efficient micro-optical light emitter.

10.2.2 LED operating characteristics

As diodes, LEDs are current-driven devices; the voltage drop is slightly larger than the bandgap voltage, so that, save for very short wavelength materials, the applied voltages are compatible with low-voltage microsystems-based drivers. Current levels vary with the required output power, but are typically in the range of several tens of milliamperes, such that total electrical power input is in the tens to hundreds of milliwatts.

Since the light-emitting diode relies on spontaneous emission of photons, it thus has a relatively broad emission spectrum, typically with $\Delta\lambda$ on the order of 50 to 100 nm.

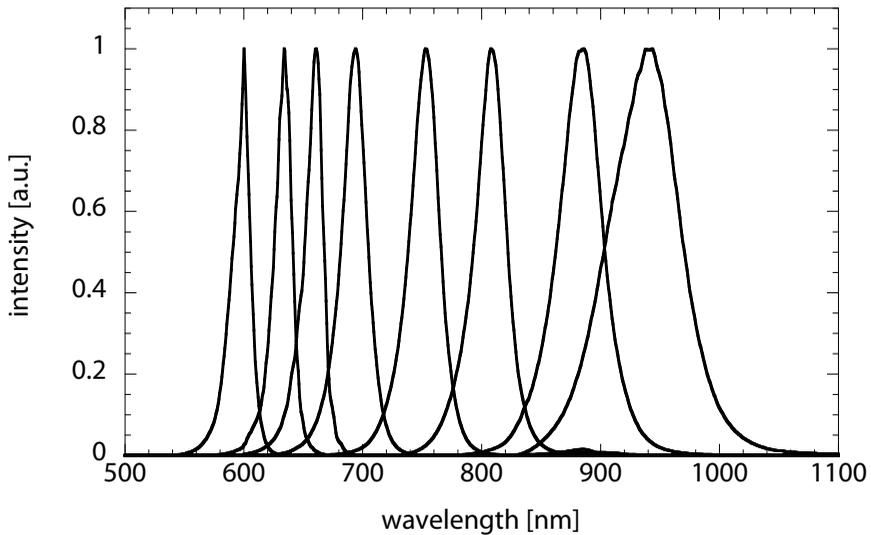


Figure 10.7 Measured emission spectra for eight LEDs in the visible to near infrared wavelength regimes; the peak intensities have been normalized to unity. Spectra courtesy of Robert Gehrke.

Typical spectra for a selection of LEDs in the visible and near-IR wavelength ranges are shown in Figure 10.7. The wavelengths available range from the near UV to the mid-IR, with these dependent on the availability of direct-bandgap materials with the appropriate bandgap energy. So-called “white” LEDs are often in practice blue LEDs on which an inorganic fluorescent layer (for example Ce-doped YAG, $Y_5Al_5O_{12}$), which emits in the yellow, has been deposited; the white is then generated by a balanced combination of blue and yellow emission (Baur et al., 1998). White emission is also generated by three LEDs (red, green and blue) in a single package, an approach popular for large displays.

Typical output power levels vary widely, but are generally in the range of several tens of milliwatts for drive currents up to 50 mA for devices useful in microsystems applications. High-power LEDs, with optical output powers up to 5 W and luminous flux of almost 200 lm, are available. Efficiency is high, with typical values of 100 lm/W (output lumens per input electrical Watts⁴). Native optical emission from most LED chips is omnidirectional, but many packaged devices are combined with reflecting and lens structures to achieve a roughly circularly-symmetric, directional emission.

10.2.3 Resonant-cavity and superluminescent LEDs

The emission intensity of LEDs can be enhanced by incorporating the light-emitting active region into a resonant cavity. Using a vertical Bragg mirror, similar to that used

⁴ The lumen (lm) is the unit of luminous flux, given by $cd \cdot sr$ (candela steradians). The candela (cd) defines luminous intensity and is a fundamental SI unit. 1 lm is then the luminous flux emitted by an isotropic point source of luminous intensity 1 cd into 1 sr.

for the VCSEL (Section 10.3.4), underneath the active region, a portion of the light is reflected back into the pn junction, increasing the optical intensity and giving rise to incipient stimulated emission (Wirth et al., 2001). As a result, the emission from the surface of a resonant-cavity LED increases strongly in intensity and directionality, and the spectral width is reduced with respect to the garden-variety LED.

Increasing the quality of the resonant cavity further, to that suitable for lasing, results in a superluminescent light emitting diode (SLED). Employing a laser cavity, in which strong feedback results in stimulated emission, the SLED is designed to inhibit lasing by anti-reflection coatings on the mirrors or coupling out a significant fraction of the generated light, so that the laser remains perpetually below threshold (Okamoto et al., 1998). The emission is, however, of considerably higher intensity and narrower spectrum than a conventional LED, with a beam quality approaching that of a laser diode.

10.2.4 Organic light emitting diodes

As part of an effort to reduce the cost and improve the mass-manufacturability of light-emitting devices, polymer-based organic LEDs (OLEDs) have seen considerable development in the past decade. Although electroluminescent emission from polymers was well known, it was not until the late 1980s that practical polymer-based light emitting devices operating at low voltage were demonstrated, first using small molecules (Tang and VanSlyke, 1987) and then conjugated polymers (Burroughes et al., 1990). OLEDs rely on quite different principles than semiconductors for light emission, but due to ease of manufacturability, particularly for large areas, organic LEDs are rapidly growing in popularity and performance, particularly for displays (Shinar, 2003), and plastics in general are playing an increasingly significant role in optoelectronics (Forrest, 2000, 2004).

Polymer materials

OLEDs may be manufactured from one of a variety of polymer materials; two of the most widely used have been PPV (poly-(para-phenylene vinylene)) and MEH-PPV (poly(2-methoxy-5-(2'-ethyl-hexoxy)-1,4-phenylene-vinylene)). PPV is an example of a π -conjugated polymer, in which the π orbitals of the monomers are delocalized thus forming hybridized orbitals along the polymer chain; these orbitals are only partially filled with electrons such that the material represents an organic semiconductor, albeit with relatively high resistance, since electronic transport is by carrier hopping. Electrons and holes injected into this material combine to form excitons, which in turn decay; singlet excitons decay radiatively whereas triplet excitons decay non-radiatively, the latter representing a limitation on the quantum efficiency of the OLED.

The polymer materials are typically deposited by spin-coating or screen printing (Birnstock et al., 2001) on a glass substrate, over areas up to $0.35 \times 0.35 \text{ m}^2$; large-area roll-to-roll OLED processes are under development. Two different materials are used for electrical contact as shown in the schematic view of the OLED structure in Figure 10.8. The cathode is generally a metal, such as Au, Al or Ca, or a fluorinated material such as LiF. The anode is typically indium tin oxide (ITO, a transparent conductor), using

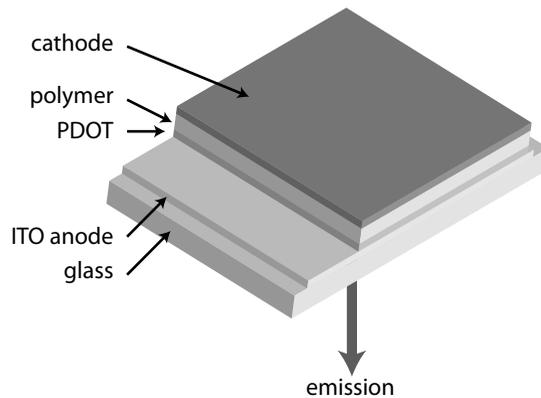


Figure 10.8 Schematic cross-section of an OLED. The cathode is typically a metal; ITO is a transparent conductor; and the PDOT layer is used to isolate the polymer from the anode material. Emission is through the glass substrate.

another polymer, such as PDOT (poly(3, 4-ethylene-dioxythiophene), sometimes also referred to as PEDOT) to improve hole injection efficiency and reduce the diffusion of tin from the ITO into the PPV layer (Patel et al., 2002). Many structures employ more complex layer sequences, particularly to enhance efficiency by, for example, increasing the ratio of singlet to triplet excitons.

Operating characteristics

Depending on the details of the polymer employed, the effective energy gap is typically in the range $1.5 \leq E_g \leq 3.5$ eV; emission from the earliest devices was in the green/yellow range, but OLEDs are now available with individual RGB pixels for full-color display applications (Fukuda et al., 2000; Müller et al., 2003). Typical operating voltages are about 8 V, with luminous intensities on the order of 10^4 cd/m² or 10 cd/A (Birnstock et al., 2001).

An issue which is still the subject of considerable research is the improvement of emission efficiency, limited by non-radiative transitions, charge balance and output coupling efficiency. The use of surface features, such as buried grids (Sun and Forrest, 2008), microlens arrays (Möller and Forrest, 2002), or a roughened glass surface, has improved the extraction efficiency to values approaching 20%, but overall quantum efficiency is still in the range of 5%. Record power efficiency values are at present in the range of those found in fluorescent tubes, about 90 lm/W (Reineke et al., 2009).

Applications

Whereas the small size of OLEDs, and their ease of fabrication (if a chemist is available), makes them attractive devices for use in compact, micro-optical systems, efficiency considerations still imply that electrical power input for a given optical output is high. The primary application for OLEDs is at present for compact displays with better contrast and improved viewing angles when compared to industrially standard LCD displays; their primary use is in consumer electronics devices such as mobile telephones.

A further promising application area is for ambient lighting. If fabrication technologies permit manufacture of OLEDs over large areas ($100 \times 100 \text{ cm}^2$), yielding devices with good efficiencies (100 lm/W) and adequate lifetime (10^6 h), all at low cost, they will likely prove to be a commercial success.

10.2.5 Quantum dot LEDs

Also currently of considerable interest are quantum dot-based LEDs. Quantum dots are spheres with diameters below about 10 nm, whose optical properties are defined by quantum confinement effects, rather than the material bandgap (Wood et al., 2009). The size of the quantum dots then defines their emission and absorption wavelengths, with narrow-band emission and broad-band absorption, making them attractive for use in new types of both LEDs and photodetectors. Due to their small size, quantum dots may easily be suspended in solution, so that liquid based deposition techniques, such as spin coating or ink-jet printing, may be used for fabrication.

Using CdSe/ZnS nanospheres, with diameters down to 3 nm, embedded in a polymer matrix and with transparent ITO contacts, LEDs with emission colors ranging from red through yellow to green have been fabricated (Sun et al., 2007). Power efficiencies of quantum dot emitters of 2.41 lm/W have been demonstrated (Cho et al., 2009) and external quantum efficiencies of 7%, almost in the range of those of OLEDs, have been shown. The advantage of quantum dots as the optical emitters for these structures is the generation of bright and saturated colors, and their use in two-dimensional displays is at present an area of intense academic and industrial research (Coe-Sullivan, 2009).

10.3 Laser diodes

Whereas the LED is a simple semiconductor light emitter, its spectral width, limited intensity and omni-directional emission restricts its use for many types of optical setups. Lasers circumvent many of these limitations and semiconductor laser diodes are often the light source of choice for micro-optical systems. Diode lasers have been the subject of considerable development in the past 50 years, and are now omnipresent in myriad applications, for the consumer primarily in optical data storage and, at the high end, in optical telecommunications systems.

We will digress again briefly to discuss the basic principles behind lasers, since much of this material is the subject of an extensive literature (Agrawal and Dutta, 1986; Chow and Koch, 1999; Liu, 2005; Zappe, 2004). We subsequently turn to a few examples of the most relevant laser diode structures for use in micro-optics, including the Fabry-Perot, DFB, DBR and vertical cavity lasers, and discuss their most salient characteristics.

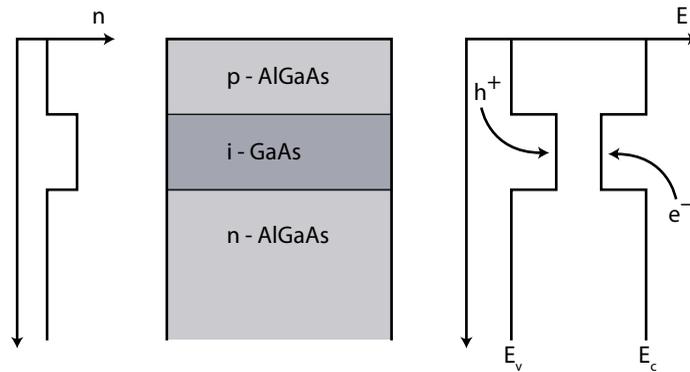


Figure 10.9 Cross-section through a double-heterostructure laser active region, using the GaAs/AlGaAs material system as an example. The GaAs layer, typically a few 100 nm thick, has a higher refractive index but lower E_g than the surrounding AlGaAs layers. As a result, injected electrons and holes are confined in the GaAs layer, leading to population inversion, and generated photons are likewise confined this layer, leading to stimulated emission and the generation of a high optical field intensity.

10.3.1 Laser structure and functionality

We were introduced to the basics of lasing in Section 10.1, with the most important effects summarized in the schematic sketch of Figure 10.4. In summary, a light-emitting material is subject to strong pumping to achieve population inversion, and the resultant optical field is amplified as it traverses the material. The field intensity is augmented by placing the material in an optical resonator, which concomitantly serves to define a resonant mode. The lasing wavelength is then given by that resonant mode for which the gain compensates all residual optical losses and the optical energy is concentrated into that single wavelength.

Laser core

The structure of the LED is well on the way to being a laser, but two important differences characterize the two devices: the laser benefits strongly from optical and electrical confinement in the active region, that volume in which the photons are generated and amplified, and also requires a resonant cavity.

Considering confinement first, the active region of a semiconductor laser diode is similar to that of an LED, which we saw schematically in Figure 10.5, in that it consists of a pn diode, typically in the plane of the semiconductor substrate. A key difference is seen in Figure 10.9: the active region is defined by a thin sheet of material with lower energy gap but higher refractive index than the surrounding material. In the example shown, the GaAs/AlGaAs system satisfies these requirements, with the added feature that GaAs and AlGaAs can be grown on one another epitaxially, due to close matching of the lattice constants. The arrangement of Figure 10.9 is termed a double heterostructure (since there are two material transitions) and thus forms a well for carriers (electrons and holes) and a waveguide for the generated photons.

The several 100 nm thick active layer, GaAs in this example, has a higher refractive index but lower E_g than the surrounding AlGaAs layers such that injected electrons and holes are confined in the GaAs layer. As a result, population inversion is more easily reached and the subsequently generated photons are likewise confined to this layer, leading to the generation of a high optical field intensity and thus stimulated emission. Such a layer sequence acts as a *separate-confinement double heterostructure*, since both carriers and photons are separately confined, and its development was key to the demonstration of semiconductor lasers which could be operated continuously at room temperature.

The laser core frequently includes one or more quantum wells, layers of material with even smaller energy gap than the core layer, and so thin that they represent a two-dimensional structure with discrete energy levels for electrons (Bastard, 1988); the positions of the energy levels are a direct function of the quantum well thickness. As the lowest- E_g component of the core, electrons and holes fall into the quantum well and recombine between two of the energy levels, such that the difference defines the emission wavelength. In addition to allowing accurate tuning of the emission wavelength, quantum wells generate considerably higher gain than bulk material and are thus almost always used in advanced semiconductor diode lasers.

The GaAs/AlGaAs layer sequence is one example of a widely used double-heterostructure. Other material systems are also popular, for different wavelength regimes. Mid-IR lasers, for the telecommunications wavelengths around 1.55 μm , use InGaAsP grown on InP substrates; red lasers are generally made using AlGaInP on GaAs; and blue lasers, using GaN and InGaN, are combined with an AlGaN cladding and grown on sapphire (Al_2O_3), SiC or spinel (MgAl_2O_4), albeit with relatively poor lattice matching.

Resonant cavities

The resonant cavity of a semiconductor laser may take one of several forms. The most basic is the Fabry-Perot etalon, which we used as the basis for our analysis of Section 10.1.3 and this may be formed by cleaved semiconductor facets, as we will see in Section 10.3.2 presently. Alternatively, wavelength-selective Bragg reflectors may be employed, as is the case for the DBR and DFB lasers of Section 10.3.3 and the VCSELs of Section 10.3.4. In all of these cases, the cavity is designed to generate sufficient optical feedback by generating an optical resonance at a wavelength corresponding to the gain peak.

Materials and emission wavelengths

As was the case above for LEDs, the materials used for laser diodes depend on the desired wavelength regime; the direct bandgap energy has to correspond to the required photon energy. Referring again to Table 3.3, we see there are a number of important material systems, including GaN for the UV and visible blue (Nakamura et al., 2000; Nakamura and Chichibu, 2000) AlGaInP for red (Yoshikawa et al., 1995), GaAs/AlGaAs for near IR and InGaAsP for the IR telecommunications wavelengths (1.3 to 1.55 μm) (Agrawal and Dutta, 1986).

A decided gap in the range of available wavelengths is the green part of the visible spectrum. The green laser pointers which have become *de rigueur* at conferences are actually YVO_4 (yttrium aluminum vanadate) solid state lasers, pumped at 808 nm and frequency doubled using a KTP (potassium titanyl phosphate, KTiOPO_4) crystal (Bai and Chen, 2002). A number of materials are being developed to fill this gap, including longer wavelength InGaN lasers (Okamoto et al., 2009) or relatively exotic compounds such as the II-VI materials BeZnSeTe on InP (Nomura et al., 2009); recent experimental work on InGaN-based semiconductor lasers has resulted in high output powers around 520 nm.

Electrical structure

As LEDs, laser diodes are electrically diodes, such that the core is also part of a pn diode. In practice, referring to Figure 10.9, the core region is intrinsic (nominally undoped) and the upper and lower layers are then p- or n-doped. As a result, current is injected to the core from the doped cladding layers, where the high carrier density is then concentrated, and the lasing process begins.

10.3.2 Fabry-Perot lasers

In this and the following three sections, we briefly consider the basic structures and operating characteristics of the most popular laser diodes, those most relevant for application in micro-optical systems. We provide references to the literature for more in-depth theoretical analysis and the author shamelessly refers the reader to (Zappe, 2004, Chapter 5) for a comprehensive look at laser diode measurement and characterization, particularly with respect to microsystems applications.

The oldest and conceptually simplest laser diode structure is that of the Fabry-Perot laser. The laser cavity is formed by a Fabry-Perot etalon, using the reflecting surfaces which result when cleaving the semiconductor chips as the resonator mirrors, as shown schematically in Figure 10.10. Due to the crystalline structure of the semiconductor from which the laser is fabricated, the cleaved facets are not only of high optical quality, with roughness in the nanometer regime, they are also automatically oriented parallel to each other, and these features come “for free” when the chip is cleaved.

Structure

As we see in Figure 10.10, the Fabry-Perot stripe laser is essentially a channel waveguide, of the type we saw in Chapter 9. The core region is formed by a double-heterostructure, usually populated by one to five quantum wells, such that vertical photon guiding is provided by the material layer structure. Lateral guiding may, as in the example of the figure, be achieved by etching a ridge structure, although a variety of other guiding mechanisms is employed, including buried, implanted and gain-guided structures.

As can be seen from the photograph of Figure 10.11, the Fabry-Perot chip is typically quite compact, making it ideal for use in optical microsystems. The ridge width is typ-

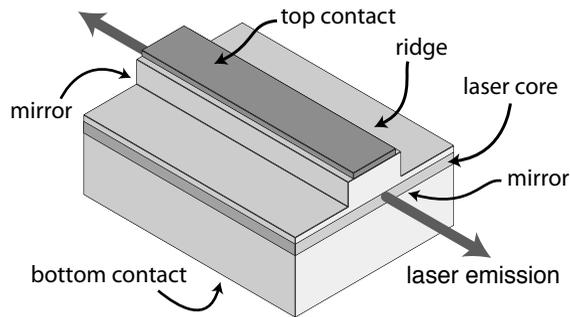


Figure 10.10 The Fabry-Perot ridge laser uses an etched waveguide structure to define the longitudinal laser cavity. The laser core is typically a double-heterostructure and the mirrors are formed by the cleaved semiconductor facets. Emission is typically from both sides of the cavity.

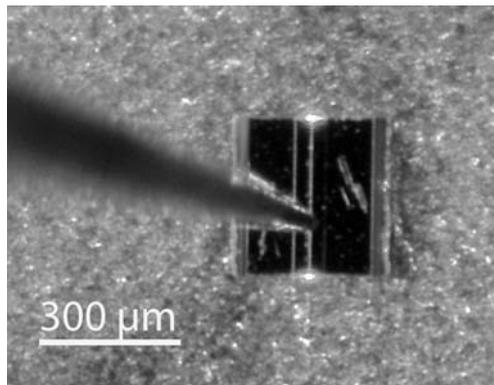


Figure 10.11 Photograph of a Fabry-Perot ridge laser, with a probe needle making electrical contact from the left. The chip emits toward the top and bottom of the figure (some diffracted light is seen from the top and bottom facets) and the ridge is visible as a vertical stripe at the center of the chip.

ically 2 to 5 μm and the cavity length 300 to 500 μm ; unthinned substrate thickness is typically 350 μm .

Since the laser diode is an active optical device, electrical contacts are required for current injection. As is more clearly seen in the schematic of Figure 10.10 than in the photo of Figure 10.11, the upper contact is usually aligned over the waveguide stripe; the second contact is usually large-area metallization on the rear side of the chip. Typical structures usually have the n-doped side of the diode closer to the surface, requiring metallization such as NiGeAu; the p-doped rear surface uses a TiPtAu sandwich.

Operating characteristics

Laser diodes are current-driven devices; due to the exponential relationship between current and voltage in a forward-biased diode, it is more practical to operate them using a current source than a voltage source. Upon current injection, the conditions for lasing are reached at a threshold current, I_{th} [A], after which point the optical output power

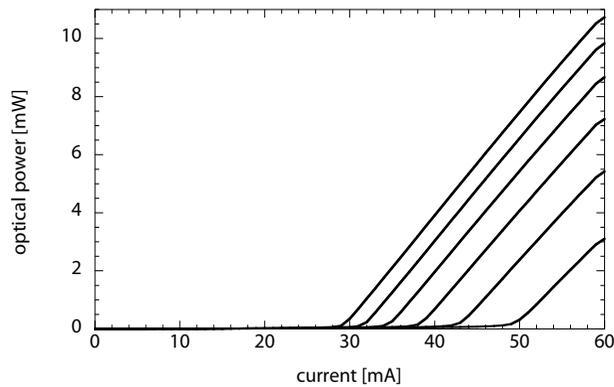


Figure 10.12 The PI (power-current) characteristic of a typical AlGaInP-based 670 nm Fabry-Perot laser diode, as a function of substrate temperature. From left to right, $T = 5, 15, 25, 35, 45, 55$ °C.

increases linearly with increasing current. The typical relationship which results is the PI (for power-current) characteristic, shown for a 670 nm AlGaInP Fabry-Perot laser in Figure 10.12. Although it cannot be seen on the scale of the figure, the transition from spontaneous (LED-like) to laser emission at threshold is accompanied by an abrupt increase in optical power by several orders of magnitude.

As can be seen from Figure 10.12, the threshold current has values of several tens of milliamperes for typical Fabry-Perot lasers, although this value varies strongly with structure, material type, and, as seen in the figure, temperature. Optical emission is from both sides of the laser chip, but usually a high-reflection (HR) coating is deposited on one facet, optimizing the efficiency of the emission from the other facet. Optical output powers vary widely, but are typically in the range of several to perhaps 100 mW, whereby considerable efforts have been expended on the development of high-power Fabry-Perot laser diodes.

The emission spectrum of this laser is seen in Figure 10.13. This AlGaInP diode is specified to emit in the visible red part of the spectrum, around 670 nm, and we see that the spectral peak at 20 °C is around 676 nm. The peak wavelength is a strong function of temperature and drive current. The spectrum shows distinctly multimode emission: there are at least four peaks near the maximum and clear neighboring maxima between 10 and 35 dB below the peak wavelength. This type of spectral behavior is typical for the Fabry-Perot laser: narrow laser ridge widths imply that only one lateral mode is supported, but the close proximity of possible resonant wavelengths to each other (recall the results of Example 10.1) implies that, usually, more than one of these lasers at any given current level. Monomode Fabry-Perot lasers are difficult to realize.

Broad-area lasers

The stripe laser of Figure 10.10 is designed to have one or, at most, a few lateral modes, dependent on the ridge width. The broad-area laser, on the other hand, is a Fabry-Perot structure which dispenses with the stripe structure altogether, and emits over the entire

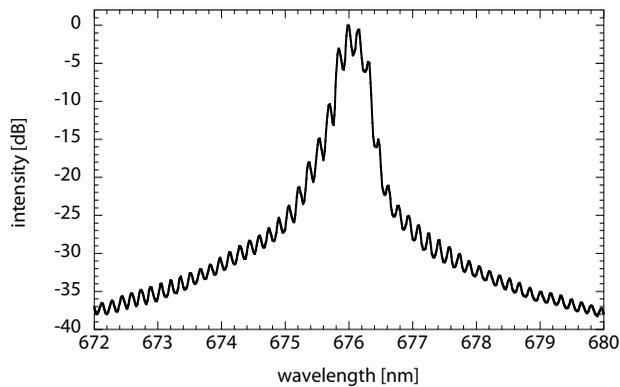


Figure 10.13 The emission spectrum of a AlGaInP-based nominally 670 nm Fabry-Perot laser diode, at 20 °C.

width of the substrate. The several hundred micrometer widths of the emitting regions implies that broad-area lasers have hundreds of lateral modes, but, due to the large emission area, they can emit tens of Watts of optical power from a millimeter-sized chip.

Fabrication of broad-area lasers is very simple, once the heterostructure layers have been grown, such that the chips themselves can be relatively inexpensive. The greater cost is in packaging: the high electrical input powers require good heat-sinking and cooling and the tossed salad of different modes results in a far-field which is difficult to focus or optimize. Broad-area lasers do have a variety of important applications, however, including pumping of solid-state lasers, driving optical amplifiers and in laser printing.

10.3.3 DBR and DFB lasers

The Fabry-Perot lasers we just considered rely on the Fresnel reflection from the cleaved laser facets for feedback into the resonant cavity; the reflectance is broad-band, meaning there is negligible wavelength-dependence of R . In contrast, the family of distributed feedback lasers, and the VCSELs we consider in the next section, use the wavelength-selective Bragg mirrors to which we were introduced in the case study of Section 8.5.1. As we learned there, a periodic corrugation of refractive index in the plane normal to the propagation of an optical field results in strongly wavelength-dependent reflection; the resulting wavelength-dependent mirrors are used to form resonant cavities with a very narrow bandwidth (Morthier and Vankwikelberge, 1997; Carroll et al., 1998).

DBR lasers

The distributed Bragg reflector (DBR) laser was already extensively introduced in Section 8.5.1, when we considered the properties and applications of wavelength-dependent mirrors. As we see again in the schematic of Figure 10.14, the DBR laser consists of

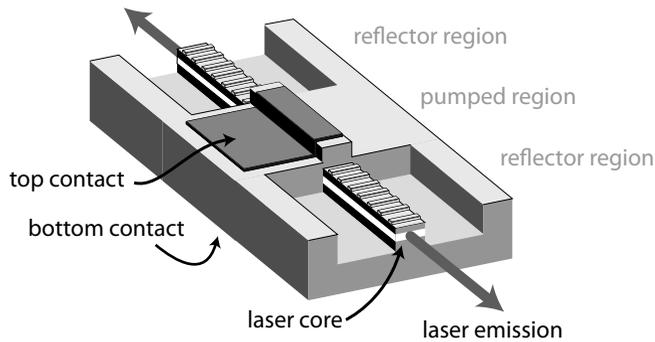


Figure 10.14 The distributed Bragg reflector (DBR) laser uses etched Bragg gratings in separate reflector regions on the laser ridge to generate wavelength-dependent reflection into the active pumped region.

three sections: the active pumped region and two periodic Bragg reflectors on either side.

Based in the discussion of this chapter, we now better understand how the laser is conceived and how it works. The laser core is formed using a heterostructure, and is fabricated using epitaxial layer deposition. The stripe laser is then defined by an etched waveguide structure and the device is separated into the three parts. The pumped region has a metal surface contact (the opposing contact is again on the substrate back side) and the Bragg reflector regions are formed on either side, frequently using a recess as shown in the figure to enhance the coupling of the Bragg grating to the modes generated in the waveguide. Emission is from the cleaved semiconductor facets, which are usually AR coated to avoid parasitic Fresnel reflections into the laser cavity.

The laser waveguide dimensions are similar to those of the Fabry-Perot laser, with a sub-micron thick active layer and core widths from 2 to 5 micrometers. The length of the pumped region is generally in the range 200 – 500 μm and that of the Bragg reflectors 50 – 200 μm . The emission spectrum, as for the DFB laser which we will consider presently, is usually monomode, due to the wavelength-selective reflection of the mirrors, and optical output powers in the tens of milliwatts are available.

DFB lasers

An alternative form of distributed feedback laser, the DFB laser⁵ uses a different arrangement of pumped and grating regions. In the DFB laser, as seen schematically in Figure 10.15, the pumped and grating reflector regions are coalesced into a single structure, such that the entire pumped region consists of a Bragg reflector.

The laser cavity feedback, which occurs continuously along the cavity length, is described by coupled mode theory, as engagingly discussed in (Agrawal and Dutta, 1986, Chapter 7.3). The grating period defines the resonance wavelength, and the cor-

⁵ DFB stands for “distributed feedback”; somewhat confusingly, the family of distributed feedback lasers, using Bragg gratings for reflection in waveguide-based edge-emitters, has the same designation as the particular “DFB laser” we consider in this sub-section.

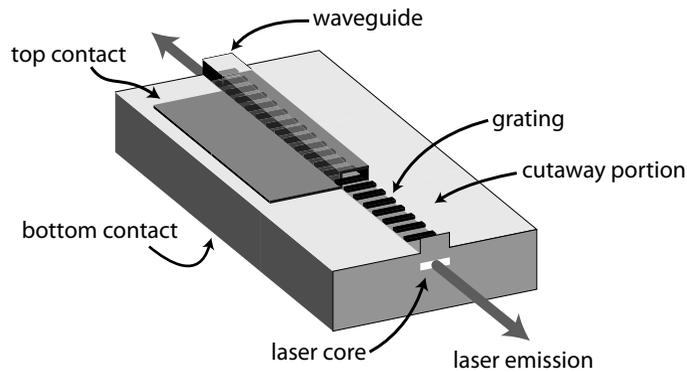


Figure 10.15 In a DFB laser, the pumped and grating regions coincide, so that the Bragg grating is etched onto the laser core (as shown in the cutaway portion of the figure) and then buried under the upper cladding of the waveguide and contact. Feedback is generated continuously in the pumped region.

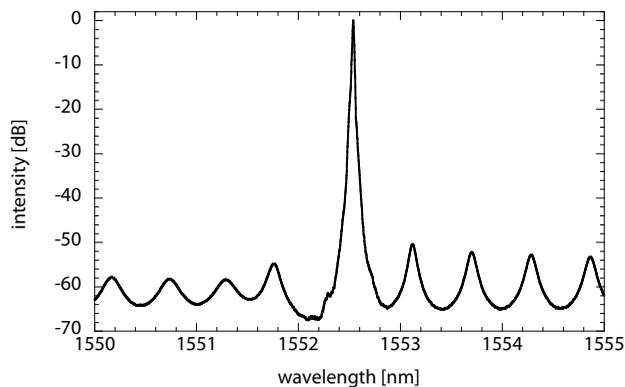


Figure 10.16 The emission spectrum of a InGaAsP-based DFB laser diode designed for the 1.55 μm optical telecommunications wavelength band, operated at 20 $^{\circ}\text{C}$.

rugation depth as well as the spacing from the heterostructure core give the coupling coefficient. Due to the resultant strongly wavelength-dependent feedback, the emission spectrum of the DFB laser is, like for the DBR laser, very narrow-band, with only a single lasing mode, as seen in Figure 10.16. We note that the parasitic resonances are at least five orders of magnitude below the lasing peak, implying a very high side-mode suppression ratio (SMSR) of -50 dB or more. Such a high value of SMSR is required for telecommunications lasers employed in long-haul fiber networks.

DBR and DFB lasers represent the high-end of diode lasers. They have excellent spectral characteristics (monomode emission and narrow linewidth), emit considerable optical power and (in more complex implementations) may be electrically tunable (Coldren, 2000; Coldren et al., 2009). They are, however, complex to manufacture and usually require involved packaging and temperature control. In addition, as edge-emitting lasers (like the Fabry-Perots), the far-field intensity distribution is usually highly astig-

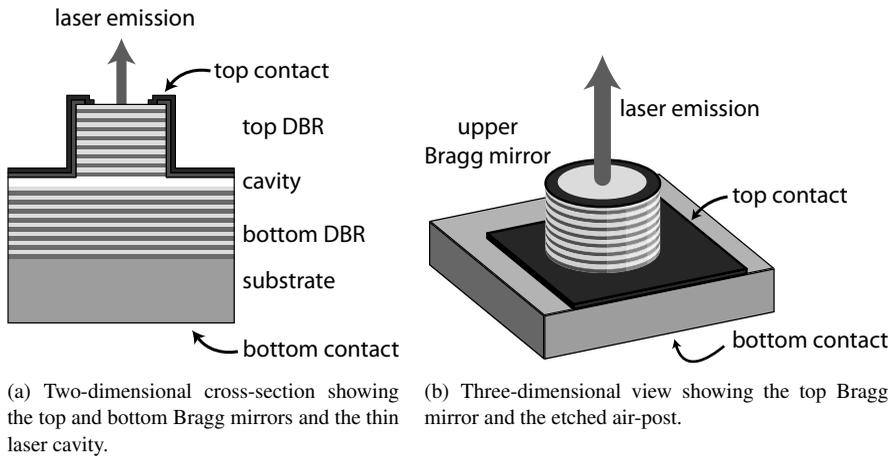


Figure 10.17 Sketches of the structure vertical-cavity surface-emitting lasers (VCSELs); this example shows an etched “air-post” VCSEL, only one of numerous possible types.

matic, requiring anamorphic optics to generate some semblance of a circularly symmetric emission pattern. As a result, DBRs and DFBs are expensive and generally only employed in optical systems which require their advanced spectral properties.

10.3.4 Vertical cavity lasers

As we also saw in Section 8.5.1, the vertical-cavity surface-emitting laser (VCSEL), or vertical cavity laser, also takes advantage of Bragg reflection for generation of a resonant laser cavity. The VCSEL turns the laser cavity onto its side, as seen in the schematic sketches of Figure 10.17 (Wilmsen et al., 2001; Li and Iga, 2003; Koyama, 2006), and emission is from the substrate surface, hence its name⁶.

Structure

In contrast to the edge-emitting laser, the VCSEL is formed of a large number of material layers, usually grown epitaxially using an optically active III-V semiconductor, as seen in the cross-section of Figure 10.17(a). Whereas edge-emitters have pumped regions several hundred micrometers long, the VCSEL has a very short cavity, typically a few half-wavelengths long (Rakic and Majewski, 2003), on the order of 200 nm. Emission from this central pumped layer is reflected by upper and lower Bragg mirrors, formed by layers of periodically-varying refractive index, each layer $\lambda/4$ thick. The VCSEL cavity reflectance characteristic, shown in Figure 10.18, then has a broad

⁶ The author admits to promulgating the less redundant designation VCL (vertical cavity laser) for the vertical-cavity surface-emitting laser in a previous book (Zappe, 2004), but since that acronym is today more widely used for *Virus Creation Laboratory* than for the description of an innocent little photon emitter, he bows to common usage and re-embraces the VCSEL.

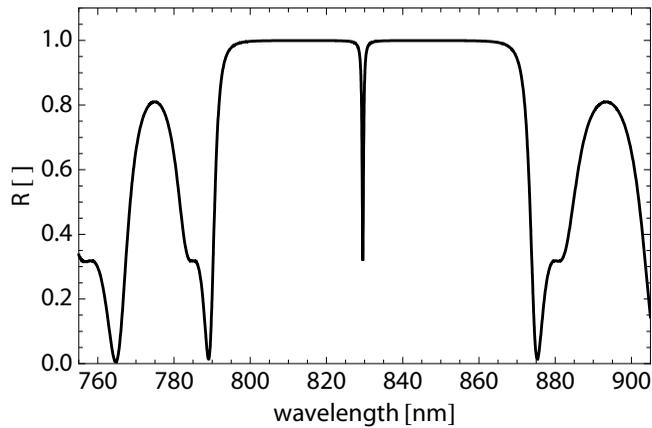


Figure 10.18 The calculated reflectance of a GaAs/AlGaAs VCSEL cavity, designed for emission at 830 nm, with 22 upper and lower Bragg mirror pairs and a cavity of thickness 214 nm.

stop-band, seen as the wide region with $R \approx 1$, due to the Bragg mirrors, and a sharp transmission resonance, defined by the cavity length, and seen as the sharp dip around 830 nm in the figure.

Light emission at a wavelength defined by the active region (usually the thickness of quantum wells in this layer) is then reflected by the Bragg mirrors, such that a good overlap of the cavity resonance and the emission wavelength is required. Since the gain volume of the active region is very small, when compared to an edge-emitting laser, very high mirror reflectance is required to reach lasing threshold; as a result, 20 – 50 mirror pairs are typically needed, leading to values of $R \approx 0.995$ or greater. The necessary high R values imply that only multi-layer reflectors, such as Bragg mirrors, can provide sufficient reflectance to generate a functional VCSEL cavity; simple Fresnel or coated mirrors have insufficient reflectivity.

Waveguide definition

Since the VCSEL cavity is oriented normal to the plane of the substrate, lateral confinement needs to be accomplished using different techniques than the waveguides used for edge-emitting lasers. One simple example is the air-post, shown in Figure 10.17(b). The top mirror is structured into a cylinder using dry or wet etch techniques, such that the emission is guided vertically in the circularly-symmetric structure. Typical post diameters vary from about 5 μm to several tens of micrometers, and their height corresponds to the thickness of the upper mirror, generally 2 to 3 μm . Electrical connection to the pn diode (the upper Bragg mirror is typically fabricated using p-doped semiconductor, the lower with n-doped material) is by one contact directly above the active region and the other on the substrate back-side.

Numerous other approaches have been demonstrated for lateral definition of a VCSEL, including implantation (Morgan et al., 1995), buried heterostructures with epitaxial regrowth (Choquette et al., 1993), buried dielectric aperture structures (Weigl

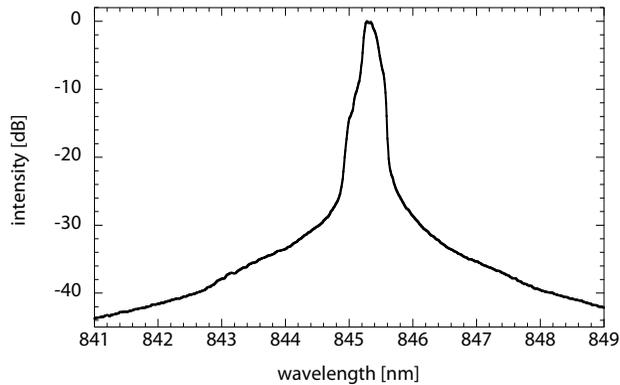


Figure 10.19 The emission spectrum of an 850 nm VCSEL used for short-range data communication, operated at 20 °C. The VCSEL is a single-longitudinal mode laser.

et al., 1997) and anti-resonant reflecting photonic crystals (Liu et al., 2004). These technologies typically result in planar surfaces, and the environmentally-sensitive Bragg mirror layers remain buried; as a result, most commercial VCSELs tend to use one of these latter techniques.

Performance

Due to the very short VCSEL cavity, the spacing between resonant wavelengths is very large and only one wavelength is found within the stop-band of the Bragg mirrors. For that reason, the VCSEL is a fundamentally single-mode laser, with a typical spectral characteristic as that shown in Figure 10.19. The single longitudinal mode is occasionally offset, however, by multiple lateral modes, if the aperture of the VCSEL is large. Even for a small device which supports only a single lateral mode, the circularly-symmetric aperture results in an unspecified polarization direction, such that the polarization of the emitted field may rotate into different directions during operation. This problem has been addressed in a number of ways, including the inclusion of gratings patterned onto the VCSEL surface (Debernardi et al., 2005; Gustavsson et al., 2005; Haglund et al., 2006; Huang et al., 2007; Chung et al., 2008), or the use of other etched features (Furukawa et al., 2004), including photonic crystals (Leisher et al., 2006), as well as the use of plasmonic nanorod arrays (Dayal and Koyama, 2007).

VCSELs are very attractive lasers for use in micro-optical systems, for a number of reasons. Their very small cavity volume (typically on the order of $30 \mu\text{m}^3$) and highly-efficient reflectors implies that the threshold current is very low, such that $I_{th} < 1 \text{ mA}$ is not uncommon. While power emission of a single VCSEL is typically limited to a few milliwatts, the ability to fabricate VCSEL arrays has resulted in high power emission from multiple vertical-cavity devices. Two-dimensional array fabrication also allows the formation of massively-parallel optical systems, and surface emission usually results in greatly simplified packaging when compared to edge emitters. Finally, due to the circularly-symmetric VCSEL aperture, the far-field is also circularly-symmetric, form-

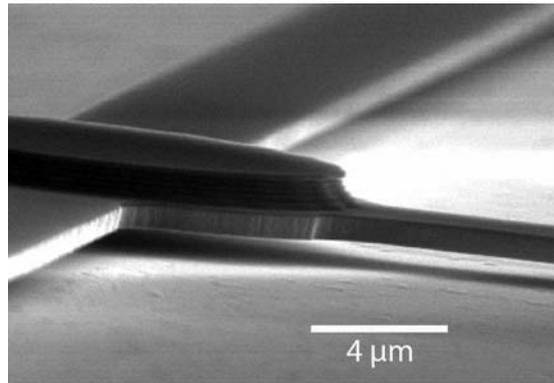


Figure 10.20 SEM photograph of a micromechanically actuated upper Bragg mirror for realization of a tunable VCSEL. Photo courtesy of James Harris, Stanford University.

ing a nice Gaussian beam which can be focussed and transmitted through an optical system using simple micro-optical components.

Tunable VCSELS

The functionality of VCSELS can be increased further by making them tunable. The high temperature-sensitivity of the resonant cavity does make VCSELS easily thermally tunable, but the range is limited. An alternative is to use an external cavity, by replacing the upper Bragg mirror with a micromechanically actuated mirror suspended over an air gap, as may be seen in Figure 10.20 (Sugihwo et al., 1998). Vertical (in the cavity direction) movement of the upper mirror, designed to be of adequately high reflectance, results in a change in the cavity resonance and thus tunes the emission wavelength. Using a GaAs/AlGaAs laser cavity structure with InGaAs quantum wells, a tuning range of 25 nm in the 950 nm wavelength range has been demonstrated with VCSELS of this type. Using an ultra-light-weight nanoelectromechanical upper mirror structure (Huang et al., 2008a), with a high-contrast sub-wavelength grating, the tuning speed can be reduced into the tens of nanoseconds range and tuning ranges of up to 60 nm have been achieved using hybridly-assembled curved Bragg mirrors (Debernardi et al., 2008).

10.3.5 Other laser types

There is quite a variety of other semiconductor laser types, some commercially viable, others the subject of ongoing research. Some of these, such as the micro-external cavity lasers we considered in the diffractive case studies of Section 8.5.1, combine established macroscopic techniques with the miniaturization capabilities of microsystems engineering to allow the development of a useful family of compact lasers suitable for use in micro-optical systems. Others, such as fiber lasers (which are not semiconductor lasers) (Digonnet, 2001) are useful light emitters, integrable into compact optical systems, but not reducible in size beyond a certain degree. We discuss two laser types, with existing and potential relevance, in greater detail here.

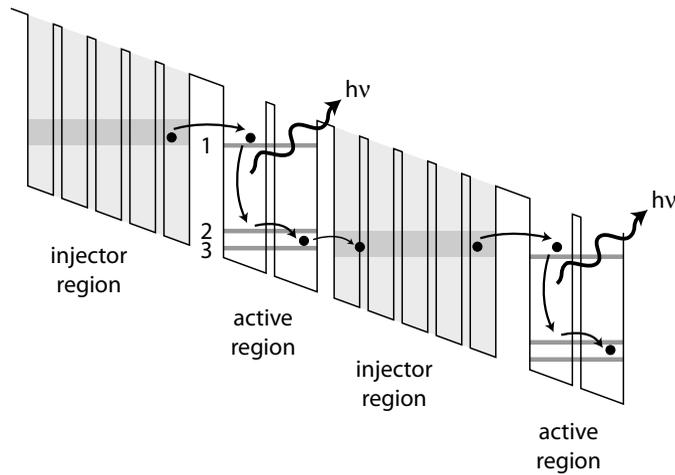


Figure 10.21 Conduction band structure of a quantum cascade laser; two stages (of a total of typically 20 to 75) are shown. The electrons are emitted from a mini-band (the gray bands) into an energy level 1, and the transition to energy level 2 results in the emission of a photon. The electron then thermalizes to energy level 3, is emitted into the next energy band, and the process repeats.

Quantum-cascade lasers

The emission wavelength of a laser, as we saw above, is given by the resonant cavity; the range of wavelengths for which resonance is possible, due to sufficient gain, is, on the other hand, defined by the energy gap of the laser material. As a result, the range of wavelengths attainable for semiconductor lasers is a function of the existence of materials with a suitable bandgap.

The quantum cascade laser (QCL) takes a different approach to defining the gain wavelength, and thus opens new wavelength regimes for semiconductor devices (Gmachl et al., 2001). The QCL uses a complex series of quantum well structures to generate “mini-bands” in the conduction band, from which electrons are injected into an active region whose energy levels are defined only by the quantum well widths there; the structure of the conduction band is shown schematically in Figure 10.21. The electronic transitions from energy level 1 to 2 in the figure results in the emission of a photon, and the energy difference is defined only by the widths of the layers in the active region; these are thus intra-band, not inter-band transitions. The electron transitions repeat by “cascading” down the conduction band, with the emission of a photon at each active region. The QCL thus allows access to gain at long wavelengths, for which no suitable material bandgaps exist, and quantum cascade lasers with wavelengths approaching $\lambda = 100 \mu\text{m}$ have been demonstrated.

The QCL structure is complex; up to 1 000 accurately-defined individual layers need to be epitaxially grown, for 20 – 75 stages, each of which emits a photon; the sketch of Figure 10.21 only shows two stages. The QCL resonator itself is similar to that of a normal edge-emitting laser, and may be a Fabry-Perot or DFB structure (Faist et al., 1997; Hofstetter et al., 1999); an example of the latter is shown in Figure 10.22, in

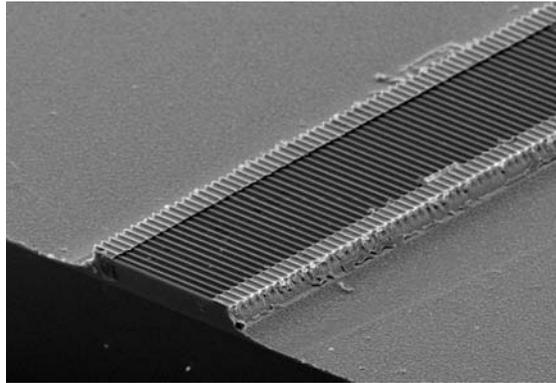


Figure 10.22 An InGaAs-based DFB QCL, showing the surface grating with a period of $1.59\ \mu\text{m}$. The ridge width is $30\ \mu\text{m}$ and the laser emits at $10.16\ \mu\text{m}$. Photo courtesy of Daniel Hofstetter, Université de Neuchâtel.

which the grating structure for an InGaAs-based DFB QCL emitting at $10.15\ \mu\text{m}$ is clearly seen.

Due to the small energy level spacings which may be achieved using the quantum cascade approach, QCLs are particularly attractive for emission at long wavelengths; the accessible wavelengths range from around $3\ \mu\text{m}$, using an InGaAs/AlAsSb/InP structure (Revin et al., 2007) to 3 THz (around $\lambda = 100\ \mu\text{m}$) in the GaAs/AlGaAs system (Williams et al., 2005), although these require operation at low temperature, 164 K. Output powers up to almost 0.5 W and operating temperatures at room temperature and approaching 400 K have been demonstrated (Diehl et al., 2006). The QCL is particularly attractive for use in optical sensors which rely on absorption spectroscopy, which we look at in a little more detail in Section 10.7.2; tunable quantum cascade lasers, which use an external cavity design such as that which we considered in Section 8.5.1, have been developed which show mode-hop-free tuning over a range of 35 nm around $\lambda = 1.95\ \mu\text{m}$ (Wysocki et al., 2005).

Microdisk lasers

The ring resonator which we briefly considered in Section 9.5.4 may also be used as a laser. The microdisk laser consists of a circular microcavity, formed as a disk mounted on a pedestal; the resonator supports so-called “whispering gallery” modes⁷, varying as Bessel functions radially and sinusoidally in the azimuthal direction. If material with optical gain is used, such as InGaAsP on InP, with InGaAs quantum wells, optical pumping can result in laser resonance, which gives rise to optical emission in the plane of the disk (McCall et al., 1992).

⁷ “Whispering gallery” is in analogy to the acoustic phenomenon, found in architectural structures as diverse as St. Paul’s cathedral in London or the temples of Angkor Wat in Cambodia, where one can engage in a whispered conversation with one’s partner across a large elliptical space due to acoustic reflections around the perimeter.

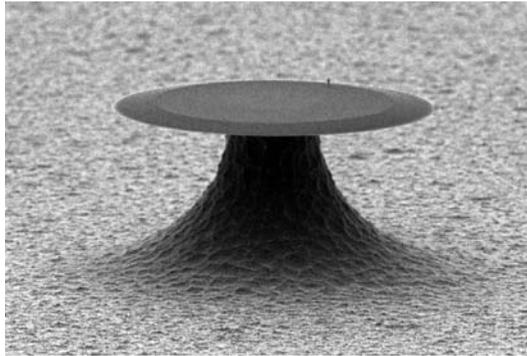


Figure 10.23 An Er-doped SiO_2 microdisk laser fabricated on a Si pedestal; the $1\ \mu\text{m}$ thick disk has a diameter of $60\ \mu\text{m}$ and the laser emits at $1.54\ \mu\text{m}$. Photo courtesy of Kerry Vahala, Caltech.

Typical microdisk laser structures have a diameter of 3 to $60\ \mu\text{m}$ and a thickness of only 50 to 150 nm; a nice example is seen in Figure 10.23. Due to their small volume, the pump power required to reach threshold is quite low, typically less than $100\ \mu\text{W}$. The emission wavelength, as for diode lasers with a standard cavity shape, is a function of the materials and layer structure of the gain region. InGaAsP-based disk lasers emit around $1.5\ \mu\text{m}$, but dye-doped polymers have also been used to generate emission in the yellow-green part of the spectrum, around 565 nm (Kuwata-Gonokami et al., 1995). Alternatively, Er-doped SiO_2 fabricated as a disk on a Si post, the device shown in Figure 10.23, has been shown to emit at a wavelength of $1.54\ \mu\text{m}$ (Kippenberg et al., 2006).

The disk resonator configuration is problematic insofar as the emission takes place along the circumference, more or less (depending on diffraction due to the thin plate) in the plane of the disk; it is thus more difficult to harness the emitted laser light from a disk laser than from a standard edge or surface emitter. A number of approaches to circumvent this difficulty have been proposed and demonstrated, including the fabrication of an InP disk on SiO_2 over a silicon waveguide, into which the light could be coupled evanescently with an efficiency of 35% (Hattori et al., 2006). Also attractive is the possibility of generating vertical emission, which can be realized using a second-order diffraction grating structured along the disk circumference, as has been demonstrated for disk lasers operating in the THz regime (Mahler et al., 2009).

10.3.6 Silicon light emitters

As we noted in Section 10.1.4 above, silicon is an indirect bandgap semiconductor and thus a poor emitter of light. The advanced state of silicon fabrication technology, its prevalence in a wide variety of other electrical and optical applications and its relatively (compared to, for example, III-V or II-VI semiconductors) low price combine to provide considerable impetus in the development of Si-based light emitters. Silicon photonics is well-established (Kimerling, 2000; Zimmermann, 2000; Reed and Knights, 2004), and if nature is uncooperative in the case of light emission from silicon, technology must

be used to help her along; even so, there is as yet no serious competition for compound semiconductor or polymer emitters.

Several techniques have been employed to enhance the optical emission capabilities of Si, either by structuring or combining silicon with optically-active materials. Light emission has been attained from silicon nanostructures; nanometer-sized volumes of material result in quantum confinement of electrons, which leads to a spreading of the energy bands in k -space, and thus the generation of a limited amount of direct-bandgap energy overlap. The first efforts in this direction, using porous silicon, a sponge-like structure fabricated by electrochemical etching (Canham, 1990), demonstrated emission in the orange-red as well as blue-green parts of the visible spectrum under electrical stimulation, but the efficiency is poor (below 0.2%), response is slow and the devices degrade rapidly during operation (Hirschman et al., 1996). More promising has been the use of nanocrystals (Iacona et al., 2000) or nanorods (Hybertsen, 1994), textured bulk silicon (Trupke et al., 2003) or Si/SiO₂ superlattices (Lu et al., 1996). The best performance has been realized through optimized microstructuring of silicon LEDs, enhancing photon emission from the indirect transitions, and leading to efficiencies approaching 1% (Green et al., 2001).

An alternative approach to nano-structuring has been to modify the composition of silicon using materials which are optically active. One example is the IV-IV combination SiGe, an alloy with composition Si_{1-x}Ge_x. In SiGe quantum wells or quantum dots, enhanced optical emission, around 1.3 μm , results from Δk due to scattering, somewhat compensating the indirect bandgap (Tang et al., 1995); efficiency is however still below 0.1%, such that this technology is at present not promising. Considerable effort has also been expended on the co-implantation of erbium and oxygen into silicon (Zheng et al., 1994). Erbium is fluorescent and is most widely used in photonics in erbium-doped fiber amplifiers (EDFAs); when pumped at shorter wavelengths (typically 980 nm or 1.48 μm), Er emits at 1.54 μm . Er is typically ion-implanted into Si, usually co-implanted with oxygen, but the optical emission has an efficiency less than 10^{-4} and decreases rapidly with increasing temperature (Polman et al., 1995). Other types of material modification of Si, including, for example, boron implantation for the generation of dislocation loops leading to optical confinement (Ng et al., 2001), have also led to promising results.

Progress in the demonstration of a silicon-based laser has been made using stimulated Raman scattering to generate optical gain in a silicon waveguide (Rong et al., 2005). In a Raman laser, an optical pump generates optical gain due to Raman scattering at a wavelength slightly (by 15.6 THz) longer than the pump wavelength; using an SOI-based waveguide as a resonant cavity, lasing emission is then generated when the gain exceeds the background loss of the material. Using a pump wavelength of 1.536 μm , lasing emission at 1.670 μm is generated, and longer wavelengths in the near infrared may also be attained using higher order scattering wavelengths (Rong et al., 2008). Whereas this is promising work, the silicon Raman laser still requires pumping, at 200 to 300 mW, with an external pump laser, to generate an optical output power of 2 to 5 mW and as such does not yet represent a micro-optically relevant light source.

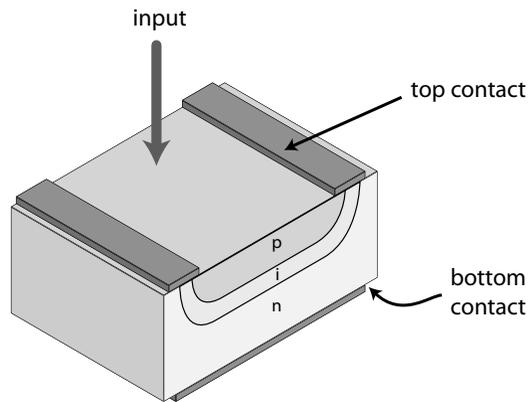


Figure 10.24 A photodetector configured for surface incidence of the optical field. Using an absorbing semiconductor, the reverse-biased p-i-n diode separates the generated electrons and holes, which then give rise to a current at the contacts. Not to scale.

10.4 Photodetectors

As important as generation of photons in a micro-optical system is their detection. The function of a photodetector is to convert optical radiation into an electrical current, with the greatest possible efficiency. Since the evaluation of the output of an optical microsystem is most often done electrically, the photodetector usually represents the final element in an optical setup.

As for the other components we consider in this chapter, there is a wide variety of detectors, using many different physical principles (Dereniak and Boreman, 1996; Lutz, 1999; Rogalski, 2000). We will focus on those relevant for integration with small-scale optical systems.

10.4.1 Photodiodes

Absorption of photons in a material leads to the generation of free carriers, electron-hole pairs. To generate an output current, these carriers must be separated so that they do not recombine, and then extracted from the contacts. Carrier separation is efficiently realized by a reverse-biased pn diode, in which the electric field pulls the electrons and holes to opposite sides of the junction. As a result, the most popular configuration for a semiconductor-based photodetector is the pn junction diode, as shown schematically in Figure 10.24.

As shown in the figure, light is incident onto the surface, under which a shallow pn or p-i-n junction is diffused; the junction depth is usually less than 1 μm . The p-i-n junction incorporates an intrinsic region between the p and n doped layers, in which a reverse bias generates a large depletion region and high electric field. Photons are efficiently absorbed in the depletion layer, and separated by the field. A relatively thick intrinsic layer in the p-i-n junction thus provides a thick absorption layer and allows a

high electric field, ensuring that an excess of unabsorbed photons does not penetrate to the substrate and thus not contribute to detector current.

Materials

A wide range of semiconductors is used for the realization of photodetectors. The most relevant characteristic is the optical absorption at the wavelength of interest. As we saw in Figure 3.10 of Chapter 3, there are numerous materials of interest in the visible and near-IR wavelength ranges. Silicon is popular for wavelengths shorter than 1.1 μm , due to its relatively low cost and highly-developed fabrication technology, as well as high purity. GaAs may be employed for wavelengths shorter than 850 nm and Ge and InGaAs are the most popular materials for near-infrared wavelengths. For the mid-IR and longer wavelengths, more exotic materials come into play, frequently II-VI compound semiconductors. At long wavelengths, the detectors usually need to be cooled to reduce the background noise, enhanced due to the small bandgap. We discuss organic and quantum dot materials below.

Structures

The photodiode of Figure 10.24 is one example of a simple pn diode structure; a large surface area, with sizes ranging from square millimeters to centimeters, leads to reduced alignment tolerances for free-space optical systems and substrates are often thinned to reduce the recombination of carriers before they reach the contacts. Variations include the use of heterostructure pn diodes for III-V or II-VI materials, leading to carrier confinement and a larger absorption region. By the use of a high applied reverse bias on a thick intrinsic region in the p-i-n diode, the resulting high electric field can lead to avalanche carrier multiplication by impact ionization, and thus multiplication of the generated carriers by factors of 100 or higher. Avalanche photodiodes are attractive for high-sensitivity applications in which every photon counts (Wang et al., 2002) but require high operating voltages.

An alternative to the junction pn diode is the Schottky diode⁸, formed by the deposition of an appropriate metal onto certain semiconductor surfaces (Rhoderick and Williams, 1988). The MSM (metal-semiconductor-metal) photodetector uses an alternating finger arrangement of metal stripes on a semiconductor surface, with open spaces to allow photon absorption. MSM detectors have very low electrical capacitance, and are thus commonly used for detection of high-speed optical signals due to their fast response time.

Photodiodes vary widely in size and structure, but are available as very compact chips, analogously to laser diodes; they are found in numerous types of packages, including surface-mount device (SMD) form, useful for incorporation into micro-optical systems. Also useful are linear photodiode arrays, with closely spaced detectors, which may also be configured as long photodiode structures used as position-

⁸ Walter Schottky (1886–1976), born in Zurich, was a German professor of physics who spent most of his career at Siemens Research Laboratories. He worked extensively on understanding the emission of electrons from surfaces and first explained the potential barrier formed between metals and semiconductors.

sensitive detectors (PSDs). Finally, the basis for most modern imaging technology is the two-dimensional photodetector arrangement of the CCD, which we consider in Section 10.4.4.

Efficiency

As mentioned above, the photodiode is operated with a reverse bias applied to the pn junction; the output signal is the generated reverse-bias current, which increases from its “dark” value (resulting from thermal generation of electron-hole pairs) due to photon absorption and generation of excess carriers. Two of the important photodiode characteristics are efficiency and noise: the former defines how many electrons are generated per incident photon and the latter the effect of non-optically generated carriers and the stochastic time-variation of the signal.

For an input optical signal with power P_{opt} [W], consisting of photons of frequency ν [Hz], the number of photons per unit time is given by

$$N_{photons} = \frac{P_{opt}}{h\nu}; \quad (10.21)$$

likewise, a generated detector signal current I_s [A] consists of

$$N_{electrons} = \frac{I_s}{q} \quad (10.22)$$

electrons per unit time, for fundamental electronic charge q [C]. The efficiency of a photodetector, η [], can then be defined as

$$\eta = \frac{N_{electrons}}{N_{photons}} = \frac{h\nu}{q} \frac{I_s}{P_{opt}}. \quad (10.23)$$

Alternatively, the detector may be characterized by the responsivity R [A/W], defined as

$$R = \frac{I_s}{P_{opt}} = \eta \frac{q}{h\nu}, \quad (10.24)$$

and thus related to efficiency. Responsivity has the advantage that it can be directly determined from the slope of the output current versus input optical power characteristic.

Careful consideration of Equation 10.23 reveals that η is *not* a function of photon energy or frequency, since optical power is a function of both photon number and photon energy. We recall that a photon with energy greater than E_g will generate an electron, so that the number of electrons will not increase as photon energy increases even further. R , on the other hand, *is* wavelength (or energy) dependent, decreasing as photon energy decreases.

The factors which limit η are loss of photons or loss of electrons: the former may be lost due to reflection at the detector surface, incomplete optical absorption or absorption far from the pn junction; the latter may be lost by recombination in the bulk, at the surface or at defects in the material. High efficiency photodiodes require good optical design and materials of high purity.

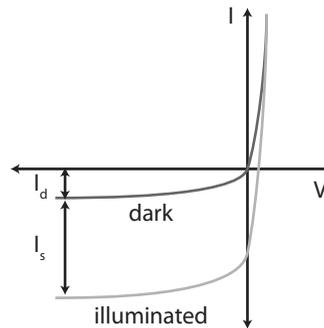


Figure 10.25 A schematic representation of the typical photodetector current/voltage characteristic; the signal is given by the reverse-bias current under illumination, I_s . The dark current, I_d , is that which flows in the absence of an illumination signal. Not to scale.

Example 10.3: Let us consider a detector with a linear output characteristic for which we measure an output current of 300 nA for an optical input power of 1 μ W at a wavelength of 780 nm. What are the responsivity and efficiency, and how would these change if the same output current had been measured using a different detector with an input wavelength of 1.55 μ m instead?

Using Equation 10.24, we see that responsivity is easily determined from the ratio of I_s to P_{opt} to be $R = 0.3$ and, from Equation 10.23, the efficiency is $\eta = 0.48$, at $\lambda = 850$ nm. Had this been a 1.55 μ m detector (made from a different material), the responsivity would be the same (since we still measure the same current for the same optical input power) but the efficiency of this detector would $\eta = 0.24$. Since the longer wavelength has lower-energy photons which yield the same output current as for the shorter wavelength case, the necessary efficiency to obtain the same current is lower for that case.

Noise

Photodetector noise is due to a number of factors, and provides a lower sensitivity limit for the detector. A fundamental limitation in the performance of photodiodes is due to dark current, the reverse-bias current which flows in the absence of an optical signal, due to thermal generation of electron-hole pairs, as shown in Figure 10.25. High-quality material with few defects usually results in low dark current, but this parameter increases in relevance as the bandgap energy, E_g , decreases; low- E_g materials are required for long-wavelength detectors, such that these have considerably higher noise due to dark current than high- E_g semiconductors. Since thermal generation of carriers is strongly temperature-dependent, long wavelength detectors (using Ge, InGaAs or II-VI semiconductors such as HgCdTe) are usually cooled to 77 K for low-noise operation.

The stochastic nature of photon incidence (since they are discrete entities) leads to shot noise, resulting in a fluctuation of the detector signal around a mean value. Shot

noise is a function of the bandwidth of the detector, BW [Hz], a measure of the highest frequencies detectable: a higher bandwidth implies that a larger range of frequencies is available to contribute to shot noise. As a result, the mean square value⁹ of the detector current may be expressed as

$$\langle I_s \rangle = 2q^2 \frac{\eta}{h\nu} BW P_{opt}; \quad (10.25)$$

if we consider the effect of the dark current, which generates a dark power P_d [W] in the detector, this last expression can be modified to yield

$$\langle I_s \rangle = 2q^2 \frac{\eta}{h\nu} BW (P_{opt} + P_d). \quad (10.26)$$

Noise performance may be practically characterized by the signal-to-noise ratio, S/N [], a measure for the ratio of signal strength to the detector noise level, and given by

$$S/N = \frac{\eta}{2h\nu} \frac{P_{opt}}{BW}; \quad (10.27)$$

again including the effect of dark current results in an alternative expression, namely

$$S/N = \frac{\eta}{2h\nu} \frac{1}{BW} \frac{P_{opt}^2}{P_{opt} + P_d}. \quad (10.28)$$

A high S/N implies that the desired signal is significantly larger than the background noise, a factor which is inversely related to bandwidth. A large S/N can be achieved by assuring good efficiency and low dark current.

A further useful parameter for characterizing photodetector noise performance is the noise-equivalent-power, NEP [W], defined as that input optical power at which $S/N = 1$. NEP is found from

$$NEP = \sqrt{2BW \frac{h\nu}{\eta} P_{opt}}, \quad (10.29)$$

for the case where signal power is much smaller than dark power. NEP essentially defines the minimum optical power detectable before the signal is lost in the noise.

Example 10.4: Given a photodiode with 40 MHz bandwidth, a measured dark current at the operating temperature of 1.5 μA , and an internal resistance of 8 Ω , what are the S/N and NEP for optical input powers of 1 μW and 0.1 nW?

From the dark current and internal resistance of the photodiode, we find that the dark power is about 27 pW. From Equation 10.28, we can then find the S/N to be about 100 dB for the higher power level, but only about 6 dB for the 0.1 nA signal. If we ignore the dark current, considering only shot noise leads to $S/N \approx 8$ dB for the lower power level (and leaves the S/N for the higher power virtually unchanged), so the dark current becomes clearly more relevant for lower input powers, as we expect.

⁹ Recall that the mean square value of a parameter ζ is given by $\langle \Delta\zeta^2 \rangle = \langle (\zeta - \langle \zeta \rangle)^2 \rangle$.

The NEP is found simply from Equation 10.29 and evaluates to about 34 pW, such that the 0.1 nA signal is still considerably above this. A S/N of 6 dB is not particularly high, however.

10.4.2 Organic detectors

We saw in Section 10.2.4 that OLEDs take advantage of the light emitting properties of organic materials, and that polymers are playing an increasingly significant role in the conception of new types of light emitting devices. As we might imagine from our knowledge of the optical behavior of materials, good optical emitters are likely to be good optical detectors; as a result, organic materials are becoming increasingly relevant for detection as well.

Materials and fabrication

The organic materials employed for photodetection are organic semiconductors, and can be roughly divided into two categories (Forrest, 2004). Molecular or “small molecule” materials are essentially monomers, which can be deposited as single-crystalline thin films; they are usually evaporated in vacuum, and can be used to make complex multi-layer sequences. Lateral structuring is accomplished using shadow masking.

Alternatively, longer chains of these monomers form polymer materials, which are most practically deposited on substrates using ink-jet techniques; lateral patterning of large two-dimensional arrays of these polymers is thus straightforward, and the approach is also extensively used for the fabrication of OLED arrays. A further approach useful for low-cost deposition and lateral patterning is through stamping or microcontact printing (Xu et al., 2009), which also allows definition on curved and flexible surfaces.

Photodetection using organic materials relies on the generation of electron/hole pairs, usually through the decay of an optically-excited exciton (Forrest, 2000). An organic photodetector usually consists of alternating layers of donor-like and acceptor-like materials; examples are CuPc (copper phthalocyanine) and PTCBI (3,4,9,10 perylene-tetracarboxylic bix-benzimidazole), respectively. 64 layers, each about 0.5 nm thick, of these materials are typically deposited on ITO, using Ag as a cathode, leading to an external quantum efficiency of 0.75 for the wavelength range between 500 and 700 nm, with a bandwidth of 0.5 GHz. Since the monomers can be easily deposited in multi-layer structures, heterostructure organic photodetectors for red wavelengths ($\lambda = 650$ nm) have also been demonstrated (Morimune et al., 2006).

Organic photodetectors have the advantage that they are potentially low cost and can be deposited, using liquid processing techniques, on different types of substrates or formed into flexible sheets and combined with polymer-based electronics; it is likely that the great flexibility in the fabrication techniques available for organic optoelectronics will lead to its successful implementation in a wide range of applications. An additional area in which these materials are applied is for solar cells, for which stacked heterostructure layers have shown efficiencies in the range of 2 to 3%. The demands of

a solar cell are somewhat different than those for a photodetector, since, whereas the latter needs to generate an output signal, usually a current, the solar cell needs to generate output power, and thus both significant current and voltage.

10.4.3 Quantum dot detectors

As for LEDs, quantum dots are poised to play in an important role in the development of new types of photodetector structures. Recalling that the absorption wavelengths of quantum dots are defined by their size, a wide range of wavelengths is accessible using dots of varying diameter. Since they may be deposited in liquid suspension, the materials may be combined with other technologies, particularly silicon-based backplanes with integrated amplifiers and readout circuitry, enabling great flexibility in the conception of new types of two-dimensional detectors and imagers. Using PbS quantum dots, for example, an imager for the near-IR range (up to 1.8 μm) integrated with a silicon active-matrix thin-film transistor (TFT) array has been demonstrated (Rauch et al., 2009), usable at wavelengths much longer than the bandgap of silicon. Quantum dot-based detector technology is also promising for use in photovoltaics (Sargent, 2009), with efficiencies approaching those of organic polymer materials.

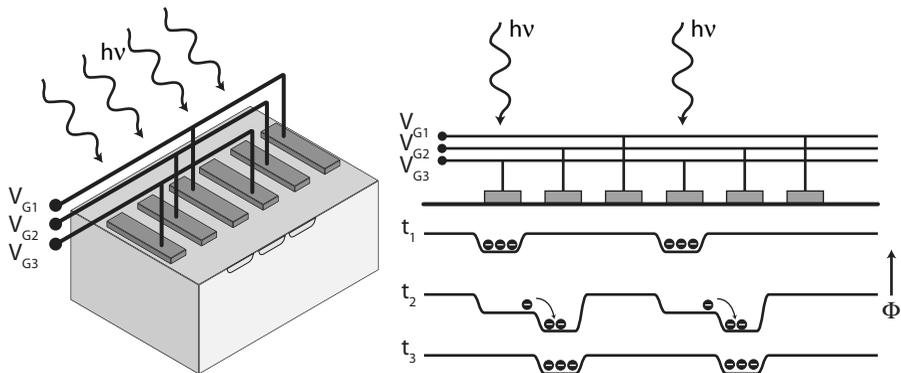
10.4.4 Photodetector arrays

From the commercial point of view, the most important photodetector products are large two-dimensional detector arrays employed for electronic imaging. Digital photography, which has all but replaced its film antecedent, completely relies on electronic imaging, as do a myriad of technical and scientific applications. Detector arrays are important tools in the optics laboratory, and also incorporate micro-optical components (in particular, lens arrays) in their structure.

Two primary technologies are used for the vast majority of one and two-dimensional photodetector imagers, the charge-coupled device (CCD) and the active pixel sensor (APS) (Abe, 2004). Both rely on the generation of electrons in silicon by incident photons and the output of electrical signals which correspond to the intensity at a particular position. Each photodetector corresponds to an individual pixel and two-dimensional arrays of several tens of millions of pixels are commercially available. Non-silicon based imagers are also available, particular those fabricated using III-V or II-VI semiconductors, but are relatively exotic and sinfully expensive.

The CCD

The CCD was developed in the 1970s and may be considered to be a form of “opto-electronic shift register”, as seen in Figure 10.26. This imager consists of one or two-dimensional arrays of metal-oxide-semiconductor (MOS) capacitors fabricated in silicon, as shown schematically in Figure 10.26(a); incident photons generate electrons in the silicon substrate, and a voltage bias on the capacitor gate (typically polysilicon over SiO_2 isolation) creates a potential well in which the generated electrons collect. Each



(a) Schematic representation of a one-dimensional CCD; the polysilicon gates on the surface generate potential minima in which photo-generated electrons are collected.

(b) Charge transfer in a CCD, cross-sectional view: the potentials on the gates are varied as shown at times t_1 , t_2 and t_3 to transfer the charge packets from left to right.

Figure 10.26 The CCD is an array of photodetectors which relies on the generation of charges due to photon absorption in silicon under an array of MOS capacitors.

capacitor represents an individual pixel, with areas down to $6 \mu\text{m}^2$ and arrays of 26 million pixels or more are available.

Once the carriers have been generated under the gate of a pixel, a sequence of voltage pulses on neighboring gates is used to transfer the charge out of the array. This process is shown in Figure 10.26(b), in which a cross-section through a CCD is sketched along with the voltages applied to the gates, V_{G1} – V_{G3} , at three times, t_1 – t_3 . At t_1 , the potential results in collection of carriers under the first and fourth gates. At t_2 , a larger potential on the second and fifth gates results in the transfer of the electrons to these positions. At t_3 , the charge is then stable under the second and fifth gates. The process continues, transferring the charge to the right to a charge amplifier, which then provides an output voltage. Typically three phases of voltages applied, with values below 1.8 V required (Tanaka et al., 1997).

One and two dimensional CCD arrays are commercially available, the former often used for scanners and fax machines, and the latter for imaging, especially in cameras. The external quantum efficiency of the CCD is high, typically > 0.7 , and Si-based devices are sensitive to wavelengths ranging from the UV to the bandgap of silicon ($1.1 \mu\text{m}$). Thermal noise (dark charge, analogous to dark current) may be reduced by cooling and very large ($30 \times 30 \mu\text{m}$) pixel CCD devices fabricated using standard CMOS technology have been developed, primarily for use in medical imaging (Rao et al., 2008).

Color CCDs are realized by the use of color filters placed over the individual pixels. Popular for this application is the Bayer filter¹⁰, which consists of a square mosaic of four filters: one red, one blue and two green filters are arranged in a square array, where two green filters are used since the human eye is most sensitive at this wavelength. In addition to the color filters, microlens arrays are placed over the pixels to concentrate the optical field onto the sensitive area of the pixel, compensating for the fact that the fill-factor for CCDs is considerably less than unity (Hojo et al., 1991). In general, CCDs feature a high dynamic range and good uniformity, and tend to be applied in high-end imaging systems.

The APS

The active pixel sensor is a more recent development than the CCD, reaching a suitable level of refinement in the 1990s (Fossum, 1997). Although the two types of array sensors are frequently confused, the APS is not a CCD; rather, each pixel of the APS consists of a MOS capacitor sensor and an amplifier circuit consisting of three to four MOSFET transistors. The APS is most often fabricated using CMOS technology, thus benefitting from advanced silicon process developments. As a result, each pixel and the image as a whole is subject to signal processing on-chip, such that the APS closely represents a “camera-on-a-chip”. Virtually all low-cost digital imaging, from webcams to digital cameras, including much high-end photographic equipment, is based on CMOS APS chips.

The structure of each pixel in the APS consists of a photodiode and transistors for selection, readout and reset of the pixel charge data. As a result, this two-dimensional array is a form of optical RAM, since each pixel can be individually addressed by the control circuitry; electronic windowing of the image is thus straightforward. Due to the circuitry required for each pixel, the design requires a tradeoff between circuit size and imaging area, and the optical aperture is at best 80%, but in some designs less than 30%, of the available area. For that reason, microlens arrays over the pixels are again used to focus a maximum of the available light onto the optically active area.

APS chips are also available as one and two dimensional arrays, the former often used as high-resolution scanning camera backs for advanced photographic systems; the image is scanned in one dimension with a linear array of 7 000 or more pixels, resulting in an image of 50 to 100 million pixels, but not requiring a two dimensional array of that size. The APS also features high response speed, since most or all of the circuitry is on-chip. Both the CCD and the APS have their market niches and, despite the commercial success of CMOS imaging, CCDs are still to be preferred for many demanding imaging applications.

¹⁰ The Bayer filter is named after its inventor, Bryce Bayer of Eastman Kodak, who patented this filter configuration in 1976.

10.5 Phase and intensity modulators

The function of most optical elements, micro or macro, is to somehow modulate a propagating optical field. Modulation implies a change in phase, intensity, propagation direction, wavefront shape, beam width, or other parameter. We have seen many forms of modulation implicitly: the effect of lenses on an optical beam was the subject of Chapter 7, modulation of propagation direction was seen for scanning micromirrors in Section 6.4.2 and wavefront modulation was the subject of Section 6.5. There are numerous other approaches for achieving various forms of optical modulation; see for example (Almeida et al., 2004) or (Solgaard, 2009, Chapters 7 – 11).

When we allude to an optical modulator, however, we generally refer to a component which affects a controlled change in the phase or intensity of an optical field. Phase modulation can be used, for example, to optimize interference between optical fields and intensity modulation may be employed for changing the intensity of laser emission, for example at GHz speeds for optical telecommunications systems. We thus look at the structure and function of the micro-optical modulators.

10.5.1 Modulation in general

We recall from Section 3.1.1 that the propagation constant of an optical wave through a material is in general complex, implying that, as in Equation 3.32, the electric field varies with propagation distance z as

$$E = E_0 e^{jk_R z} e^{-k_I z}. \quad (10.30)$$

Using the notation we employed in the discussion of laser resonance in Section 10.1.3 above, namely the real propagation constant

$$k_R = \beta = \frac{2\pi}{\lambda} = \frac{2\pi n}{\lambda_0}, \quad (10.31)$$

and the imaginary part of the refractive index, or absorption coefficient,

$$k_I = \frac{\alpha}{2}, \quad (10.32)$$

we may use Equation 10.30 to write the intensity variation with propagation as

$$I(z) = I_0 \exp \left[j \frac{4\pi}{\lambda_0} n z \right] \exp [-\alpha z]. \quad (10.33)$$

If we propagate through a material of length L whose real and imaginary refractive indices can be changed from their original values n_0 and α_0 by an amount Δn and $\Delta\alpha$, respectively, the previous expression becomes

$$I(z) = I_0 \exp \left[j \frac{4\pi}{\lambda_0} (n_0 + \Delta n) L \right] \exp [-(\alpha_0 + \Delta\alpha) L]. \quad (10.34)$$

The function of a modulator is to generate a desired value of Δn or $\Delta\alpha$ in the optical path of a system. A phase modulator generates a controlled change of Δn , and results

in an induced phase shift of

$$\Delta\phi = \frac{4\pi L}{\lambda_0} \Delta n \quad (10.35)$$

and an intensity modulator changes the intensity of a field through an induced $\Delta\alpha$, resulting in a ratio of output to input intensities given by

$$\frac{I_{out}}{I_{in}} = e^{-\Delta\alpha L}. \quad (10.36)$$

For most applications, purely phase or intensity modulation, but generally not a mixture of both, is desired.

10.5.2 Modulation mechanisms

The induced changes in real (Mias and Camon, 2008a) or imaginary (Li and Yu, 2003) refractive index, Δn or $\Delta\alpha$, may be accomplished using any one of a variety of physical mechanisms. The structure of a modulator is typically designed so that the optical wave propagates a known distance through a modulator material, and an external stimulus, often an electric field, is used to generate the desired Δn or $\Delta\alpha$. Many of these effects are not only a function of the materials used, but also often of wavelength, temperature, polarization and other factors.

Electro-optic effects

Electro-optical effects are based on the non-linear behavior of certain optically-active materials, which we examined in Section 3.2.5 and we discuss in detail in Section 10.5.3 below. In Chapter 3, we were introduced to the Pockels and Kerr effects (Narasimhamurty, 1981; Eimerl, 1987), for which we saw that the application of an electric field to an electro-optic material results in change of the susceptibility. We can relate this change to that of the real refractive index, leading to electrorefraction.

For the Pockels, or linear, electro-optic effect, we obtain a shift of the form

$$\Delta n = \frac{1}{2} n_0^3 r \mathbf{E} \quad (10.37)$$

where n_0 represents the real refractive index in the absence of an applied field and r [m/V] is the linear electro-optic, or Pockels, coefficient¹¹. We have written the applied electric field \mathbf{E} explicitly in vector form to emphasize that the Pockels effect is in general a function of the optical field polarization direction with respect to the non-linear material, typically a crystal; in addition, certain crystalline directions can lead to a change in the sign of r . Typical values for GaAs, for example, are $r = -1.44 \cdot 10^{-12}$ m/V.

¹¹ The Pockels and Kerr coefficients are often written with subscripts, such as r_{41} or s_{33} , resulting from the fact that their values derive from a particular element of the 6×3 electro-optic tensor, which relates the electric field vector to the direction-dependent refractive index tensor; see for example (Iizuka, 2002a, Chapter 5.2).

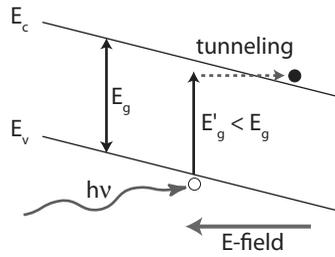


Figure 10.27 The Franz-Keldysh effect relies on a tilting of the energy bands under application of an electric field and the subsequent tunneling of electrons with energy $E < E_g$ into the conduction band, thereby reducing the effective bandgap to E'_g .

The Kerr effect, also referred to as the quadratic electro-optic effect, is in turn related to the applied field as

$$\Delta n = \frac{1}{2} n_0^3 s \mathbf{E}^2 \quad (10.38)$$

where s [m^2/V^2] defines the Kerr coefficient. For the quadratic electro-optic effect, the factor s does not vary with crystalline orientation, but is a strong function of wavelength. Again considering GaAs, for example, $s = -5.4 \cdot 10^{-20} \text{ m}^2/\text{V}^2$ at $\lambda = 1090 \text{ nm}$, changing to $s = -4.0 \cdot 10^{-20} \text{ m}^2/\text{V}^2$ at $\lambda = 1150 \text{ nm}$.

Franz-Keldysh effect

For an electrically-induced change in imaginary refractive index, electroabsorption, the Franz-Keldysh effect¹² is of utility. When a high electric field E (typically 10^4 V/cm or greater) is applied to a semiconductor, the resultant tilting of the energy bands gives rise to enhanced electron tunneling between valence and conduction bands and thus increased absorption, also for photon energies below E_g ; the effect is shown schematically in Figure 10.27. The effective shift of the bandgap which results, ΔE_g , is given by

$$\Delta E_g = -\frac{3}{2} \left(\frac{q\hbar E}{\sqrt{m^*}} \right)^{2/3} \quad (10.39)$$

for electron effective mass m^* [kg]. For an incident optical field with a wavelength slightly longer than the bandgap wavelength, the Franz-Keldysh effect can then give rise to an increase in absorption by several orders of magnitude upon application of the electric field to the material, forming the basis for an electroabsorption modulator.

The strong change in absorption leads to a concomitant change in real refractive index, such that electroabsorption is always accompanied by electrorefraction; recall the Kramers-Kronig relation of Section 3.1.1 in which we saw this explicitly. The con-

¹² Walter Franz (1911 – 1992), a German theoretical physicist, and Leonid Veniaminovich Keldysh (b. 1913), a Russian theoretician, appear to have independently discovered the effect now named after them in 1958 (Franz, 1958; Keldysh, 1958). Keldysh has made a wide variety of contributions to solid-state physics, including having first proposed the idea of the semiconductor superlattice.

comitant change of Δn with an induced $\Delta\alpha$ leads to chirp in a modulator; we again use the term “chirp” as we did for the description of a chirped grating, since Δn in the modulator gives rise to a shift in the wavelength in the material. As a result, the Franz-Keldysh effect also gives rise to a shift Δn proportional to the square of the applied electric field,

$$\Delta n = F_{FK} E^2, \quad (10.40)$$

where F_{FK} [m^2/V^2] is the Franz-Keldysh coefficient. F_{FK} is again strongly wavelength-dependent and is dichroic, implying a dependence on polarization; a typical value for GaAs would be $F_{FK} = 7.9 \cdot 10^{-20} \text{ m}^2/\text{V}^2$ at $\lambda = 1.55 \mu\text{m}$

Other mechanisms

There is a variety of other mechanisms for achieving optical modulation. In semiconductors, a high electron density leads to a shift in the real and imaginary refractive indices, such that the resultant so-called plasma effect may be used in non-electro-optically active materials such as silicon (Soref and Lorenzo, 1986). Relatively high carrier densities are required, since $n = 10^{18} \text{ cm}^{-3}$ leads to $\Delta n = 9 \cdot 10^{-4}$ in Si, but interferometer-based silicon modulators operating at 40 Gbit/s have been demonstrated using this mechanism (Liao et al., 2007).

Thermo-optic effects can also be used for modulation using materials in which the electro-optic effect is small or in-existent. In single-crystalline silicon, the change in refractive index with temperature, $\Delta n/\Delta T$, is about $2 \cdot 10^{-4} \text{ K}^{-1}$ and for amorphous silicon about $3.25 \cdot 10^{-4} \text{ K}^{-1}$. For thermo-optic modulation, relatively high electrical powers have to be applied to metallic thin-film heaters on bulk semiconductor substrates, since long path lengths are required to achieve a significant Δn ; the response also tends to be slow.

The use of quantum wells in a semiconductor heterostructure can lead to highly effective electroabsorption by means of the quantum-confined Stark effect (QCSE)¹³ (Miller et al., 1984; Kuo et al., 2006). For wavelengths close to the bandgap, an electric field applied normal to the quantum wells tilts the energy bands and leads to a reduction of the effective bandgap energy; thus a strong change in absorption with applied bias is seen, along with considerable electrorefraction, and thus chirp. The quantum effects associated with the wells thus significantly enhance electrorefraction and electroabsorption (Zappe, 1995, Chapter 11.4), such that considerably lower voltages are required to achieve a certain Δn or $\Delta\alpha$. In addition, quantum confinement effects can be combined with, for example, the Franz-Keldysh effect, to realize high-performance modulators with high dynamic range and a linearized electroabsorption characteristic (Welstand et al., 1999).

¹³ Johannes Stark (1874-1957), German physicist and 1919 Nobel laureate in physics, is best known for his discovery of the splitting of spectral lines in high electric fields, an effect which today bears his name. His lack of success in an academic career likely led to his unsavory associations with the Nazis and resulted in a sentence of four years in a labor camp after WWII.

10.5.3 Modulator materials

The most efficient modulators are made from materials which exhibit a strong non-linear response in the susceptibility, as we saw in Section 3.2.5. For practical free-space optical modulators, in macro- or micro-optical implementations, artificial (man-made) materials are most popular, due to their strong electro-optic effects. Lithium niobate (LiNbO_3) or lithium tantalate (LiTaO_3), as well as potassium niobate (KNbO_3), show excellent non-linear behavior in the visible and NIR wavelength regions. A typical value for the Pockels coefficient in LiNbO_3 is $r = 3.1 \cdot 10^{-11}$ m/V, and thus an order of magnitude larger than in GaAs.

Also popular, due to their lower cost, are KDP (monopotassium dihydrogen phosphate, KH_2PO_4) and KTP (potassium titanyl phosphate, KTiOPO_4); these form the basis for most discrete modulators available commercially. These materials exhibit primarily the linear electro-optic effect, and can be chosen to have low chromatic dispersion and negligible electroabsorption.

We already saw in the discussion above that some semiconductors exhibit considerable non-linear behavior. GaAs was used as an example repeatedly; it exhibits both Pockels and Kerr effects, where we saw that $r = -1.44 \cdot 10^{-12}$ m/V and $s = -5.4 \cdot 10^{-20}$ m²/V² at $\lambda = 1090$ nm. The relative strength of the Pockels and Kerr effects varies widely between materials (Bennett and Soref, 1987; Soref and Bennett, 1987). The Pockels effect is prominent in non-centro-symmetric materials, such as GaAs or InP, but it is negligible in Si. The Kerr effect dominates in centro-symmetric or amorphous materials; it is thus weak in III-V semiconductors but present in silicon. In semiconductors, several electro-optical effects can be present simultaneously, sometimes leading to strongly non-linear behavior in the phase shift as a function of applied electric field.

Example 10.5: Let us compare the relative strengths of the Pockels, Kerr and Franz-Keldysh effects in GaAs at an operating wavelength of 1.55 μm . We take a 1 cm long, 1 mm thick block of this material and apply 150 V across the short dimension. What are the phase shifts which result from the three effects?

Taking the magnitudes of the coefficients at 1.55 μm , we find for GaAs that, for the Pockels effect, $|r| = 1.44 \cdot 10^{-12}$ m/V, for the Kerr effect $|s| = 8 \cdot 10^{-22}$ m²/V², and for the Franz-Keldysh effect $|F_{FK}| = 7.9 \cdot 10^{-20}$ m²/V². Taking $n_{\text{GaAs}} = 3.064$ at that wavelength, and determining the electric field to be 150 V/1 mm = $1.5 \cdot 10^5$ V/m, we obtain the values for real refractive index shift as

$$\Delta n|_{\text{Pockels}} = 3.1 \cdot 10^{-6}$$

$$\Delta n|_{\text{Kerr}} = 2.6 \cdot 10^{-10}$$

$$\Delta n|_{\text{Franz-Keldysh}} = 1.8 \cdot 10^{-9}.$$

Calculating the resultant phase modulation for the 1 cm length using Equation 10.35, we obtain

$$\Delta\phi|_{Pockels} = 0.25 \text{ rad}$$

$$\Delta\phi|_{Kerr} = 2.1 \cdot 10^{-5} \text{ rad}$$

$$\Delta\phi|_{Franz-Keldysh} = 1.4 \cdot 10^{-4} \text{ rad.}$$

We see thus that the linear electro-optic effect dominates in GaAs at this wavelength. Recall that both the Kerr and Franz-Keldysh effects are strongly wavelength dependent, and they increase in magnitude for shorter wavelengths, particularly as these approach the bandgap.

10.5.4 Modulator structures

Now that we know how to obtain a shift in real or imaginary refractive index in a material, we can consider the device configurations used in practice to generate a desired phase or intensity shift (Mias and Camon, 2008a,b). Practical optical modulators may be divided into two categories, bulk or integrated optical, as seen in Figure 10.28. The former may take on very macroscopic dimensions, and are generally used in free-space optical systems; the latter are most applicable to guided-wave implementations.

Bulk

The simplest form of modulator is the bulk structure shown schematically in Figure 10.28(a); bulk modulators are predominantly employed for phase modulation. The optical field traverses a length L of the electro-optically active material, and an electric field normal to the propagation direction is generated by the application of a voltage across the entire substrate, with thickness t . For an electrically-induced Δn or $\Delta\alpha$, the resultant phase shift or intensity change is then given by Equations 10.35 or 10.36.

For the materials typically employed, frequently crystals of KDP or LiNbO_3 , the phase shift as a function of applied bias V is then

$$\Delta\phi = \frac{2\pi L}{\lambda_0} n_0^3 s \left(\frac{V}{t} \right)^2. \quad (10.41)$$

We see from this expression that significant phase shifts for a given voltage thus require large electro-optic coefficients or a long modulator length, L , or both; typical bulk modulator dimensions are in the millimeter to centimeter ranges. Since we saw above that electrorefraction and electroabsorption are always to some extent present, pure phase or intensity modulators should have as large a difference between Δn and $\Delta\alpha$ as possible. A longer modulator length exacerbates the effect of undesired residual electroabsorption in a phase modulator.

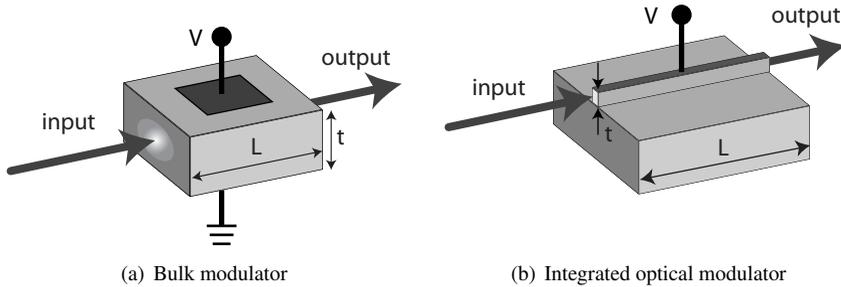


Figure 10.28 Modulators rely on the propagation of the optical field through a length L of electro-optically active material; the electric field for modulation is generated by a voltage applied normal to the propagation direction over the thickness t . The bulk modulator uses a propagation through a block of material, whereas the integrated optical modulator confines the optical field to a waveguide.

Example 10.6: We can determine the voltage required to achieve a phase shift of $\Delta\phi = \pi$, V_π [V], for a bulk modulator of length L by setting

$$\Delta\phi = \Delta k \cdot L = \frac{2\pi}{\lambda_0} \Delta n \cdot L = \pi.$$

If we consider the linear electro-optic effect, for which we know

$$\Delta n = \frac{1}{2} n_0^3 r \mathbf{E},$$

we obtain

$$\frac{2\pi}{\lambda_0} \frac{1}{2} n_0^3 r \frac{V_\pi}{t} \cdot L = \pi$$

from which we determine

$$V_{\pi(\text{bulk})} = \frac{\lambda_0}{n_0^3 r} \frac{t}{L}$$

for material thickness t , across which the bias is applied.

If we consider a 5 mm long, 100 μm thick piece of GaAs as a bulk modulator for $\lambda_0 = 850$ nm, for which $|r| = 1.44 \cdot 10^{-12}$ m/V and $n_0 = 3.66$, we find the required applied voltage for a π phase shift of $V_\pi = 241$ V. This relatively high bias is due to the relative inefficiency of GaAs as a modulator material and the fact that the bulk structure requires a high voltage to reach a given electric field.

Integrated optical

Waveguide-based integrated optical modulators have two advantages over their bulk counterparts: the optical field is confined in a thin waveguide structure, over which a

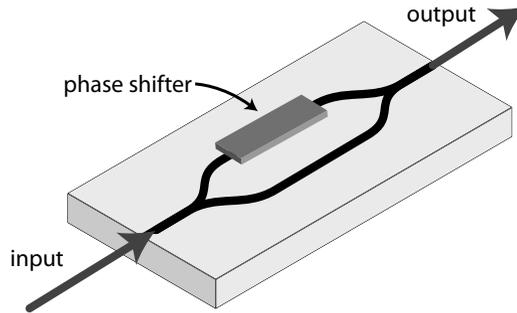


Figure 10.29 An integrated optical Mach-Zehnder interferometer used for intensity modulation. The phase shifter in one arm leads to a shift in the interference pattern and thus the output intensity, thereby converting phase to intensity modulation.

high electric field may be generated at relatively low voltages; and guided-wave systems may employ more complex arrangements for achieving modulation, while retaining a relatively small size. The latter consideration implies that, for example, integrated optical interferometers may be used to generate highly effective modulation.

Waveguide-based intensity modulators commonly employ the Franz-Keldysh effect or the QCSE; since integrated optical modulator structures are generally fabricated using semiconductors, these effects are usually dominant in the materials employed, Si or III-Vs. As seen in Figure 10.28(b), the waveguide modulator is operated by the application of the voltage only across the waveguide, with thickness t , which is typically in the range of several micrometers. As a result, the waveguide modulator can attain the electric fields necessary for modulation at considerably lower voltages than the millimeter-thick bulk structures.

Intensity modulation in waveguide-based modulators may also very effectively be achieved using phase modulation and an integrated-optical interferometer. An example is shown in Figure 10.29, in which a Mach-Zehnder interferometer is used for intensity modulation. As we learned in Section 5.2.3, the Mach-Zehnder configuration has an intensity output which is a sensitive function of the relative phase shifts in the two arms of the interferometer. By placing a phase shifter, which relies on the electro-optical or perhaps thermo-optical characteristics of the waveguide substrate, in one arm, the induced $\Delta\phi$ resulting from real Δn may thus be converted into a shift in the output intensity, and thus effectively $\Delta\alpha$.

Recalling from Section 5.2.3 that the relationship between input and output intensity for the Mach-Zehnder interferometer is given by

$$\frac{I_{out}}{I_{in}} = \left[1 + \cos \left(2\pi \frac{L\Delta n}{\lambda_0} \right) \right], \quad (10.42)$$

we see again that large values for Δn and/or L are desirable for modulation at low biases. We also realize that the output intensity varies periodically with phase shift, and thus Δn . Assuming a balanced interferometer with no applied bias to the phase shifter (i.e., relative $\Delta\phi$ between the arms is zero, so that output intensity is maximum), an

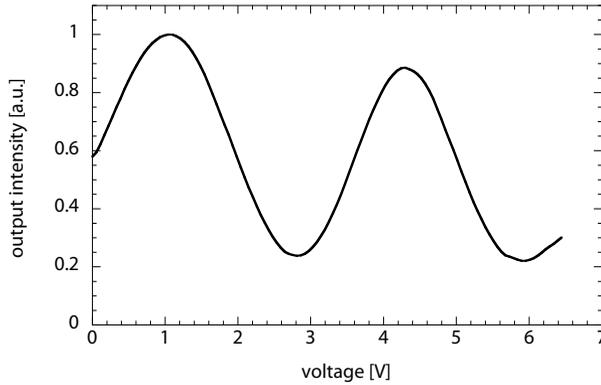


Figure 10.30 Transmission through a waveguide-based Mach-Zehnder intensity modulator. The quantum well-based GaAs device used a 1 mm long push-pull arrangement, operated at 850 nm.

induced real refractive index shift of magnitude

$$\Delta n_{\pi} = \frac{\lambda_0}{2L} \quad (10.43)$$

will result in $\Delta\phi = \pi$ and thus a minimum of output intensity. Larger values of induced Δn will result in a subsequent increase of interference intensity, until, for $\Delta n = 2\Delta n_{\pi}$, the maximum is reached again. For guided-wave implementations, employing III-V or other electro-optically active materials and benefitting from the use of quantum wells, the interferometer lengths required for reaching the first interference minimum are typically less than several millimeters and the voltages below 10 V.

A typical output characteristic of such a Mach-Zehnder based intensity modulator is shown in Figure 10.30, in which a GaAs-based push-pull interferometer (in that both arms are modulated, with opposing phase shift) traverses the periodic interference characteristic. We see a number of interesting aspects in the behavior. First, the Mach-Zehnder is not balanced in its unbiased state, since at $V = 0$, the intensity is not at its maximum; thus a residual phase shift, due to processing non-idealities in the waveguide, is present in the structure. Furthermore, the interference is not complete, resulting in a minimum intensity of about 0.2 (rather than 0), and thus reduced contrast when compared to that ideally possible. Finally, it is clear that residual electroabsorption is present in the modulators, since the maximum of the characteristic decreases as the bias increases, further reducing the contrast for higher bias. We see thus the combination of electro-optical effects which typically occur in real-life devices.

Example 10.7: For an integrated optical Mach-Zehnder interferometer, we can determine typical values for V_{π} and compare these with the values we found for the bulk modulator of Example 10.6. Rewriting Equation 10.42 with its mathematical equiva-

lent,

$$\frac{I}{I_0} = \cos^2 \left(\pi \frac{L \Delta n}{\lambda_0} \right) = \cos^2 \left(\pi \frac{V}{V_\pi} \right)$$

we can solve this expression for V_π to yield

$$V_{\pi(MZ)} = \frac{2\lambda_0}{n_0^3 r} \frac{t}{L}$$

for the linear electro-optical effect.

Again using the typical electro-optical coefficient value for the Pockels effect in GaAs (at $\lambda_0 = 850$ nm), we determine V_π for a typical waveguide modulator, of thickness $t = 1$ μm and length $L = 1$ mm, to be $V_\pi = 24$ V. We see that this value is considerably smaller than that calculated for the bulk modulator in the last example, due to the considerably smaller thickness of the material over which the bias is applied: 1 μm for the waveguide versus 100 μm for the bulk structure.

We saw that semiconductor-based waveguide modulators may benefit from the use of quantum well structures and thus the QCSE. For GaAs in the 850 nm wavelength range, a triple quantum well structure can lead to an effective electro-optic coefficient of $r_{QCSE} \approx 19 \cdot 10^{-12}$ m/V; from this value, we find that $V_{\pi(QCSE)} = 1.8$ V is expected. Such an applied bias is well within the range of typical electronic driving circuitry, explaining why such waveguide-based modulators are widely employed.

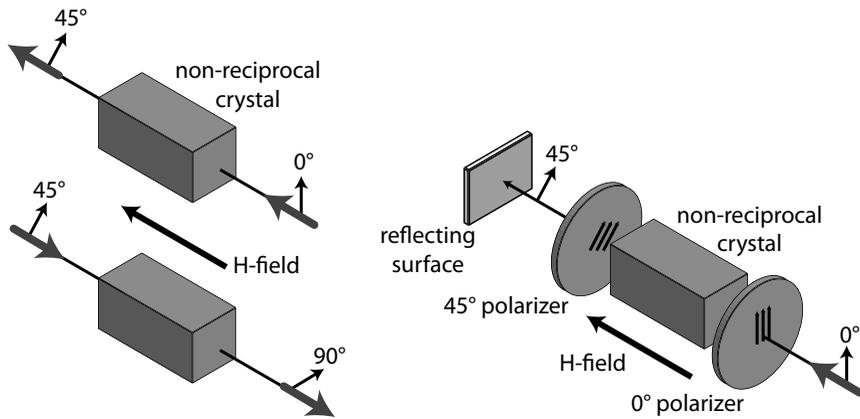
10.6 Other active components

In the catalogue of other active micro-optical components beyond those we have considered above, there are several which are particularly relevant for use in optical microsystems; some of these tend to be part of macroscopic optical systems which nevertheless play an important role in realizing a certain optical functionality or are important in laboratory optical setups. We consider optical isolators and liquid crystals here; the former are essential in high-performance laser diode modules and the latter are widely employed in displays and for tunable micro-optics.

10.6.1 Optical isolators

Some optically active devices, particularly laser diodes and optical amplifiers, are very sensitive to light reflected into themselves. If a portion of the emission from a laser diode is fed back into the cavity, the emission stability can be adversely affected, to the point of coherence collapse. Optical feedback can have have undesirable effects even for feedback levels of -60 dB or less.

The function of the optical isolator is to allow light to be transmitted in only one direction. Placed in the optical path of emission from a laser diode, the isolator then



(a) The Faraday effect results in the rotation of the polarization in the same direction (clockwise, in the frame of reference of the beam incident from the right) for propagation in either direction through the crystal.

(b) An optical isolator uses a non-reciprocal magneto-optic crystal and two polarizers, typically oriented at 45° with respect to each other; reflected light is rotated to 90° in this example and thus blocked by the input polarizer.

Figure 10.31 Optical isolation may be accomplished by a magneto-optic material which gives rise to non-reciprocal polarization rotation in the presence of a magnetic field.

prevents light reflected from other optical surfaces from re-entering the laser; such a configuration is standard in all applications, particularly telecommunications and spectroscopy, where absolute stability of laser emission under all conditions is required.

Magneto-optic isolators

The most popular means to achieve optical isolation is by the use of the Faraday effect, present in magneto-optic materials. As illustrated in Figure 10.31(a), the Faraday effect gives rise to non-reciprocal optical transmission through a medium by rotating the polarization of an incident optical field in the same direction regardless of propagation direction. Thus in Figure 10.31(a), linearly-polarized light incident from the right with a polarization angle of, for example, 0° , is rotated clockwise by 45° . Light incident from the left, however, is also rotated by 45° *in the same direction*, also clockwise as seen from the frame of reference of the beam incident from the right (counterclockwise as seen from the frame of reference of the incident field from the left).

The polarization rotation angle in a magneto-optic material, Θ [rad], is given by

$$\Theta = V \cdot H \cdot L \quad (10.44)$$

where V [rad/Tm] is the Verdet constant¹⁴ of the material and L its length.

¹⁴ Émile Verdet (1824–1866), French physicist and professor at the École Polytechnique in Paris, was noted for primarily for his editing and publication of the collected works of Fresnel, although he was also the first to accurately calculate the coherence of sunlight.

Optical isolation using magneto-optic materials is thus accomplished as shown in Figure 10.31(b). Incident linearly-polarized light is coupled into the non-reciprocal material through a polarizer, oriented at 0° in this example. In the presence of a magnetic field, the polarization is rotated by 45° and emerges through the second polarizer, oriented at 45° . Reflected light enters from the left through this 45° polarizer, is rotated by another 45° in the same direction, and the resultant 90° polarized light is blocked by the 0° input polarizer.

Isolators of this form are somewhat limited in that they require linearly-polarized light to function properly. Whereas this is usually the case for emission from laser diodes, unpolarized emission requires alternative concepts, usually using birefringent materials to split the two polarization components and rotating these separately.

Magneto-optic materials

A number of magneto-optic materials which exhibit the Faraday effect is available, including glasses (typically diamagnetic or paramagnetic) and crystals (typically ferri-magnetic) (Borrelli, 2005, Chapter 8.5.2). Examples include the diamagnetic SF59 glass from Schott, for which $V = 27^\circ/\text{Tm}$ or the crystalline $\text{Bi}_3\text{Fe}_5\text{O}_{12}$ with a polarization rotation of $-55 \cdot 10^3^\circ/\text{cm}$ for an applied magnetic field above the saturation value, several tenths of a Tesla. Most popular are likely the YIG materials, yttrium iron garnets, with a typical polarization rotation of $870^\circ/\text{cm}$, again for H greater than several 0.1 T. All of the previous values are valid for the wavelengths around $\lambda = 633 \text{ nm}$.

Waveguide-based isolators

The optical isolator described in Figure 10.31 tends to be a macroscopic device (free-space isolators are about the size of a water glass and can weigh a kilogram, due to the magnets) and as such are essential components in many optics laboratories. For use in integrated and micro-optical systems, however, waveguide-based isolators tend to be more suitable. Also practical are fiber-based isolators, which use Faraday materials but can be considerably smaller than free-space components due to the much smaller aperture required.

A number of integrated optical approaches has been demonstrated. Conceptually simple is to fabricate the waveguide from a magneto-optic material; Bi substituted YIG (Bi:YIG) has been grown epitaxially on GGG (gadolinium gallium garnet, $\text{Gd}_3\text{Ga}_5\text{O}_{12}$) and integrated with a sputter-deposited $\text{Sm}_{13}\text{Co}_{58}\text{Fe}_{20}\text{Cu}_7\text{Zr}_2$ thin-film magnet, leading to 29 dB of optical isolation (Levy et al., 1996). An interferometric approach using an integrated-optical Mach-Zehnder interferometer splits the TE and TM components and isolates these separately, thereby relaxing the polarization requirement (Fujita et al., 2000); the structure was based on a $1.65 \mu\text{m}$ thick $(\text{Bi, Lu, Nd})_3(\text{Fe, Al})_5\text{O}_{12}$ ((bismuth, lutetium, neodymium) iron garnet) layer, also grown on GGG, in an 8 mm long device. Its design led to destructive interference of the backward-traveling waves, and thus an isolation of about -20 dB in the $1.55 \mu\text{m}$ wavelength range.

Research is active on improving the performance of integrated optical isolators, and numerous alternative concepts have been proposed. One structure uses a magneto-optic

Co layer deposited on a GaAsP/AlGaAs waveguide structure configured as an optical amplifier (Zaets and Ando, 1999); the exponential field in the Co layer leads to direction-dependent loss in the near-IR wavelength regime, and the amplifier compensates for further residual losses due to absorption in the metal. Alternatively, a one-dimensional magneto-phonic crystal, using a film structure forming a Bi:YIG cavity between layers of SiO₂ and Ta₂O₅ has been employed (Kato et al., 2002); a large Faraday rotation could be achieved with only two to three such cavity layers in the optical path. Non-reciprocal loss shift has also been seen for InGaAsP waveguides on InP with Co, Fe or Ni/Fe coated on the sidewalls, leading to 6 – 9 dB/mm extinction ratios (Shimizu and Nakano, 2006).

10.6.2 Liquid crystal devices

Liquid crystals are phases of matter with characteristics between those of liquids and crystalline solids, and may often be switched between these phases by an external stimulus (Chandrasekhar and Madhusudana, 1980; de Gennes and Prost, 1995). Consisting in general of small, linearly-shaped molecules, there is a very wide variety of materials which exhibits liquid crystal behavior, both man-made and natural; the first such molecule studied was a cholesterol acetate extracted from a carrot (Reinitzer, 1888). Key to the utility of liquid crystals is that the orientation of the molecules may easily be re-oriented by application of electric or magnetic fields, temperature changes or pressure. When the molecules are periodically oriented, a crystalline structure results, but the binding forces between the molecules are very weak, making external change of crystalline orientation relatively easy.

Liquid crystals are very useful elements for realizing a number of optical functions, particularly polarization rotation and phase shifting, and as displays. The liquid crystal display (LCD) is an extremely low-cost component used as an indicator in an endless variety of consumer products but also forms the basis for optically more complex devices.

Liquid crystal types

There are numerous types of liquid crystals, based on the configuration of the individual crystalline elements; these are represented by the little olive-shaped elements in Figure 10.32. The nematic liquid crystal, shown in the figure, is one of the most useful types and consists of the crystals oriented in parallel, in a direction defined by the so-called director. Optically, nematic liquid crystals behave as uniaxial crystals, birefringent materials to which we were introduced in Section 3.2.4, so that they are characterized by ordinary and extraordinary refractive indices. n_e is typically in the direction of the director, and n_o perpendicular to this axis.

When used in an encapsulated system, the orientation of the liquid crystals is defined by microscopic grooves formed on the plates which enclose the crystals, usually generated by rubbing a cloth in a particular direction on the cover material, typically a polyimide coated plastic or glass. Three useful configurations are shown in Figure 10.33:

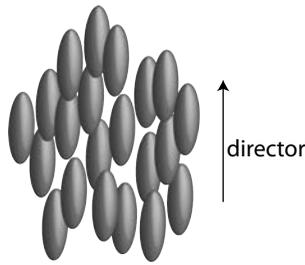


Figure 10.32 Liquid crystals consist of elongated crystallites in suspension; in a nematic liquid crystal, all crystals are oriented in the same direction, defined by the director, with no layering in the vertical direction.

Homogeneous (Figure 10.33(a)), in which grooves in the same direction on top and bottom plates result in the crystals oriented parallel to this direction;

Homeotropic (Figure 10.33(b)) which uses chemical anchors (such as lecithin) on the top and bottom surfaces to orient the crystals normal to the surfaces; and

Twisted (Figure 10.33(c)), which relies on grooves on the top and bottom surfaces oriented at an angle (usually 90°) with respect to each other, to result in a gradual twist of crystal orientation from one surface to the other.

The twisted nematic configuration is usually the basis for LCDs and may also be used for controlled polarization rotation.

Liquid crystal displays

The utility of liquid crystals lies in the ability to change their orientation by application of an external stimulus, most practically an electric field. Using nematic liquid crystals, the orientation of the director may be switched by an applied bias. For $n_e > n_o$, the director orients itself parallel to the field (positive dielectric anisotropy), and for $n_e < n_o$, it switches orientation to perpendicular to the field (negative dielectric anisotropy).

Switching of liquid crystals forms the basis for their most widely used application, namely that of the liquid crystal display. The LCD consists of a roughly $10\ \mu\text{m}$ thick twisted nematic liquid crystal layer between two cover plates; these are in turn sandwiched between two crossed polarizers. In the absence of an applied electric field, the twisted orientation of the liquid crystals rotates the polarization of the light, incident through one polarizer, by 90° , such that the light emerges from the second (crossed) polarizer; the cell is thus transparent.

If an electric field is applied normal to the plane of the LCD, the liquid crystals, typically chosen to have positive birefringence, orient themselves parallel to the field, such that the polarization rotation is suppressed. As a result, the linear polarized light is not transmitted by the second polarizer, and the cell is opaque, or dark. This arrangement is widely used for monochrome and color displays, the latter realized by the use of color filters on individual LCD pixels.

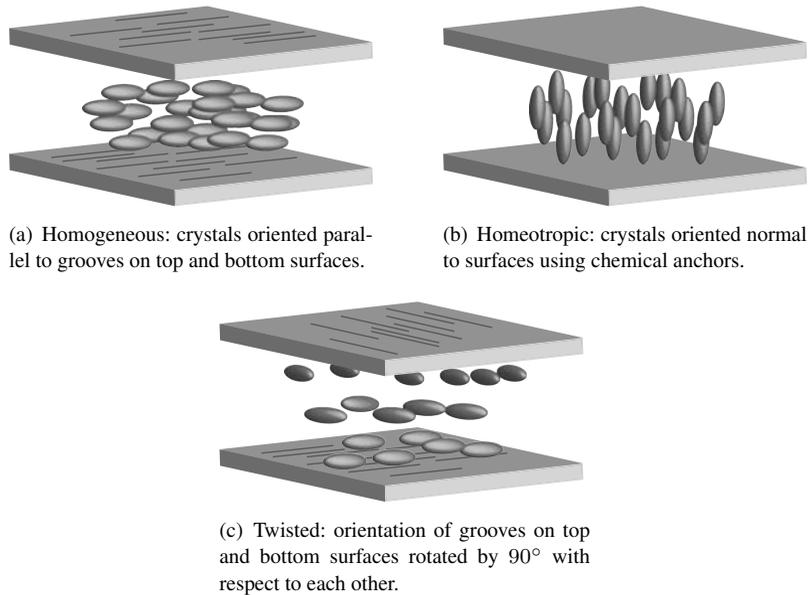


Figure 10.33 Various liquid crystal configurations are defined by the nature of the features defined in the upper and lower confining plates; the grooves shown are on the inside of the cavity, in contact with the crystals.

Liquid crystal micro-optics

In addition to displays, liquid crystals have a wide variety of other uses relevant for micro-optics. In addition to digital switching such as that used in an LCD, two-dimensional arrays of liquid crystals, usually fabricated as an active pixel structure on CMOS (liquid crystal on silicon, LCOS), can be used to generate two-dimensional phase shift patterns, thus allowing the generation of dynamic holograms (Glaser et al., 2007); liquid crystal arrays and LCDs may also be used for arbitrary phase modulation (Soutar and Lu, 1994; Remenyi et al., 2003). The performance of LCD-based displays is enhanced by the use of micro-optical structures and surface features (Shieh et al., 2005), predominantly designed to enhance the out-coupling efficiency of the emitted light and to improve the viewing angle, all of this achromatically. Liquid crystal technology has also proven useful for the realization of micro-optical functions themselves, including switchable gratings (Yeralan et al., 2002) and reconfigurable microlenses (Masuda et al., 1997), which we will examine in greater detail in Chapter 12.

10.7 Active micro-optics: case studies

Active micro-optical devices, particularly laser diodes, are probably the most extensively used micro-optical components. The driving force behind the development of

diode lasers was their use in optical data communications systems and it is still this application which likely places the most stringent technical demands on lasers, modulators, and detectors. We have looked at optical telecommunications as part of two case studies in past chapters: optical switching in Section 6.6.3 and optical data transmission using integrated optical components in Section 9.7.1. The active micro-optical and integrated-optical components we encountered in those discussions represent very advanced optoelectronic technology.

The developments spurred by the demands of optical telecommunications have, however, resulted in laser diodes and photodetectors becoming available for myriad other applications. Both consumer products, such as optical data storage, and niche markets, such as absorption spectroscopy, have benefitted from the fact that telecommunications lasers may be mass-fabricated and employed for uses other than the ones for which they were originally intended. We look at these two areas as representative examples.

10.7.1 Optical data storage

The application which consumes the largest volume of lasers annually is optical data storage. With an estimated 714 million CD, DVD or Blu-ray lasers sold in 2008, corresponding to a total market value of \$1.7 billion, most of these were Fabry-Perot devices, either dual wavelength red/infrared (650/780 nm) or blue (405 nm) devices (Steele, 2008). Of all micro-optical components, the diode laser used in optical storage is that which has the largest consumer market penetration; it is likely that every reader of this book has purchased several of these active micro-optical components in her or his lifetime, only possible due to their low price which derives from the very high volumes fabricated and sold.

Optical disc structure

Most optical data storage is based on CD, DVD or BD (Blu-ray disc) technologies, all of which rely on optical detection of patterns defined in a 12 cm diameter, 1.2 mm thick polycarbonate (PC) disc for data readout (Bouwhuis et al., 1985). The encoded data on the disc is a single track formed as a continuous spiral; for the CD, this spiral is about 5.76 km, for the DVD, 27.85 km long. The data is encoded into pits of various sizes along this track, varying from 0.833 μm to 3.56 μm in length and with a uniform width of 0.5 to 0.6 μm , and the track is read by scanning it with a constant linear velocity of 1.4 m/s past an optical reader head.

For all three media, the data is encoded into this pit pattern of the track (Davies, 1988). Contrary to popular belief, the digital 1 and 0 are not expressed by the presence or absence of pit: a 1 is read at the beginning or end of a pit (at the sidewall representing the transition), and the rest of the data represents a string of 0s. This concept is the reason for the constant linear read speed (1.4 m/s) and the precise definition of the pit length: for the CD, a 0.833 μm long pit represents 000 (the shortest 0 sequence) and a 3.56 μm long pit represents 0000000 (the longest 0 sequence). These 0 sequences are separated by a 1, and the data is mathematically converted into a digital form corresponding to this coding scheme. Data is read from the CD in “frames”, 588 bits long, of which only

Table 10.1. Comparison of some salient characteristics of the CD, DVD and BD; the data for single-layer discs is given.

parameter	CD	DVD	BD
wavelength [nm]	780	650	405
NA	0.45	0.6	0.85
spot focus [μm]	1.3	1.08	0.47
scan speed [m/s]	1.4	3.49	4.917
track pitch [μm]	1.6	0.74	0.32
minimum pit length [μm]	0.83	0.4	0.16
data read rate [Mb/s]	1.41	11.08	35.965
capacity [bytes]	650 M	4.7 G	25 G

192 bits are the actual data. The remaining bits are overhead, used for error correction, tracking and command and display. Thus only about 1/3 of the bits written onto a CD represent the data to be recorded; the rest is required for the system to work reliably.

The CD, DVD and BD represent the first, second and third generations of optical data storage technologies, respectively; a comparison of some of their relevant features is given in Table 10.1. The DVD and BD have a number of variations, especially dual layer structures (which roughly double the capacity to 8.5 Gb and 50 Gb, respectively). There are a number of other optical recording technologies in use, notably magneto-optical discs (Awano and Ohta, 1998), but these are niche products in comparison to these established technologies; writable discs use the color change of a dye (not a surface relief) to record data, and read/write discs often use phase changes due to crystallization or annealing of a AgInSbTe layer, also a different technology.

CD reader head

The CD, DVD or BD pick-up head represents a compact, ultra-low cost optical microsystem, which performs astonishing functions, especially when considering that it is mass-fabricated from plastic parts; the laser diode is likely the costliest component of the system, and has a price averaging about € 1; the entire head costs about € 3. The CD pick-up system is shown schematically in Figure 10.34: emission from the laser diode is directed through a polarizing beam splitter through an objective lens and focussed onto the data layer of the disc. The reflected light is directed by the beam splitter onto a detector arrangement, usually a quadrant detector (four photodetectors arranged such that each represents one quadrant of the Cartesian plane).

The optics of the pick-up head is designed (Bouwhuis et al., 1985) so that the intensity distribution on the four detectors allows determination of the tracking error (keeping the laser spot on the spiral data track), the focus error (since the disc is hardly flat and wobbles anyway) and of course the data to be read. The tracking data is used as a feedback signal to move the head radially as the track is read and the focus data is used to dynamically move the entire system vertically, keeping the focus spot on the data layer. Tracking is typically maintained to $\pm 0.2 \mu\text{m}$ and focus to $\pm 2 \mu\text{m}$.

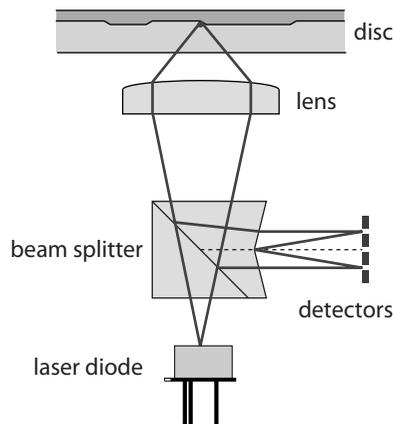


Figure 10.34 This representative example of the CD pick-up head consists of a laser diode, a polarizing beam splitter, one or more objective lenses and a photodetector array or quadrant detector; the distribution of light reflected from the disc surface on the photodetectors allows determination the tracking error, focus error and the data to be read. Not to scale

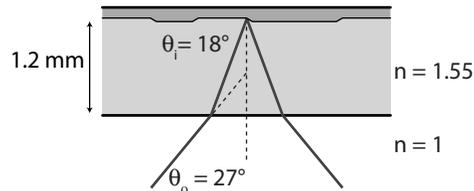


Figure 10.35 The data on a CD lies below the surface, on the back side of the disc. As a result, refraction at the front results in a focus spot well below this exposed surface, reducing the optical effect of scratches or dirt on the top surface.

The optical disc is designed so that the data layer (the pits) is found below the surface; for a video DVD, it is essentially just below the printed label. As a result, the rays of the optical field focussed onto the data layer look like those shown in Figure 10.35: refraction at the disc/air interface is used to achieve a focus point under 1.2 mm of polycarbonate. As can be seen from the figure, the effect of the surface properties of the disc (scratches, dust, children's fingerprints) is thus mitigated, explaining why an audio CD often still works after being tossed around in the back seat of the car for a year. Coupled with the redundancy programmed into the data, this robustness implies that a 2 mm stripe can be scratched into the surface and the data (usually) still recovered.

The pick-up heads for the DVD were designed to be backwards compatible with the CD (Shinoda et al., 1996) and share many of the same features. The most important difference is that of optical spot size; we compared the expected resolutions of the lens systems for CD, DVD and BD readers in Table 8.1. This resolution is achieved using a low-cost plastic optical system fabricated using a molding process; the CD lens is low weight (< 2 g) and has a typical focal length of 8 mm. The entire pick-up head assembly

is actuated vertically (for focus) using an electromagnetic voice coil and radially (for tracking) using a dc motor or magnetic actuator.

Lasers

The laser diodes form the heart of the optical disc reader system, and it is on both the availability and performance of these devices that the technology relies. The reason that red and near-IR laser diodes are available in huge numbers and at such low cost is due to the fact that they are fabricated in quantities of hundreds of millions; many other optical and consumer applications unrelated to data storage have benefitted as a result. In contrast, the Blu-ray disc technology had no chance of success until the GaN-based blue laser diode technology was sufficiently advanced to provide lasers with lifetimes of 5 000 hours or more at acceptable cost.

The CD lasers were originally gain-guided, single-mode Fabry-Perot lasers, emitting 5 mW at 780 nm. For read/write discs, which use either a dye or a specialized alloy (AgInSbTe) as the data layer, 30 mW devices are used. The DVD moved from the IR to the visible, and employs 650 nm lasers, at the same power levels, for the same applications. The eponymous Blu-ray lasers are indeed blue, GaN devices emitting at 405 nm. Whereas laser technology has thus strongly been pushed by this consumer market, the writing is currently (2009) on the wall that cheap flash memory, the staple of iPods and MacBook Airs, is beginning to encroach on the optical storage market.

10.7.2 Absorption spectroscopy

In contrast to the wide consumer market for optical storage microsystems, this second case study represents a specialized application which relies on highly developed laser diodes with tightly specified optical characteristics. Absorption spectroscopy is a sensor technique, used primarily for detection of low concentrations of particular compounds based on their optical absorption.

Many atomic and molecular species have optical absorption resonances at mid-IR wavelengths, in the range 3 – 30 μm ; overtones of these resonances, or rotational and vibrational transitions, often result in weaker absorption peaks at near-IR or even visible wavelengths. Tunable diode laser absorption spectroscopy (TDLAS) is a sensor technique in which the emission wavelength of a tunable laser diode is scanned over the absorption lines of the substance to be detected; from the strength of the resultant absorption features in the spectrum, the concentration can be determined. The technique requires narrow-linewidth lasers which emit in the wavelength regime of interest, and whose emission wavelength can be tuned, electrically, by temperature or through external means.

Gas sensing

TDLAS is well-established for trace-gas sensing (Ried et al., 1978; Cassidy and Reid, 1982; Kroll et al., 1987), since most gasses have extensive absorption spectra. An example spectrum is shown for O_2 in Figure 10.36: these absorption features correspond to rotational transitions of O_2 and we see that there are numerous closely spaced peaks

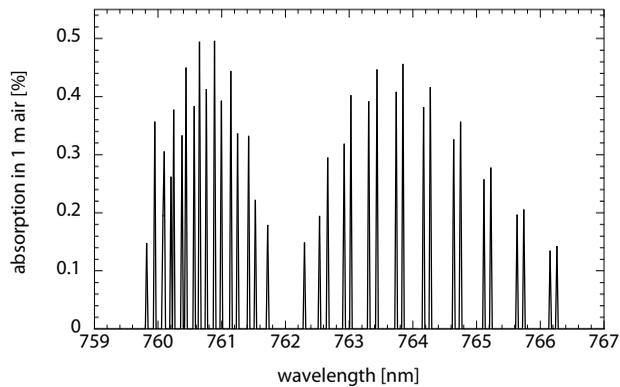


Figure 10.36 Absorption spectrum of O_2 around 762 nm, corresponding to the blue end of the P-branch of rotational transitions in the oxygen A band. The ordinate value represents the percent absorption in 1 m of air, with the standard mole fraction of oxygen (20.9 volume % of O_2 at 300 mbar).

around 762 nm. Absorption spectra such as these are typically available in the widely-disseminated HITRAN (“high-resolution transmission molecular absorption”) database (Rothman et al., 2003).

Laser diode-based absorption spectroscopy uses a measurement configuration such as that shown in Figure 10.37. Emission from a tunable laser diode (a VCSEL is shown) is directed through a gaseous medium and onto a photodetector. The emission wavelength of the laser is scanned over the absorption peaks of the species to be detected, typically using either current or temperature tuning. The dip in the transmission through the cell due to absorption is measured during the scan by the photodetector, and the strength of this dip allows determination of the concentration. As with the other optical absorption mechanisms we have discussed in this text, the intensity decreases exponentially with distance and absorption coefficient as predicted by Beer’s law, Equation 10.2, so that, for a small absorption coefficient (due, perhaps, to low concentration of the absorbing species), long path lengths are required to obtain good detector signals.

The result of such a measurement is given in Figure 10.38, in which a temperature-stabilized VCSEL was scanned across two O_2 absorption lines around 763 nm (Zappe et al., 2000). Since the emission wavelength of a VCSEL changes strongly (due to internal temperature shifts) with drive current, the current scan from 3 to 4.5 mA corresponds to a wavelength shift of about 0.3 nm. The two dips in the output power curve, due to oxygen absorption after transmission through 1 m of ambient air (20.9 volume % of O_2 at 300 mbar), are clearly seen, and the positions may be more precisely determined by examining the second derivative of the PI curve, a standard analysis technique. By comparing the peak positions with those expected from the known absorption spectrum of Figure 10.36, we see that the two peaks found here correspond to O_2 at 763.312 and 763.427 nm.

The TDLAS approach, using variations of the setup just described, has been used for sensing of a wide variety of gas species, including CO_2 and H_2S (Weldon et al., 1995),

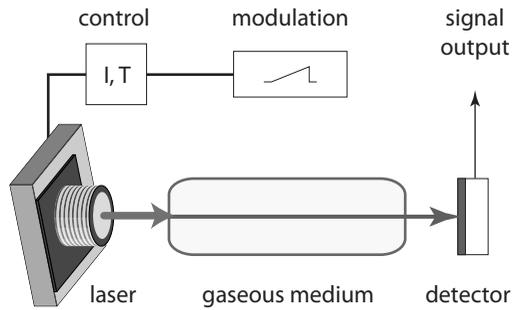


Figure 10.37 The measurement configuration for absorption spectroscopy: emission from a wavelength-tunable diode laser is directed through a gas-filled measurement cell, whereupon the detector registers a change in intensity as the laser wavelength sweeps across the absorption peak.

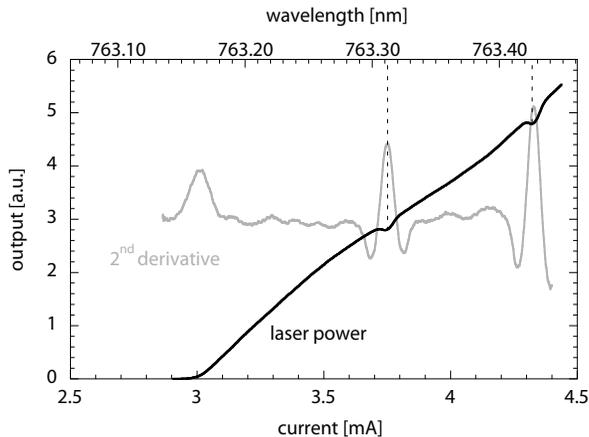


Figure 10.38 The scan across two O₂ absorption lines using current-tuning of a VCSEL. Two dips in the PI characteristic are seen, features made more clear by examining the second derivative of the scan. The absorption peaks correspond to those at 763.312 and 763.427 nm.

CH₄ (Nagali et al., 1996), NO₂ (Gianfrani et al., 1997) and NH₃ (Larson et al., 1997) among others. For high gas concentrations, a centimeter to meter long path through the measurement cell is often sufficient to generate an acceptable sensor signal; for high sensitivities requiring a long analyte path, multi-pass cells which achieve a long path length through multiple reflections in an enclosed cavity are often used (Helou et al., 2006). The sensor technique also works for long paths using atmospheric transmission, and has been used in the field for measurement of CO₂, for which the isotopes ¹²CO₂ and ¹³CO₂ could be distinguished (Bowling et al., 2003; Griffis et al., 2004).

Laser types

Since most atomic and molecular resonances are found in the mid-IR range, many original efforts in the use of diode lasers for absorption spectroscopy involved the use

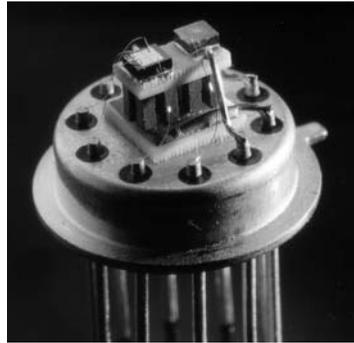


Figure 10.39 A 4×4 VCSEL array (the chip on the left) used for absorption spectroscopy, mounted on a micro-thermoelectric cooler with an integrated thermistor (the chip on the right). Photo courtesy of CSEM.

of lasers at these wavelengths. Due to the fact that the fundamental transitions have the highest absorption coefficient, using mid-IR wavelengths also has the advantage of providing the highest sensitivity. Lead-salt lasers (Bomse et al., 1992) have been extensively deployed for these wavelengths, although these have mostly been displaced by quantum cascade lasers. QCL-based absorption sensing has been demonstrated, for example, for measurement of N_2O and CH_4 around $\lambda = 8 \mu\text{m}$, yielding a sensitivity of 0.25 ppm (Namjou et al., 1998).

The use of resonance overtones, which are mostly found at wavelengths below $2 \mu\text{m}$, on the other hand, allow the use of standard telecommunications lasers or VCSELs at shorter wavelengths, which tend to be less expensive and not require cooling. DFB lasers in the $1.55 \mu\text{m}$ range have been used for oxygen sensing (Weldon et al., 1996), as have the 762 nm VCSELs we saw above. Long wavelength VCSELs, also emitting around $1.55 \mu\text{m}$, have been used for sensing of NH_3 (Totschnig et al., 2003) as well as CO and CO_2 (Kögel et al., 2006).

The laser diodes used have demanding performance requirements. To allow scanning over the, typically small, absorption features, they must operate single-mode, have a narrow linewidth and low noise, and be operated with good temperature stabilization. An example of the kind of optical microsystem which is advantageous for this application is shown in Figure 10.39, in which a 4×4 VCSEL array is mounted directly on a micro-thermoelectric cooler with an integrated thermistor for monitoring the substrate temperature and thereby allowing fast, low-power temperature regulation; temperature stability of better than $\pm 0.1 \text{ }^\circ\text{C}$ is typically required.

As an example, the VCSELs used in the measurements of Figure 10.38 had a linewidth of about 3.3 MHz and relative intensity noise (RIN) of about $3 \cdot 10^{-13} / \sqrt{\text{Hz}}$ above laser threshold, in the frequency range 5 – 20 kHz (Zappe et al., 2000). Not to be underestimated is the role of the current source used to drive the laser, which must also have very low noise.

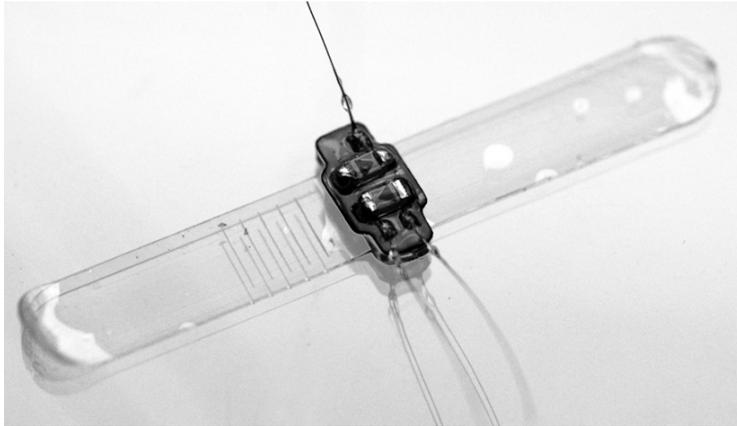


Figure 10.40 An implantable optical absorption sensor for measurement of blood oxygenation; the LED emitters are shown mounted on a 4 mm wide silicone stripe which is wrapped around a blood vessel. Photo courtesy of Jens Fiala and Niklas Weber.

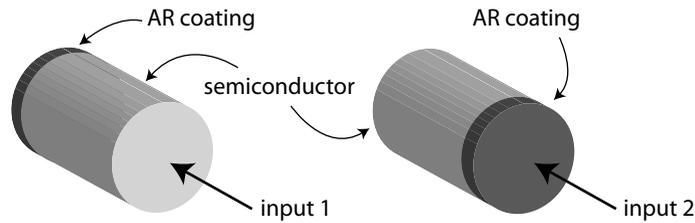
Spectroscopy in liquids

Absorption spectroscopy is employed in a wide variety of configurations for sensing of a large palette of compounds. Wavelength-dependent optical absorption in liquids is also of considerable utility, for example in determination of the composition of blood. Various molecular components of blood have strongly varying absorption spectra in the visible and near-IR; in particular, derivatives of hemoglobin, including oxy-hemoglobin, deoxygenated hemoglobin, carboxyhemoglobin and methemoglobin have different absorption coefficients in the wavelength range between 700 and 1 000 nm, such that optical transmission measurements in this wavelength range can be used to determine, for example, the level of blood oxygenation using pulse oximetry (Reichelt et al., 2008).

An example of an optical microsystem for performing blood oxygenation measurements *in vivo* is shown in Figure 10.40, in which a pair of LEDs is mounted and encapsulated on a silicone stripe, which is intended to be wrapped around a blood vessel; a second stripe with a photodetector completes the system. This measurement configuration uses a slightly different approach than that used for gas sensing using discrete absorption peaks, since hemoglobin and its derivatives have a broad absorption spectrum. Instead, absorption is measured at two or more wavelengths, in correlation with the pulse, and the intensity differences used to determine, for example, blood oxygenation; the entire sensor is implanted and the measurement can be performed continuously, important for high-risk patients. The concept has been extended to 8 LED wavelengths and can also be performed in reflection from perfused tissue (Fiala et al., 2009).

Problems

1. Estimate the efficiency of an incandescent light bulb; i.e., what fraction of emission is at visible wavelengths? Compare the value you obtain with that you expect for an LED.
2. Given a nominally 300 μm long GaAs-based laser with a (broad) gain peak at 862 nm, estimate the uncertainty in the lasing wavelength if the cavity length is only known within an accuracy of 2%. Assume $n = 3.65$ for this wavelength.
3. Determine the finesse of a GaAs-based Fabry-Perot laser cavity without any reflection coatings on the facets.
4. Derive an expression for the mode spacing of a Fabry-Perot laser cavity, including the effects of dispersion.
5. Using the result of the previous problem, or a suitable estimate, determine the mode spacing for a 200 μm long GaAs laser cavity? For a 100 μm long cavity? A 1.5 μm long cavity? Assume emission at 862 nm.
6. Based on the results of the previous problem, what are the advantages of a short-cavity laser? What are the disadvantages?
7. Given a GaAs-based laser with the gain characteristics shown in Figure 10.2, a cavity length of 500 μm and a residual background loss of $\alpha_0 = 5 \text{ cm}^{-1}$, calculate the threshold gain and estimate the wavelength of the first mode which lases. How many modes could theoretically lase at this pumping level?
8. Re-calculate the emission efficiency of the LED structure of Example 10.2 for a thinned substrate which has been etched to 10 μm below the pn junction and compare the values obtained with those for the standard un-etched substrate.
9. Estimate the coherence lengths of the 950 nm LED of Figure 10.7 and the DFB laser of Figure 10.16.
10. Estimate the overall efficiency (output photons per input electron) at an operating current of 50 mA for the AlGaInP laser whose PI characteristic is shown in Figure 10.12 at $T = 5$ and 45°C .
11. Given a DBR laser reflector region defined in a $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{Al}_{0.7}\text{Ga}_{0.3}\text{As}$ waveguide, estimate the grating period required for a third-order Bragg reflector for an 850 nm laser. You know that refractive index for $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ is 3.45 and for $\text{Al}_{0.7}\text{Ga}_{0.3}\text{As}$ is 3.18.
12. Plot the output characteristic of a photodetector (current versus optical input power) with efficiency $\eta = 0.8$, for $\lambda = 405, 514, 633, 850 \text{ nm}$.
13. Plot the IV characteristic of a photodiode with a reverse bias leakage current of 1 nA under illumination of 0, 2 and 5 nW optical power. Assume an efficiency of 0.85 and operation at room temperature.
14. You have two cylindrical photodetectors, configured as shown in the figure below: one is made from silicon ($n = 3.44$), the other GaAs ($n = 3.71$). The detectors have a length L and a diameter of 100 μm ; one facet has a (perfect) AR coating.



Couple a 1 mW, $\lambda = 800$ nm signal into the non-coated facet, shown on the left. Calculate the generated current density, assuming that internal $\eta = 1$, for $L = 1, 10, 100$ μm .

15. Now turn the previous photodetectors around and couple the signal into the side with the coated facet, as shown on the right; repeat the calculation, ignoring any internal reflections.
16. Plot the interference characteristic for a AlGaAs-based Mach-Zehnder-based interferometric intensity modulator operating at 852 nm using the QCSE; modulator length is 2.5 mm. Assume that the 1 μm thick waveguide has a refractive index of 3.3 and that the QCSE results in $r_{QCSE} = 19 \cdot 10^{-12}$ m/V. What is the value of V_π ?
17. Repeat the previous problem for the case that the modulator has a parasitic electroabsorption component, with $\alpha = 35$ $\text{m}^{-1}\text{V}^{-1}$, and the Mach-Zehnder is unbalanced, leading to a zero-bias phase shift of $\pi/5$. Compare your results with those found in the previous problem.
18. Design a waveguide phase modulator using GaAs for operation at $\lambda = 1.55$ μm , at which wavelength $n = 3.06$ for this material. The waveguide is 1 μm thick and the material has a breakdown field of $E_{BD} = 4 \cdot 10^5$ V/cm. What is the shortest modulator length you can use to achieve $\Delta\phi = \pi$ using the Pockels and Kerr effects while avoiding breakdown? What is V_π in both cases?
19. Sketch the structure of a simple black & white LCD display and explain how it works. What is the structure and where does the light come from if there is no backlight?

11 Micro-optical fabrication

*A fool, Mr. Edgeworth, you know, is a man who never tried an experiment in his life.*¹

Erasmus Darwin, 1792

One of the distinguishing characteristics of micro-optics is the means by which micro-optical components and systems are fabricated. Classical optical elements, predominantly lenses and prisms, have been manufactured for centuries using well-established and highly-refined grinding and polishing processes. As the size of lenses decreases below the millimeter range, however, these techniques reach their limits: grinding a lens with a diameter below 1 mm is difficult and economically impracticable. As a result, micro-optics uses completely different fabrication techniques, namely those based on semiconductor manufacturing.

Semiconductor fabrication technology is extremely highly developed, and the factories which produce acres of processed silicon likely represent the most advanced manufacturing technologies on the planet. Whereas most of this industrial capacity is still used for generating the electronic chips on which much of human civilization relies, the past two to three decades have seen an expansion of these technologies into other realms, notably microelectromechanical systems (MEMS) (Menz et al., 2001; Senturia, 2001; Korvink and Paul, 2006; Liu, 2006; Gianchandani et al., 2007), a field which has diversified to include micromechanics, microfluidics, microacoustics and micro-optics. These sub-fields have benefitted from semiconductor technology, not only in that extremely small structures may be defined but also that these may be mass-fabricated, in marked contrast to the fabrication of individual components which characterizes classical optics. Thus microsystems components, including micro-optical ones, may be made in quantities of thousands as easily as individually, and the resultant economies of scale contribute to the success of microsystems in a wide variety of new applications.

Most micro-optical devices are not off-the-shelf components, such that most micro-optical components tend to be custom-designed and fabricated for a particular application. As a result, we discuss here a few of the fabrication methods used to accomplish this (Herzig et al., 2005; Suleski and Te Kolste, 2005). We begin with a look at surface-tension-based lenses and consider several specialized forms of lithography as well as

¹ Erasmus Darwin (1711–1802), English polymath and poet, grandfather of Charles Darwin. Quoted in (Krause, 1880), p. 102.

graded-index structures. Subsequently, we focus on replication as key technology for fabrication of high-quality micro-optics in large quantities and then turn to a few specialized processes of increasing relevance.

11.1 Basic semiconductor processing

As semiconductor process technology is a very broad area which is taught as part of many (particularly electrical) engineering curricula, we will not rehash the field here, but refer to the reader to some of the excellent weighty tomes which consider the subject in breadth and depth (Madou, 1997; Nishi and Doering, 2000; Campbell, 2001). To allow those readers who have no exposure to processing technology to follow some of the specialized topics which follow, we will briefly summarize the essence of semiconductor manufacturing techniques here, for which the literature provides more in-depth resources. Most microsystems or EE students are well versed in this.

11.1.1 General concept

Semiconductor manufacturing generally employs silicon wafers, such that most of the processes are planar; as a result, most micro-optics is relatively flat, as we saw in the discussion of Section 7.6.1. The fabrication processes are designed to generate large numbers of identical components simultaneously, in a succession of process steps performed on one or more substrates in parallel. After completion of planar processing, the individual components are separated into chips and packaged. This type of mass-fabrication is key to making micro-components economically viable, and is one reason that micro-optics has found several market niches.

Whereas the semiconductor world is predominantly silicon, with an occasional excursion into the relatively exotic III-V materials, micro-optics also makes extensive use of glasses, polymers and other types of semiconductors or dielectrics. In addition, even though packaging for electronics is an important post-process step, the assembly and packaging issues for microsystems and micro-optics are often considerably more complex and demanding; we consider these in greater detail in Section 11.6 below.

11.1.2 Essential processes

Semiconductor fabrication processes generally consist of a long list of relatively simple steps which are performed and often repeated in a well-defined sequence; typical processes may consist of hundreds of sequential steps. Whereas any single step may be considered “simple” in its effect, the extreme demands on the accuracy of its results usually implies that each such step was subject to extensive development, and the equipment required can be very complex and expensive.

Substrates

The standard starting substrate for electronics is usually the silicon wafer, with diameters up to 300 mm (Tsuchiya, 2007). Micro-optical fabrication also employs Si (most research cleanrooms use 100 or 150 mm wafers) but also relies heavily on glass, often Pyrex, fused silica (SiO_2) or any one of the specialized glasses, such as BK7, considered in Section 3.4. Compound semiconductors are of interest for active micro-optical components (Dharmarasu and Hillmer, 2007) and polymer substrates are also occasionally used, although most polymers are subsequently deposited on other substrates (Selvaganapathy, 2007).

Cleaning

Due to the very small structures defined and fabricated in microsystems processes, cleanliness is of the utmost important. Fabrication is usually performed in the isolated environments of cleanrooms and the substrates are repeatedly cleaned using chemical or plasma means, often after every step, to remove any contaminants generated during a previous process. Many fabrication failures or low yields may be attributed to inadequate cleaning.

Lithography

Definition of the structural patterns on the substrate is done by means of lithography, in which a photo-sensitive layer (typically photoresist) is exposed and developed, leaving covered and uncovered areas on the substrate; electron beams or patterned stamps may also be used for this step, as we will see below. Again due to the small sizes involved, photolithography generally requires extremely advanced equipment to define sub-micrometer patterns reliably on the substrate surface. Definable structure sizes vary from 30 μm using cheap foil masks to several nanometers using electron beam equipment.

Deposition

Any one of a multitude of materials may be deposited as layers onto the substrate. Dielectrics (such as SiO_2 , Si_3N_4 or a host of others); metals (particularly Al, Au, Ti, Pt or Cr); semiconductors (Si, Ge, or III-Vs, either as amorphous films or epitaxially); glasses; or polymers may be defined as thin films. Deposition technologies also vary: evaporation; oxidation; sputtering; chemical vapor deposition (CVD) or low pressure chemical vapor deposition (LPCVD); plasma-enhanced CVD (PECVD); and spin or dip coating are widely employed. Layer thickness and composition are usually of primary importance, the tolerances of both a function of the process employed; thicknesses of several tens of nanometers to several micrometers are typical.

Etching

Surface layers may also be removed using wet (chemical) or dry (plasma) etch processes, usually in the regions uncovered by photoresist as defined by the photolithography step. Depending on the material intended to be etched, with possible requirements for selectivity over lower-lying layers, a wide spectrum of wet chemical etchants may be

employed. Plasma-based processes include plasma etching (in a barrel etcher); reactive ion etching (RIE) or reactive ion beam etching (RIBE) (generally using a F or Cl-based chemistry); or ion-beam milling (IBM) (typically using inert Ar ions). As an alternative to etching, photoresist liftoff may be used to define surface layers by depositing them on patterned photoresist structures, which are then removed by dissolution, leaving the material deposited in the exposed areas.

Post-processing

Somewhat oxymoronic in its terminology, post-processing refers to everything which happens after the fabrication of the individual devices, including chip separation (using cleaving, sawing, or laser dicing); passivation (often with polymer surface layers); electrical wire bonding (using ball or wedge bonders); mounting and packaging. Since these are often “dirty” processes, they are usually done after removal of the substrates from the cleanroom, sometimes giving them an aura of being less technologically advanced than the previous steps; nevertheless, post-processing is not only essential but also often represents a significant fraction of the time and costs of a complete manufacturing cycle.

Assembly

Less of an issue for electronics fabrication, the assembly of microsystems or micro-optical components into a complete system often follows post-processing and is frequently as complex, important and expensive as the microfabrication steps (Malshe and O’Conner, 2005; Brand, 2007). A considerable topic in itself, micro-optical system assembly will be discussed in Section 11.6, where we will outline some of the important concepts involved.

The essentials of microfabrication technology as just nanoscopically outlined are those on which the microsystems, MEMS, and micro-optics fields are based, whereas of course a myriad process variations are employed. In the remaining sections of this chapter, we will look at some of the processes which have evolved from this technology for specific use in micro-optics fabrication.

11.2 Self-assembly lenses

Surface tension in liquids can form excellent spherical surfaces and these, if the liquids are transparent, lead to natural microlenses. We will see these effects again when we discuss tunable liquid lenses in Chapter 12, but surface tension has long been used as a means to easily produce microlenses and lens arrays.

The effects we consider here fall under the heading of “self-assembly”, since the forces of nature (surface tension, in this case) define the optical surface, rather than having it defined by a controlled manufacturing process. Self-assembly effects are attractive since, once the required boundary conditions are set, the device is formed without any further explicit structuring, and this repeatedly and reproducibly. As a result, self-assembly effects, when they apply, can strongly simplify micro-optics processing.

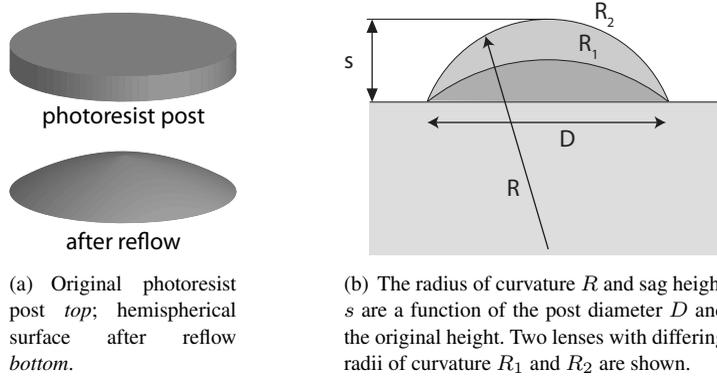


Figure 11.1 The high-temperature reflow of a cylindrical post of photoresist causes the cylinder to assume a hemispherical shape and thus a segment of a spherical lens.

11.2.1 Photoresist reflow lenses

One of the most established and simple means to fabricate refractive microlenses is by photoresist reflow. Developed in the 1980s, this process uses the controlled melting of cylindrical photoresist posts, where surface tension results in a hemispherical liquid surface which, upon cooling, becomes a solid microlens (Daly et al., 1990; Hutley, 1997; Daly, 2001). This process works well for lens diameters less than about 1 mm.

The process is illustrated schematically in Figure 11.1. Cylindrical photoresist posts with diameter D are defined photolithographically on a suitable substrate, usually glass or a semiconductor. From basic geometric considerations, we can relate the sag height of the lens, s , to the aperture diameter D and radius of curvature R , as defined in Figure 11.1(b), as

$$s = R - \sqrt{R^2 - \frac{D^2}{4}}. \quad (11.1)$$

Since the radius of curvature ultimately defines the focal length, and the two variables that the lens designer has at her disposal are the initial photoresist thickness, t_{PR} [m], and the photolithographically-defined lens diameter D , we can use conservation of photoresist volume to determine the required initial photoresist thickness. The volume of the photoresist cylinder, V_{cyl} [m³], is given by

$$V_{cyl} = \pi \left(\frac{D}{2} \right)^2 t_{PR} \quad (11.2)$$

and the volume of a hemispherical lens, V_{hemi} [m³], by (Råde and Westergren, 1999, Chapter 3.2)

$$V_{hemi} = \frac{1}{3} \pi s^2 (3R - s). \quad (11.3)$$

Since photoresist volume is conserved during reflow,

$$V_{hemi} = V_{cyl}, \quad (11.4)$$

which leads to the required initial thickness of the photoresist

$$t_{PR} = \frac{4}{3} (3R - s) \left(\frac{s}{D} \right)^2. \quad (11.5)$$

By solving Equation 11.1 for R and entering the result into this last expression, we are left with a result which is only a function of sag height and aperture diameter,

$$t_{PR} = \frac{s}{6} \left[3 + 4 \left(\frac{s}{D} \right)^2 \right]. \quad (11.6)$$

The designer thus has simple geometric relationships between photoresist thickness, lens diameter, sag height and radius of curvature to allow the design of a lens with the desired characteristics.

Example 11.1: How does the required height of the photoresist post vary with desired focal length of a photoresist lens? We can use the above relationships to plot the variation of t_{PR} with focal length for various aperture diameters.

We recall (Equation 7.19) that for a plano-convex lens such as these we are considering here, radius of curvature R and focal length f are related as

$$R = (n_L - 1) f$$

where photoresist refractive index is typically on the order of $n_L = 1.56$. From Equation 11.1, we thus have

$$s = R - \sqrt{R^2 - \frac{D^2}{4}} = (n_L - 1) f - \sqrt{[(n_L - 1) f]^2 - \frac{D^2}{4}}$$

which we can then in turn plug directly into Equation 11.6.

Evaluating the above for focal lengths ranging from 100 μm to 1 mm for aperture diameters of 50, 100 and 150 μm , the required initial thickness of photoresist for the cylindrical posts, t_{PR} , varies as shown in Figure 11.2. We see that, not unexpectedly, short focal lengths require a thick photoresist layer, the thicker the larger the aperture. Initial photoresist layer thicknesses larger than about 10 μm are sometimes difficult to achieve, so we cut off the plot at that value. We also see that control of focal length for small apertures is difficult: the $D = 50 \mu\text{m}$ curve is very flat for focal lengths above about 300 μm , so precise control of initial thickness is essential.

Process

Fabrication of reflow microlenses is relatively simple and straightforward. Standard lithography is used to define photoresist cylinders with diameter D and resist thickness t_{PR} on an appropriate substrate, usually a glass such as BK7, silicon, or fused silica. In an oven or on a hotplate, the substrates are then subjected to a high-temperature

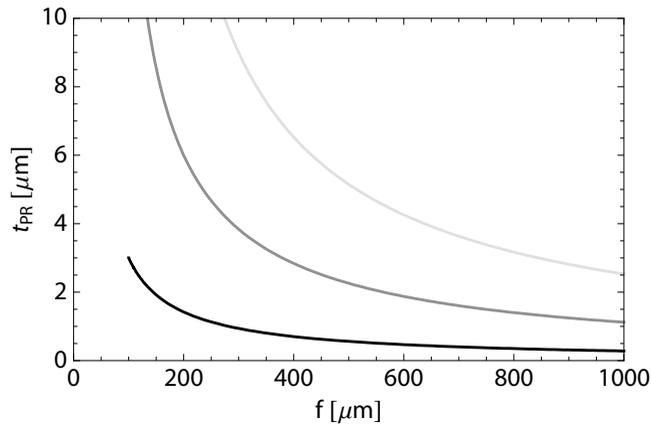


Figure 11.2 Variation of required photoresist post height t_{PR} with desired focal length of photoresist reflow lenses using a resist with $n_L = 1.56$; the plots are calculated for lens aperture diameters (cylindrical post diameter), D , of 50 μm *black*; 100 μm *gray*; and 150 μm *light gray*.

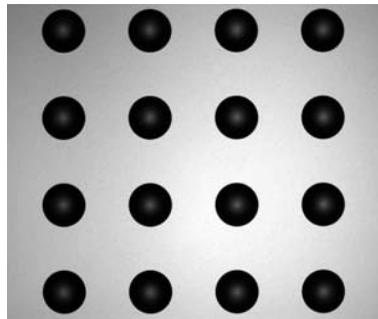


Figure 11.3 Microscope image of an array of 300 μm diameter photoresist reflow lenses, fabricated on glass. Photo courtesy of Wolfgang Mönch.

sequence, ramping to the glass transition temperature of the photoresist, usually in the range 100 – 140 °C; the heating and cooling ramps generally need to be precisely defined to achieve controlled and reproducible reflow.

The lens shape is then defined by surface tension conditions and thus the contact angle at the photoresist/substrate/ambient interface; the contact angle is a function of photoresist viscosity and temperature, recalling that surface tension generally decreases with increasing temperature. For complete melting, the original shape of the cylinder is irrelevant: surface tension always yields a hemisphere. Photoresist “sculpting” is nevertheless possible, by limited melting of binary or continuous-profile structures defined by gray-scale masks, for example; the profiles can be smoothed without completely reflowing the resist.

The resulting photoresist hemispheres can be used as spherical microlenses themselves, for wavelengths longer than about 650 nm, because photoresist absorbs strongly for shorter wavelengths. Since photoresist is also rather delicate, the more popular

approach is to use the resist as an etch mask and transfer the lens shape into the underlying substrate. Using a directional dry etch, such as RIE, RIBE or IBM, the profile may be transferred into the underlying substrate, etching away the photoresist layer in the process. The result is a refractive microlens etched into the surface of the substrate, as we saw schematically on the right side of Figure 7.38. Alternatively, the photoresist can be used as the starting profile for a mold used in replication processes, as we will see in Section 11.4.

Performance and limitations

Typical reflow microlenses have diameters below 500 μm and sag heights below about 20 μm ; since initial photoresist thicknesses greater than 100 μm are usually not possible, the range of possible radii of curvatures for larger aperture lenses is limited. A photoresist reflow microlens with $R = 240 \mu\text{m}$ and sag height $s = 2.6 \mu\text{m}$, for example, will have a focal length of about 400 μm , which is typical for this type of lens. Numerical apertures range from $0.15 \leq NA \leq 0.45$ and $f/\#$ is typically $f/1$ to $f/3$. Due to the high quality of the liquid surface, roughness is very low and aberrations are generally below $\lambda/4$.

11.2.2 Microcontact printing

The use of surface effects to promote self-assembly for microlens fabrication can be extended to other materials and given considerably greater flexibility through the use of microcontact printing. Also known as “soft lithography” (Wilbur et al., 1994; Xia and Whitesides, 1998) this process uses an elastomeric mechanical stamp, impregnated with a wetting or other coupling agent. Using this stamp on a glass, semiconductor or polymer substrate, a wettability pattern may be printed on the surface, thus defining hydrophobic and hydrophilic regions on the substrate. By subsequently depositing liquids on this patterned surface, these are self-assembled onto the regions with the appropriate wettability, forming structured patterns of the liquid (Wilbur and Whitesides, 1999).

Microcontact printing may be used for the fabrication of microlenses and lens arrays, again relying on the surface tension of a liquid to define the hemispherical shape of the liquid lens surface (Biebuyck and Whitesides, 1994; Hartmann et al., 2001). This process is highly-parallel and requires only a single master structure, which may be used to replicate a large number of stamps which in turn may be stepped repeatedly on a surface. Patterning on planar and non-planar substrates is possible and the process is both low-cost, high-resolution and relatively fast.

Process

Microcontact printing relies on replication (see Section 11.4 below) to fabricate the stamp which is used to print surface patterns. As shown in the process outline of Figure 11.4, a master is generated, generally in silicon, using standard photolithography. Employing a molding process, a stamp (with the negative relief pattern) is generated, usually using soft materials such as silicone elastomers. This stamp is then “inked”

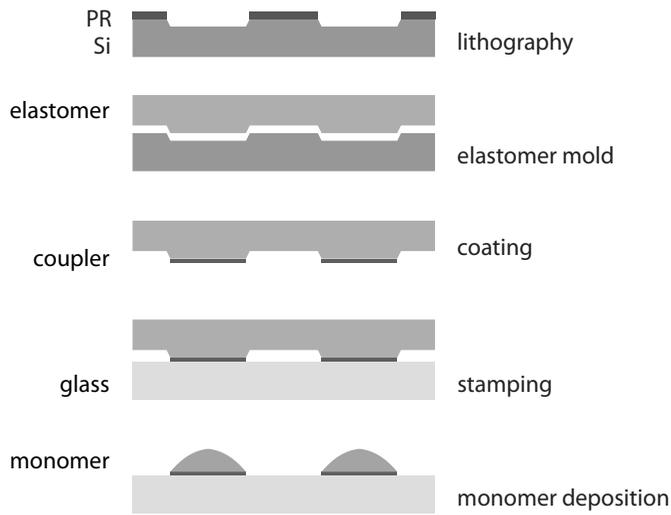


Figure 11.4 Microcontact printing process for fabrication of self-assembled microlenses: *top to bottom* origination of a master in Si using standard lithography; molding the elastomer stamp; coating the stamp with a hydrophilic coupling agent; stamping on a glass substrate; lenses subsequently formed in a monomer by surface tension. Not to scale.

using the desired coupling agent (either by depositing the agent on the stamp or using what amounts to a stamp pad) and then stamped on the substrate to be patterned; typical coupling agents are dimethyl-monochloro-silane or octadecyl-trichloro-silane, which generate hydrophobic surfaces on SiO_2 .

For fabrication of microlenses, circular or elliptical patterns are thus stamped onto the surface. The substrate is then coated with a monomer layer, such as 1,6-hexanediol-dimethacrylate or an optical adhesive, using dip-coating or a similar process (Mönch and Zappe, 2004); the liquid monomer forms hemispherical caps on the hydrophilic areas, yielding microlenses in the same manner as for photoresist reflow lenses. The monomer is then cured, typically using UV-exposure, to form solid polymer microlenses. Although fabrication of the stamp requires a lithography step and the lens characteristics may vary depending on which method is used to deposit the monomer, this self-assembly lens fabrication process is relatively simple.

Lens characteristics

An array of microcontact printed microlenses is shown in Figure 11.5; these lenses have an aperture diameter of $90 \mu\text{m}$ and a focal length of $870 \pm 30 \mu\text{m}$. Typical refractive indices for the monomers employed are 1.56 for optical adhesives and 1.458 for 1,6-hexanediol-dimethacrylate. The lenses are generally characterized using scanning white-light interferometry to determine the overall profile, and a motorized microscope stage can be used to determine the focal length for the lens under illumination. Focal length varies with monomer deposition method, but typical ranges are 200 to $400 \mu\text{m}$ for optical adhesives and 800 to $1\,200 \mu\text{m}$ for other monomers. Aperture diameters of $10 \leq D \leq 100 \mu\text{m}$ are easily achieved, yielding $4 \leq f/\# \leq 13$.

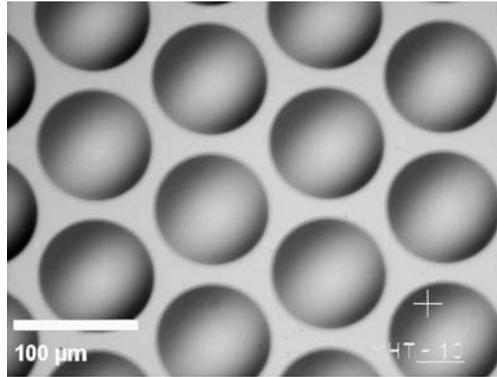


Figure 11.5 Microcontact printed microlenses, formed in poly-alkyl-methacrylate; lens diameter is 90 μm . Photo courtesy of Wolfgang Mönch.

Due to the very smooth liquid surface before curing, this type of microlens has very low surface roughness, generally 10 to 30 nm. Aberrations can be measured using wave-front analysis, in which the Zernike coefficients are determined from the details of the lens profile as measured by interferometry. Spherical aberration tends to dominate this type of microlens, and the value is a function of aperture; both positive and negative SA is seen, with values ranging from 0 to about 200 nm. Coma and astigmatism are generally below $\lambda/10$ for visible wavelengths.

11.3 Lithography for microlenses

The microlenses based on self-assembly processes require only rudimentary or no photolithographic definition; nature, in the guise of surface tension, forms the lens automatically. Refractive and diffractive micro-optical structures may, however, also be defined photolithographically. In contrast to the strictly two-dimensional patterning of standard photolithography, most micro-optical structures require a variation in the pattern in the third (vertical) direction. As a result, a number of techniques have been developed to yield fully three-dimensional structures.

The optical profiles are defined in photoresist by the methods we consider here. As for the photoresist reflow lenses, one can use the photoresist itself as an optical component, but more robust structures result if the profile is transferred vertically into the glass or semiconductor substrate by dry etching or used as the first step in the definition of a master for replication. Thus, whereas we confine our discussion here to the origination of the profile, fabrication of the final optical structure usually requires further processing (Hutley, 1997).

11.3.1 Direct-write lithography

Continuous-profile micro-optics may consist of standard refractive lenses, Fresnel microlenses, Fresnel-like diffractive/refractive structures as well as DOEs and kino-forms; they all have in common that the optical surface is defined by a continuous relief. Photolithography as employed for semiconductor manufacturing can be used to define continuous surface profiles; even though photoresist and its processes are typically optimized for digital response (the resist is either developed away or not), techniques have been developed to generate intermediate responses using either electron beam (e-beam) (Kress et al., 1993) or optical lithography (Gale, 1998).

E-beam approaches

Electron beam exposure of photoresist, while also traditionally optimized for digital response, has been used to make high-resolution continuous-profile optical structures (Fujita et al., 1982). By adjusting the electron dose and through accurate calibration of the photoresist response and development characteristics, features with resolution down to 3 - 5 μm have been defined in thick photoresist and used to defined Fresnel-like microlenses (Haruna et al., 1990). Although the e-beam spot size may be as small as 10 nm, proximity effects limit the achievable resolution to several micrometers.

Using electron-beam-sensitive resists, such as PMMA, the exposure beam is either scanned across the surface using a vector-scan approach, or the stage is moved with respect to a stationary beam. Both the beam spot size and shape may be varied using electron optics, providing flexibility in the exposure pattern. Development depth is then a function of dose, such that continuous profiles with typical depths of up to 5 μm have been demonstrated.

Laser approaches

Electron-beam systems have high resolution but also a high price tag, such that these are frequently impractical for both research laboratories and manufacturing facilities. An alternative is laser direct-writing of continuous profile optical structures using an intensity-modulated laser beam (Haruna et al., 1987; Gale et al., 1994). As for the electron beam, the response of the photoresist to exposure dose and development characteristics have to be well characterized; the substrate is then mechanically scanned under the variable intensity beam. A HeCd laser emitting at 442 nm is generally used, its intensity modulated using an acousto-optical modulator allowing up to 256 gray levels. The 1.5 μm diameter spot can then be scanned with a write speed of up to 10 mm/s, as shown schematically in the setup of Figure 11.6.

Positioning accuracy and optical spot size limit the minimum feature size of laser direct-written optical structures to about 5 μm and the depth-of-field of the projection optics limits the usable photoresist height to about the same level, again 5 μm ; the performance of this optical approach is thus not inferior to the electron-beam variant. Optimized exposure and development processes can yield very smooth optical surfaces, with roughness below 25 nm; the Fresnel-like lens of Figure 8.28 is fabricated using this approach.

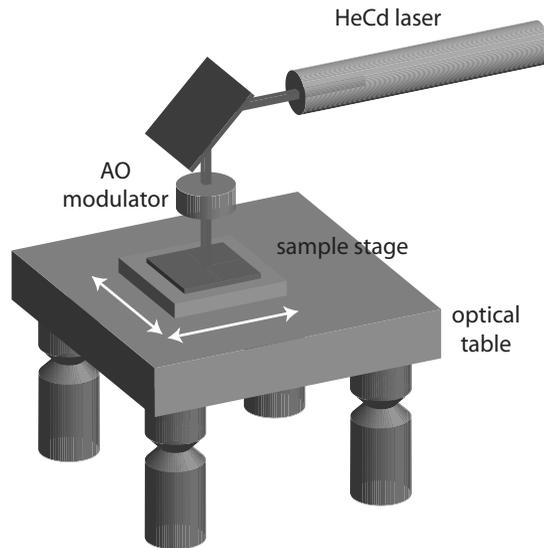


Figure 11.6 A schematic depiction of a laser direct-write system for generating continuous-profile micro-optics. Laser emission is modulated in intensity by the acousto-optical modulator, and the photoresist-coated substrate scanned underneath the beam. The entire setup must be mechanically isolated from vibrations using an optical table.

11.3.2 Gray-scale lithography

An alternative to the sequential process of direct-writing continuous-relief profiles is the use of gray-scale lithography, which retains the advantages of a single flood exposure as is commonly done in standard photolithography. In gray-scale lithography, the patterns on the photomasks are not “digital” (i.e., either transparent or opaque) but use arrays of closely spaced dots to generate effectively “gray” regions, thus producing a continuous range of transmission values (Oppliger et al., 1994). Again using well-calibrated photoresist exposure and development characteristics, the continuous range of exposure values results in continuous-relief profiles in photoresist.

Gray-scale masks can be used in standard contact aligners or projection printers, but require a customized optical system with limited MTF or spatial filters such that the individual dot patterns on the mask are not imaged but rather become a gray region (with differing densities) when projected onto the substrate. The dot-based masks usually provide 8 gray levels with a resolution better than 4 μm . Gray-scale masks may also be made using variable-intensity exposure during mask fabrication (Suleski and O’Shea, 1995), using means similar to those employed for direct-write lithography in the mask-making process.

A further alternative for the generation of gray-scale masks is through the use of High-Energy Beam Sensitive (HEBS) glass, which has spatially-defined variable optical density. HEBS glass is a speciality material whose optical absorption may be modified by electron beam exposure (Däschner et al., 1997); the glass is manufactured using Ag^+ ion exchange (see Section 11.5.2 below) and the silver ions are reduced (thus becoming

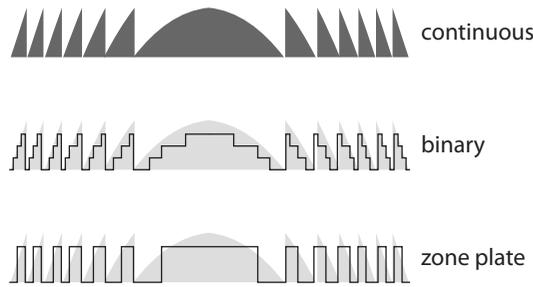


Figure 11.7 The binary approximation of a continuous Fresnel lens profile *top* using a multi-level *center* or two-level *bottom* structure; the latter corresponds to a Fresnel zone plate.

absorbing) when exposed to an electron beam. The general concept is similar to that used for traditional photographic film but at considerably higher resolution, typically 5 to 8 μm . Photomasks based on HEBS glass may then be used in gray-scale lithography exposure systems, and have been used for the definition of two-dimensional continuous-relief optics as well as three-dimensional structures (Mosher et al., 2009).

11.3.3 Binary optics

As we mentioned in Section 8.4.3, as the dimensions of microlenses shrink, the curved surface of the lens becomes increasingly difficult to define; this problem occurs for most refractive lens structures, including Fresnel-like lenses and kinoforms. As a way around this limitation, the shape of a continuous-relief structure may be approximated by a stepped profile, generated by discrete photolithography steps, a fabrication concept known as binary optics (Veldkamp and Allen, 1983; Stern, 1997). Thus a Fresnel-like lens (which we discussed in Section 8.4.2), for example, may be approximated during fabrication in the manner shown in Figure 11.7.

Concept

Using multiple lithographic exposures, the continuous profile is approximated by a multi-level stepped structure, as shown in the middle sketch of Figure 11.7. The number of achievable steps, or phase levels, P , scales with the number of required lithography steps N as

$$P = 2^N. \quad (11.7)$$

As result, the number of phase levels increases rapidly with each step; in the example shown, $N = 2$ so that there are 4 phase levels. The number of phase levels ultimately defines the diffraction efficiency of the structure; for a given P , the efficiency of the first diffraction order is given by (Stern et al., 1991)

$$\eta = \frac{\sin^2\left(\frac{\pi}{P}\right)}{\left(\frac{\pi}{P}\right)^2}. \quad (11.8)$$

The efficiency increases with the number of phase levels (as does, of course, the fabrication complexity) and is highest for the paraxial case, for rays parallel to the optical axis, with angle $\theta = 0$. η decreases to zero for $\theta \rightarrow \theta_{max}$, where θ_{max} defines the NA of the lens. While a continuous profile can reach efficiencies $\eta \approx 1$, a system with $P = 4$ as shown in the example typically has $\eta \approx 0.8$. If only one photolithography exposure step is made, the resulting lens has $P = 2$ as shown in the bottom example of Figure 11.7, yielding a Fresnel zone plate.

N photolithography exposures result in $P = 2^N$ phase levels, or steps in the profile; as a result, $N - 1$ alignment steps are required such that the demands on alignment tolerance, as we will see below, also increase rapidly.

Fabrication

For continuous-profile micro-optics, such as the Fresnel-like lens, the thickness of the optical layer (the surface feature height) is a critical parameter. To generate a phase shift of 2π , the thickness of a layer with refractive index n_L has a maximum value of

$$t_{max} = \frac{\lambda_0}{2\pi(n_L - 1)} \quad (11.9)$$

for vacuum wavelength λ_0 . t_{max} defines the zone height of a Fresnel-like lens, such that, for P phase levels, the height of each phase step is given by

$$t_{step} = \frac{t_{max}}{P} = \frac{\lambda_0}{2\pi P(n_L - 1)}. \quad (11.10)$$

Figure 11.8 shows how these levels are achieved in a typical binary optics process. Three masks are used in this example ($N = 3$), so that, after three exposure and two alignment steps, eight phase levels ($P = 8$) are generated. Each mask has a pattern roughly twice as large as the previous one, so that half of the features generated are transferred to a lower level by the subsequent etch step. The total lens height is t_{max} , with steps of height t_{step} , as shown. The structure is generated by repeating the following steps:

1. preparation of the substrate, typically glass or a semiconductor;
2. spin-on of photoresist;
3. alignment to the previous structure (not necessary for the first step);
4. exposure and development;
5. etch to a depth of t_{step} ;
6. strip photoresist;
7. repeat from step 2.

Tolerances

We recall from Section 8.4.2 that, for a Fresnel-like lens, the radial (lateral) position of zone m is given by (Equation 8.79)

$$x_m = \sqrt{\frac{2m\lambda_0 f}{n_L}}. \quad (11.11)$$

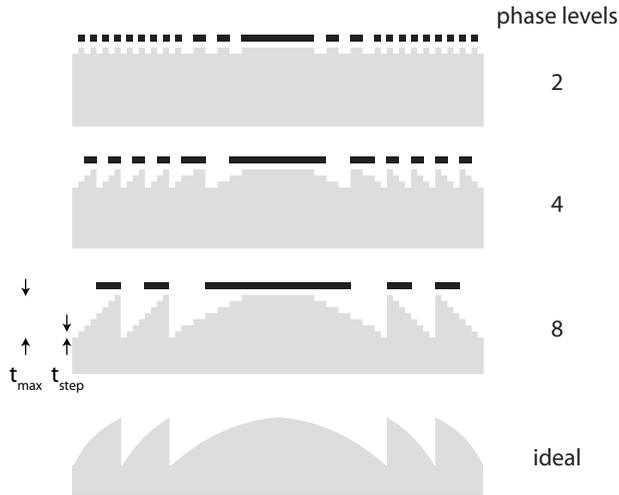


Figure 11.8 Photolithography steps to generate a Fresnel-like lens using the techniques of binary optics. Three lithography steps, using three masks (*black*) are shown: *top to bottom* 1 step yields 2 phase levels; 2 steps yield 4 phase levels; 3 steps yield 8 phase levels; the ideal structure is shown at the bottom.

so that the spacing between zones is given by

$$x_{m+1} - x_m = \sqrt{\frac{2\lambda_0 f}{n_L} (\sqrt{m+1} - \sqrt{m})} = \Delta x. \tag{11.12}$$

The spacing Δx , which decreases as m increases, defines the smallest required mask dimension. Typical lateral tolerances are on the order of 100 to 250 nm, thus providing a limit on the size and number of Fresnel zones in a binary optical approximation. Mask linewidth errors of 0.1 μm or alignment errors of 0.3 μm may lead to reductions in η by 10%. Etch depth tolerances are also in the 100 nm range, where etch depth errors of $\pm 2.5\%$ can reduce η by 1%. Total etch depths may be up to 10 μm , important for longer-wavelength applications.

Example 11.2: We wish to design a Fresnel-like lens in BK7 glass ($n_L = 1.5168$) for use with a green HeNe laser, with $\lambda_0 = 543 \text{ nm}$; focal length should be 100 μm . If the lateral mask and alignment resolution is 1 μm , what is the maximum number of zones and what is the aperture diameter and $f/\#$ of the lens?

Given that $\Delta x = 1 \mu\text{m}$, we can determine the zone spacing from Equation 11.12. Solving this equation for m , we obtain

$$m = \frac{n_L \Delta x^4 - 4f \Delta x^2 \lambda_0 + \frac{4}{n_L} f^2 \lambda_0^2}{8f \Delta x^2 \lambda_0};$$

if we enter our minimal achievable value for Δx into this expression, we obtain the maximum attainable value for m . Using the values given above, we calculate, taking the integer value, $m = 17$: the Fresnel-like lens can thus have a maximum of 17 zones.

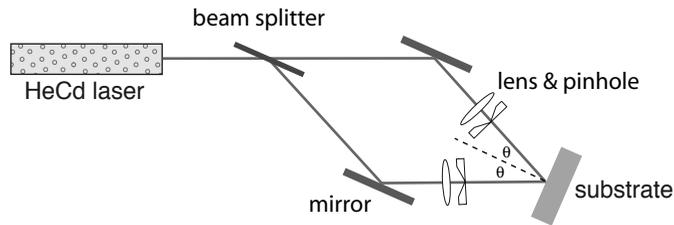


Figure 11.9 Layout of an interference lithography system used for definition of diffraction gratings. Emission from the laser is split into two arms, the beams widened and cleaned using the lens and pinhole, and these generate an interference pattern on the photoresist-coated substrate.

Using Equation 11.11, we find that the position of the last zone (# 17) is at $35.1 \mu\text{m}$ from the center, and using Equation 11.12 again, we determine that its width is just slightly over our minimum resolution, namely $1.02 \mu\text{m}$. The sum of the last two values is then the radius of the lens, so that we finally have

$$D = 2(35.1 + 1.02) \mu\text{m} = 72.2 \mu\text{m}$$

for the lens aperture. From this we can easily find

$$f/\# = \frac{f}{D} = 1.38$$

which is a reasonably good value for a lens of this type.

11.3.4 Interference lithography

Likely the oldest and most established lithography process used for micro-optics is that of interference lithography, sometimes somewhat incorrectly referred to simply as holography. Nevertheless related conceptually to the process for generating classical holograms, interference lithography has primarily been used for the fabrication of photoresist-based diffraction gratings, these patterns then also often transferred by etching into an underlying substrate. Interference lithography is attractive since it is a highly-parallel flood exposure process which may be employed over large areas.

The basics of an interference lithography system are shown in Figure 11.9. Emission from a short-wavelength laser (HeCd, with $\lambda_0 = 442 \text{ nm}$, or Ar, with $\lambda_0 = 468 \text{ nm}$, lasers are popular) is split into two paths, each directed through a large-aperture lens to expand the beam and a pinhole to act as a spatial filter, removing high-frequency aberrations in the beam and generating a Gaussian profile. The two beams recombine at the position of the photoresist-coated substrate, generating an interference pattern which usually takes the form of a series of lines of well-defined period, and thus a one-dimensional grating; the structure of Figure 8.22 was fabricated in this manner and interference lithography is one of the most popular means to generate diffraction gratings.

Based on the laser wavelength λ_0 and the incidence angles of the two beams θ , the period of the interference pattern Λ is given by

$$\Lambda = \frac{\lambda_0}{2 \sin \theta}; \quad (11.13)$$

thus by changing the angles of the arms, the generated grating periods may be varied over a considerable range, using a single laser for exposure. With wide beam expansion, interference lithography can be used to expose large areas with high resolution, and patterns with areas of up to 4 800 cm² have been fabricated using this approach (Gombert et al., 2004).

By appropriately adjusting the relative positions and orientations of the two interfering beams, curved and chirped gratings may also be originated using an interference lithography exposure system. Through the use of three-or four-beam setups, three-dimensional photonic crystals have been defined using a single-shot exposure from an appropriately configured setup (Campbell et al., 2000); we will see an example in Section 14.2.3.

Example 11.3: What incidence angles are required to expose a grating with a period of $\Lambda = 500$ nm using a Nd:YAG laser, assuming we can frequency double and frequency triple the wavelength?

We can simply apply Equation 11.13 using the wavelength of the Nd:YAG laser, $\lambda_0 = 1\,064$ nm. We see immediately, of course, that we can never obtain a period below $\frac{1\,064 \text{ nm}}{2} = 532$ nm.

Plotting the variation of Λ with interference angle θ in Figure 11.10, we see that frequency doubling (to 532 nm) or tripling (to 355 nm) the Nd:YAG laser yields a 500 nm period interference pattern for 32° and 21°, respectively. We see from this simple example why short wavelength lasers are desirable for interference lithography and that the nature of the exposed pattern is sensitive to the geometric arrangement of the interference lithography setup.

Near-field holography

A variation on the interference lithography theme is near-field holography (Suleski et al., 1999). This technique is essentially interference lithography using a computer-generated phase mask designed so that all the incident energy is directed into the +1 or -1 diffraction orders; placed in a standard contact aligner, these masks generate periodic patterns with a period one-half of the structure on the mask. As a result, periods of 0.5 μm and feature sizes of 0.1 μm have been demonstrated using this approach, which may be extended to the definition of circular as well as linear patterns.

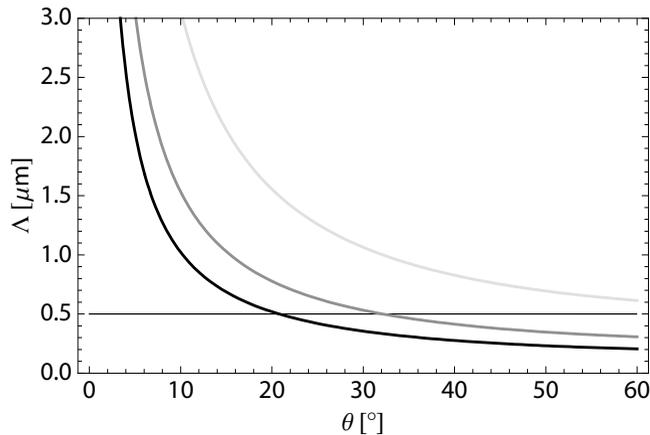


Figure 11.10 Variation of the grating exposure period Λ with beam incidence angle in an interference lithography setup, for a Nd:YAG laser (1 064 nm) *light gray* as well as frequency-doubled (532 nm) *gray* and tripled (355 nm) *black* Nd:YAG emission.

11.4 Replication

In the previous sections, we saw numerous means for generating micro-optical lenses and other devices, either refractive, diffractive or mixed structures, primarily using soft materials such as monomers, polymers or photoresist. These approaches generate the optical component directly, either individually or in arrays. One of the most attractive features of micro-optics, however, is the ability to mass fabricate optical structures by replication: using precisely defined masters, the components are replicated using molding techniques. Replication, which is used on large scale industrially for the manufacture of CDs and DVDs, allows the fabrication of complex microlenses, DOEs and other micro-optical devices in very large volumes at relatively low cost, and represents a key micro-optical process technology.

11.4.1 Molding and embossing

The basic concept behind replication is given in Figure 11.11. A precisely-defined mold is used to define a three-dimensional pattern in a soft substrate, usually a polymer. Molded optics is not new (Miller et al., 1951) and has long been used to manufacture plastic macro-optics, often with astonishing quality; the cheap plastic objectives used in disposable film cameras (popular before digital cameras became ubiquitous) were highly-refined, aspheric objective lenses fabricated by high-volume injection molding.

Mold generation

One of the lithography processes we described above, either classical (two-level), binary, continuous-profile or holographic, is used to generate a pattern on a photoresist-coated substrate, typically glass. We show a two-level pattern (such as that of the surface of a CD or DVD) in the example of the figure, but continuous relief lens structures

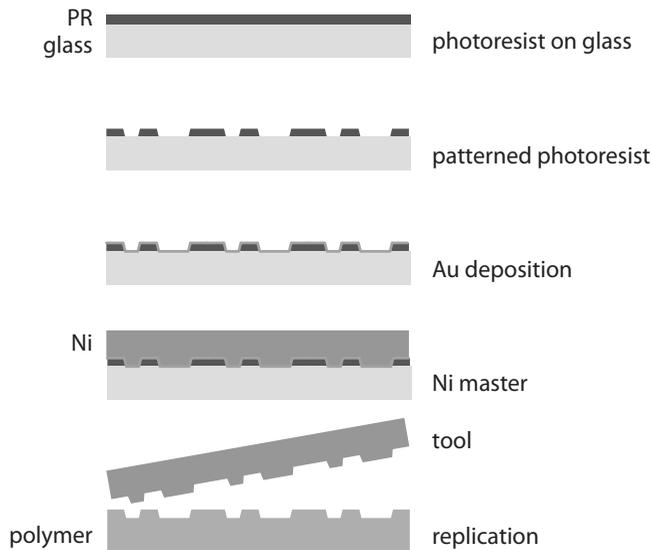


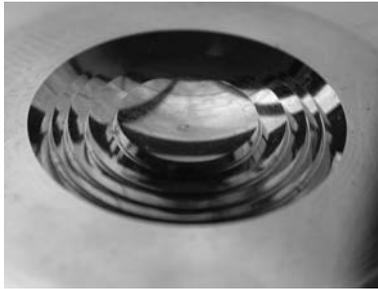
Figure 11.11 Basics of replication: *top to bottom* photoresist on a substrate, typically glass; exposed pattern, binary or continuous relief; deposition of Au contact layer; electroplating a Ni master; replication in a polymer using the resultant molding tool.

(such as those defined by photoresist reflow) may also form the original surface pattern. A gold contacting layer is then evaporated onto the photoresist pattern and a thick nickel layer is deposited using electroplating. This Ni shim is subsequently separated from the glass and represents the mold master.

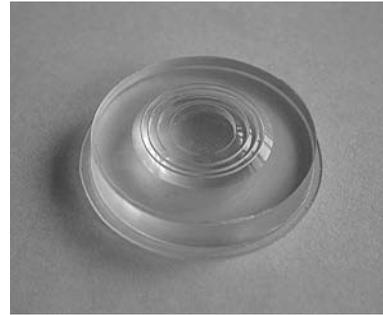
In industrial replication processes (as those for optical disc manufacture), the electroplated Ni master is used to generate a further inverse structure, a replica referred to in the industry as a “mother”. This “mother” is subsequently used to define a third, inverse of the inverse, mold, the “father”, which has the same profile as the original and then serves as the replication tool in the actual molding process. For structures which are fabricated in volumes of hundreds of thousands or millions, wear becomes a factor in the quality of the molded components, and using several generations of molds obviates the need to eventually fabricate a new master, which is typically the most expensive part of the process.

A variety of other techniques is occasionally employed for definition of the master. Ultra-high-precision machining, in brass or nickel, can yield surfaces with optical quality, but is usually limited to circularly-symmetric structures; an example of a machined mold for a micro-Fresnel lens is shown in Figure 11.12(a). Also employed is LIGA², which uses x-ray lithography for the generation of high-aspect-ratio structures which are subsequently used in molding processes (Wallrabe et al., 2007).

² LIGA is a German acronym, *Lithografie, Galvanoformung, Abformung* (lithography, electroplating, molding).



(a) A master mold for replication of a Fresnel lens, machined into brass. Diameter = 25 mm.



(b) Injection-molded discrete Fresnel lens, fabricated using the mold on the left; diameter of the inner lens structure (the Fresnel zones) is 25 mm.

Figure 11.12 Molding tool and replica of a discrete Fresnel lens, fabricated by injection-molded using polycarbonate. Photos courtesy of Michael Engler.

Injection molding

Once the molding tool has been defined, one of several possible processes is used to replicate the optical structures, usually into a thermoplastic polymer (Heckele and Schomburg, 2004). The highest surface quality is achieved with injection molding, the standard process used for manufacture of optical discs and much macroscopic plastic optics. In this process, a liquid polymer, typically polycarbonate (PC), PMMA, polystyrene (PS) or poly vinyl chloride (PVC), is injected into a heated mold cavity with the structured tool on one surface. After cooling below the glass transition temperature of the polymer, the mold, consisting of two or more sections, is then opened and the optical structure removed. Done industrially, this process can achieve cycle times of a few seconds.

Injection molding can yield optical structures with resolved surface features down to 10 nm; structure heights can be greater than 1 μm (Brenner et al., 1993); an example of an injection-molded discrete Fresnel lens is shown in Figure 11.12(b). The technique is commonly used for diffractive components such as DOEs and gratings, as well as waveguides and Fresnel optics. Using PVC as a lens material, these optics are usable in visible to mid-IR wavelength ranges, up to $\lambda = 2.6 \mu\text{m}$.

Embossing

In an embossing process, the heated mold is pressed into a solid polymer surface; at the high temperatures employed, the polymer deforms and attains the inverse form of the tool. The use of release films helps the de-molding process and the technique is typically done either using a flat bed or on rollers, the latter for long continuous films. The same polymers as for injection molding (PC, PMMA and PVC) are used for embossing, although sol-gel materials (which we discuss in Section 11.4.3 below) may also be embossed. Cycle times vary, but can be on the order of minutes; typical mold temper-

atures are in the range of $80 - 140^\circ$ (Tsurumi et al., 2005). Figure 7.45 of Chapter 7 showed a nice example of an embossed Fresnel lens in PMMA.

Whereas the resolution of hot embossing is also in the range of 25 nm, the achievable aspect ratio of the embossed structures is limited when compared to injection molding; structure heights are usually below 1 μm . However, embossing can be considerably cheaper (per area fabricated) than other molding techniques, so it is often used for gratings, DOEs and holograms and low cost diffractive foils; the popular metallic silvery iridescent wrapping paper is, for example, a simple diffractive structure fabricated hundreds of meters at a time using roller embossing.

UV casting

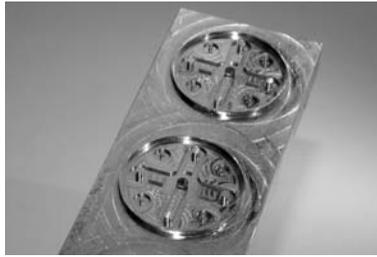
The most industrially-relevant replication process for large-scale micro-optical components is at present UV casting (Rudmann and Rossi, 2004). This process uses a two-component resin (epoxies, a class of synthetic thermosetting polymers) deposited into a mold (Pietarinen et al., 2008). After being sealed by a glass substrate, the epoxy is cured using UV exposure, usually in the wavelength range $375 \leq \lambda \leq 395 \text{ nm}$, for several tens of seconds to minutes. UV casting allows large feature heights (typically $> 1 \mu\text{m}$) with high resolution, and is done industrially using up to 300 mm Si wafers as mold masters. Due to its high resolution, this technique is commonly applied for the replication of DOEs and other complex optical structures, as well as gratings (Tsurumi et al., 2005) and large-area microlens arrays for displays (Huang et al., 2008b) and can also be used for double-sided optical components.

11.4.2 Other micromolding techniques

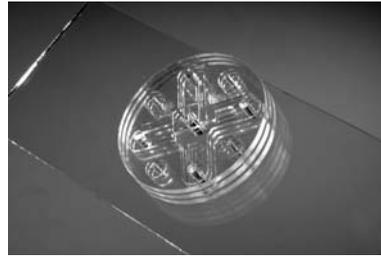
There is a wide variety of other molding techniques available, some developed for high-precision replication of three-dimensional components, many of which are applicable to micro-optics (Heckele and Schomburg, 2004). An example of combined optical and fluidic structures fabricated by molding of a silicone elastomer (PDMS) is shown in the photos of Figure 11.13: a machined aluminum tool (Figure 11.13(a)) is used to define optical surfaces and microfluidic channels for a tunable achromatic microlens system, yielding the molded parts shown in Figure 11.13(b).

Of considerable interest is nano-imprint lithography, which uses a mold (for example, e-beam structured SiO_2 on Si) to emboss a pattern directly into a photoresist, such as PMMA (Chou et al., 1996); pattern sizes down to 6 nm have been “printed” in this manner (Chou et al., 1997). A variant on this technology, reminiscent of UV casting, uses a photo-curable resist, set by UV exposure after application of the mold, with which 5 nm patterns on a 14 nm pitch have been demonstrated (Austin et al., 2004).

Although the definition of features on this scale is less of an issue for micro-optics, nano-contact printing can be used for definition of high-resolution, short-period grating structures without the need to employ expensive serial lithography tools, such as e-beam, except for the definition of the stamp. In addition, nano-imprint lithography may be applied over an existing surface topography if a soft, pliable stamp is used. Although the technology is still in development (Schift, 2008), applications such as the



(a) Machined aluminum tool; the diameter of the circular structure is 24 mm.



(b) A stack of three molded parts in PDMS made using the tool at the left. At the center is a 2 mm diameter microlens.

Figure 11.13 Machined molding tool and replicated micro-optical/microfluidic structure; the components are part of a tunable achromatic microlens system. Photos courtesy of Daniel Mader.

replication of sub-wavelength optical structures or printing of DFB laser gratings have been demonstrated (Viheriälä et al., 2009).

Direct molding of aspheric refractive microlenses (Kuo and Lin, 2008) has also been demonstrated in SU-8, a photostructurable polymer popular for MEMS fabrication due to its ability to generate thick layers (up to 100 μm) and steep sidewalls (Conradie and Moore, 2002). Using a cured SU-8 stamp, liquid SU-8 heated to above the glass transition temperature is stamped, and the temperature ramps are used to define the lens profile. Microlenses with aperture diameters ranging from 50 to 500 μm , with NA up to 0.58, have been defined using this replication technique.

Finally, most replication processes, particularly those relevant for micro-optics, involve molding of thermoplastic materials, usually polymers. Since glasses represent materials of considerable interest for optics, including micro-optics, molding of glass lenses is also relevant (Hug et al., 2005). However, molding of glass requires significantly higher temperatures (up to 1 000 $^{\circ}\text{C}$) and pressures (up to 50 kN), such that specialized equipment with tight control of temperature and pressure ramps is required. In addition, the tools are typically made of Mo, ceramics or Si. While the molding of test microstructures in borosilicate glass has seen promising results (Schubert et al., 2006), the technology is still under development for micro-optical components (Kleer et al., 2003).

11.4.3 Sol-gel microlenses

We learned in Section 3.6.3 that sol-gel materials are essentially glasses which can be formed from liquid pre-cursors. Since their composition and refractive index can be controlled during fabrication, sol-gels have proven useful for the fabrication of microlenses (Yuan et al., 2005), diffraction gratings (Blanc et al., 1999) and for fluorescent dyes used in lasers (Lo et al., 1993).

Depending on their composition, two different types of sol-gels may be defined: inorganic and organic/inorganic hybrids. The hybrids are attractive since one may incorpo-

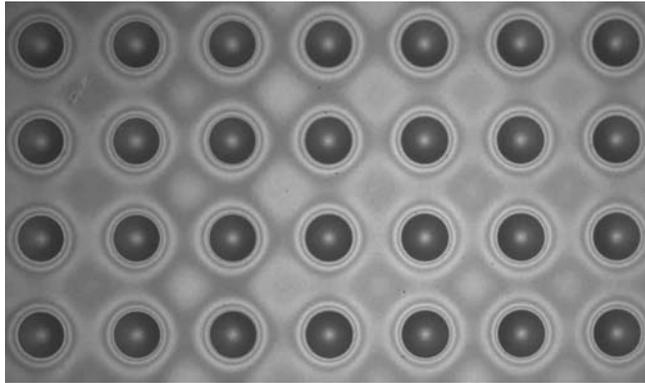


Figure 11.14 An array of 200 μm diameter molded sol-gel microlenses on a glass substrate. Photo courtesy of Holger Krause.

rate photo-initiators into the material, thus making them suitable for photolithographic structuring, employing binary or phase-shift masks (Kärkkäinen et al., 2002), or using laser direct writing (Yu et al., 2002); the useful spectrum is limited, however, since they absorb strongly near the UV. Hybrid sol-gel processes are also used to form microlenses in quartz glass (Kreuzberger et al., 2009)

Inorganic sol-gels, which have good transparency for wavelengths below 300 nm, have been used to fabricate replicated microlens arrays using a molding process (Krause et al., 2006). Lens profile origination can employ photoresist reflow, ink jetting, or gray-scale lithography: one of these approaches is used to generate a mold, using a silicone elastomer such as PDMS. These molds, which represent the negative (hollow) profile of the lens, are filled with the sol, sealed with a glass plate and subsequently cured using UV exposure, in a manner similar to UV embossing.

Microlenses with diameters ranging from 40 to 400 μm , radii of curvature in the range $30 \leq R \leq 1\,300 \mu\text{m}$ and sag heights of typically 28 μm were fabricated using this approach. These sol-gel materials have refractive indices on the order of 1.42 to 1.44 (at $\lambda = 633 \text{ nm}$), such that a focal length range of $60 \leq f \leq 4\,100 \mu\text{m}$ was achieved, corresponding to $1.4 \leq f/\# \leq 12.8$. Using sol-gel materials has the advantage that the optical components are fabricated directly in robust glass-like material and the molding processes allow inexpensive generation of lens arrays. However, the sol-gels often suffer from shrinkage and cracking of the materials during curing, such that precise definition of the required dimensions can be difficult; nevertheless, well-developed sol-gel processes are promising for the manufacture of a variety of micro-optical components.

11.5 Specialized processes

As a glance into the contemporary and historical literature will show, there is a colorful spectrum of micro-optical fabrication ideas, schemes and procedures which have been proposed and demonstrated, some more of academic interest and others industri-

ally applied. We take a look at a small selection of these, encouraging the reader to occasionally browse through technical publications for further inspiration.

11.5.1 MOEMS

As we mentioned in the introduction to this chapter, much of micro-optical fabrication technology is based on that developed for microelectronics and subsequently expanded to embrace the demands of MEMS. Whereas the MEMS field was originally limited to microelectromechanical systems, as the name stipulates, the acronym today refers to the fabulously interdisciplinary spectrum of microsystems in general, a flavor of which may be gleaned from perusal of, for example, (Gianchandani et al., 2007). Microsystems today may include electrical, mechanical, magnetic, acoustic, chemical, biological and, of course, optical components.

The combination of micro-optical with micromechanical components and/or the fabrication of micro-optical devices using silicon-based fabrication technology, has led to the designation MOEMS (micro-opto-electromechanical systems), optical MEMS (OMEMS), or occasionally MOMS (micro-opto-mechanical systems) (Motamedi, 2005). The dividing lines between MOEMS and micro-optics in general are, as we might expect, quite vague, and much of what we have presented in this text could just as well have been classified as MOEMS. Even though a precise categorization is irrelevant, we look at a few salient aspects of what is considered trademark optical MEMS.

Fabrication

MOEMS technology was originally based on silicon micromachining as it was developed for MEMS technology (Wu, 1997); the increasingly wide spectrum of available fabrication techniques which involve other materials and processes, including polymers, glasses, metals and ceramics, has also broadened the technology base for optical microsystems. Silicon-based machining processes (Lang, 1996) may be divided into two categories: surface and bulk micromachining. Surface micromachining typically employs structured surface layers, particularly polysilicon and dielectrics, and results in roughly planar devices. Bulk micromachining, in contrast, uses wet or dry etch techniques to etch partially or entirely through a silicon wafer, allowing the manufacture of three-dimensional semiconductor structures.

Many of these processes are relatively complex, a fact which has, however, never been an obstacle for the dedicated graduate student. For research laboratories lacking the appropriate infrastructure, foundry processes are available which allow MEMS and MOEMS designs to be fabricated professionally; SUMMiT (Jorgensen and Yarberry, 2001) and MUMPS (Markus et al., 1995) are two popular surface micromachining foundry services³. Many micromachining processes are now designed with CMOS-compatibility in mind, such that hybrid integration with electronics is possible for the completed system, a increasingly important consideration for complete systems. Finally,

³ SUMMiT (Sandia Ultra-Planar Multilevel MEMS Technology) is a service provided by Sandia Laboratories; MUMPS (Multi User MEMS Processes) uses processing facilities at various sites in the USA.

III-V-based MOEMS are also becoming more relevant, primarily with the intention of generating tunable active components, such as multi-wavelength laser diode arrays.

Components and systems

MOEMS often combine micromechanics with micro-optics. These systems use various means of mechanical actuation, of which electrostatic (Toshiyoshi, 2007), magnetic (Oh and Ahn, 2007) and thermal (Que, 2007) actuators are the most relevant in MEMS systems. A wide variety of optical components has been subject to actuation, including mirrors, lenses, apertures and gratings, some of these configured as silicon-based free-space micro-optical systems (Lin et al., 1994).

We have already seen a number of MOEMS configurations in previous chapters: the renowned DMD (Figures 1.1 and 1.2) in Chapter 1; the adaptive optical mirror (Figure 6.18) and scanning micromirrors (Figures 6.19 through 6.23) of Chapter 6; and the external-cavity tunable laser (Figure 8.33) of Chapter 8 are all devices which combine micro-optical with micromechanical structures and represent in part the apotheosis of this technology.

MOEMS are employed in a similar spectrum of applications as is the case for micro-optics in general; configurations where high functionality and integration density are required tend to be attractive for this technology. Optical switching, which we discussed in some detail in Section 6.6.3, has provided significant impetus for the development of MOEMS (Marom, 2007), primarily using switchable or reconfigurable micromirrors; also of relevance for optical communications are MOEMS-based variable optical attenuators (Lee and Yeh, 2008). Other areas in which MOEMS are playing a significant role are in displays and imaging (Urey and Dickensheets, 2005), optical scanners (Bourouina et al., 2005), endoscopic microscopy and imaging (Dickensheets, 2008; Aljaseem et al., 2008) and even atomic optics (Chuang et al., 2008).

Tunable micro-optics, finally, which uses many of the concepts of MOEMS to realize individually tunable optical components, is the subject of Chapter 12.

11.5.2 Ion-exchange

As has diffused into our consciences after reading many of the chapters above, much of optics relies on steps or gradients in the refractive index of materials; lenses, diffractive structures and waveguides all depend on a well-defined index distribution. One means to generate continuous refractive index profiles for waveguiding or lensing is through the use of diffusion processes (Borrelli et al., 1986), of which ion exchange is likely to be the most established.

Ion exchange processes are generally employed with glass or LiNbO_3 substrates and are based on an “exchange” of Na^+ or K^+ ions with heavier metal ions which diffuse from the surface; Ag or Ti are typically used (Houde-Walter et al., 1993). The procedure is schematically shown in Figure 11.15: an aluminum masking layer is deposited on the substrate and patterned using standard photolithography; the aluminum is subsequently etched. The thus structured substrate is then immersed in a molten salt containing the ion of interest, for example AgNO_3 or TiNO_3 , usually at temperatures between 200 and

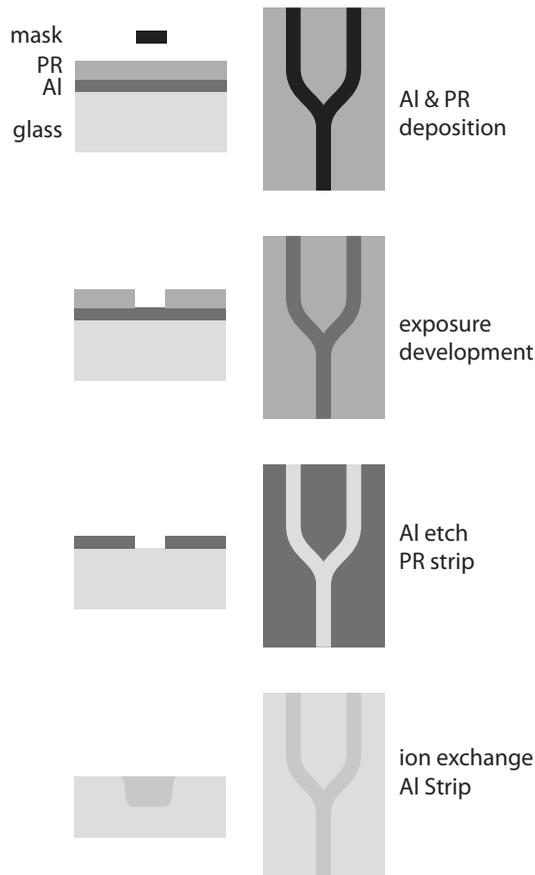


Figure 11.15 A typical ion exchange process, here shown for the fabrication of a waveguide coupler, in cross-sectional *left* and top *right* views. *Top to bottom* deposition of aluminum and photoresist on a glass substrate; exposure and development; Al etch and stripping photoresist; the ion exchange process in a hot salt, and subsequent stripping of the Al masking layer.

230°C. The diffusion process, which takes 2 to 4 hours, results in an incorporation of the ion into the glass, generating a diffusion profile with an increased refractive index.

The ion exchange process is industrially well established for the fabrication of glass-based waveguide circuits (Findakly, 1985). The waveguides typically have a depth of 1 μm , and the 4 μm wide guides are generally single mode since the refractive index increase is only on the order of 0.1 for Ag. The diffusions have, as we expect for this type of process, a Gaussian or erfc^4 profile. One irritating problem in the subsequent alignment and system assembly is that ion exchange waveguides are practically invisible, since the diffusion does not significantly change the absorption properties; since the diffused areas are slightly rougher than the non-diffused areas, one can usually find them by looking for scattering from an oblique light source.

⁴ Complementary error function

11.5.3 Graded index optics

Ion exchange and related diffusion processes are also used for the fabrication of graded-index optics, particularly GRIN lenses (Moore, 1980; Messerschmidt et al., 1996). Rods of glass are immersed in large ion exchange baths to generate a diffusion profile radially toward the center, allowing thus the manufacture of GRIN rod lenses, which we discussed in Section 7.6.2.

These processes are also used to fabricate planar optical lenses, by diffusion into a glass substrate through a circular masked opening. Microlenses with diameters ranging from 10 to 1 000 μm , with a focal length range of $20 \leq f \leq 2\,000 \mu\text{m}$ have been demonstrated using this approach; values of $f/\#2.5$ and numerical apertures of $NA \approx 0.15$ are typical.

Diffusion-based graded index lenses have the advantage that they are planar, making them easily integrable into more complex microsystems. On the other hand, they have strong aberrations at the lens edges, requiring apertures for good performance. In addition, the process times are long, require high temperatures and tend to be expensive, so that these types of microlenses have limited popularity.

11.6 Assembly

Fabricating micro-optical components is one thing; assembling them into a usable system is often quite another and the size scales typical of microsystems lead to difficulties not generally encountered in macroscopic setups. Optical alignment is an issue for all optical systems and is particularly demanding in micro-optics due to the micrometer positioning accuracies required. Also demanding in assembly of micro-optics is the manipulation, positioning, mounting and fixing of the very small components, many of them, as we have seen, no larger than salt grains. A sharp pair of tweezers and a steady hand are usually not sufficient.

11.6.1 Monolithic versus hybrid integration

In the move to increasingly complex optical systems, two conceptual approaches for integration have evolved. Monolithic integration combines all the necessary components on a single, usually semiconductor, substrate and uses a single technology to fabricate all the devices simultaneously. Monolithic optical systems have been extensively developed, primarily for optical telecommunications applications as we saw in Section 9.7.1, usually employing III-V semiconductors. With this approach, all active and passive, as well as electronic, components, can be fabricated on a single chip; an integrated laser transmitter, for example, was seen in Figure 9.40.

Most of these monolithic systems are waveguide-based optical circuits fabricated on InP. Alternatively, planar lightwave circuits using the technologies of silicon photonics have monolithically integrated all but the light-emitting components on a silicon chip. Whereas the appeal of monolithic integration is strong, since it yields highly

compact systems without the need for any post-processing alignment or assembly, it is also highly complex, expensive and subject to low yields. Although a wide spectrum of monolithic systems has been demonstrated, the industrial relevance of monolithic integration is limited; few manufacturers have monolithic systems on the market more complex than a single InP-based laser integrated with an electroabsorption modulator.

Hybrid assembly, the second conceptual approach for optical systems integration, has proven to be more versatile and ultimately economically feasible than monolithic integration. Hybrid systems are assembled using components fabricated using disparate technologies, allowing the optimum material for each device to be used; as a result, hybrid systems can combine silicon and III-V semiconductors with polymers, glass, metals, LiNbO_3 and a host of other materials. However, the alignment, assembly and packaging of these different components, as their long-term stability, are the challenging issues that hybrid integration faces.

For micro-optics, with the exception of waveguide-based integrated optics, most components and assemblies require complex three-dimensional configurations, making monolithic integration less appealing at any rate. As a result, hybrid assembly techniques play a dominant role in micro-optical systems, and we will look at a few key relevant technology steps here.

11.6.2 Active and passive alignment

Aligning optical components to one another is typically the most critical assembly step in any optical system. Due to the small size scales which define micro-optical systems, the required alignment tolerances are in the micrometer to sub-micrometer ranges. Two alignment principles, active and passive, are commonly used to achieve this level of accuracy.

Active alignment

In active alignment, light is coupled into the optical system and the components maneuvered into their optimal orientation, as determined by the measurement of the output field; when optimized, the optics are then fixed into position. The emission from a laser diode, for example, may be optimally coupled into a microlens by monitoring the position of the ensuing focal point.

Active alignment has the advantage that it can lead to high coupling efficiencies, albeit at the expense of considerable effort and relatively long assembly times; the components need to be positioned in roughly the correct place, light needs to be coupled in and out and the positions of the components maintained until they can be secured in place. The alignment procedure usually requires that the component is positioned using a six-axis manipulator (allowing movement in $x, y, z, \theta_x, \theta_y, \theta_z$) and, after alignment, fixing the component using glue, UV-curable epoxy or laser welding. Alternatively, active alignment can be performed by a micro-optical component itself, using a tiltable micromirror, for instance, to optimize coupling from one component to another.

Such active alignment schemes have been used, for example, to couple emission from a VCSEL into an optical fiber, using a thermally-actuated micromirror (Ishikawa et al.,

2003); initial misalignment tolerances of up to 25 μm could be compensated using this approach. Long-term stability of optical systems can also be assured using active alignment, by employing beam steerers to optimally couple light into a free-space optical system over longer time (Boisset et al., 1995). Nevertheless, the overhead required for active alignment has resulted in it remaining more of a niche rather than a mainstream technology.

Passive alignment

In contrast to active alignment, passive alignment uses physical alignment features and occasionally self-alignment processes to arrange the components with respect to each other, usually on a machined semiconductor, metal, glass or ceramic substrate. The physical alignment features, typically etched grooves, edges or posts, behave as mechanical “stops” against which the optical components are positioned. Whereas the maximum achievable positioning accuracy may be lower than that achievable with active alignment, passive techniques tend to be highly parallel, relatively inexpensive and fast. As a result, passive alignment processes are usually preferred for micro-optical systems assembly if at all possible.

As an example, passive hybrid assembly approaches were developed for aligning optoelectronic devices (laser diodes and photodetectors) as well as optical filters on a Si PLC substrate (Himeno et al., 1998). The filter plates could be mounted vertically into the substrate using etched alignment grooves. The approach is also suitable for other substrates, such as low-temperature co-fired ceramic (LTCC), which was used to optically couple a laser diode to a multimode fiber with less than -1.5 dB coupling loss (Hiltunen et al., 2002). Alternatively, a succession of silicon “motherboards” and “daughterboards” has been used as a passively-assembled optical regenerator system (coupling light from an input fiber to a waveguide, then to an optical amplifier and subsequently to an output fiber) in which the subsystems are assembled into a high-functionality microsystem (Maxwell et al., 2002). Such passive alignment approaches have proven to result in hybrid systems sufficiently robust for use in aerospace applications (Beranek et al., 2000).

The best of both worlds (the active and the passive) can also be employed in hybrid active/passive assembly. For example, a fiber can be coarsely aligned using a V-groove and then fine-positioned using a thin-film heater for actuation in the micrometer regime (Datta et al., 2003). We will look at a few technologies which have proven to be useful for facilitating passive alignment techniques below.

11.6.3 V-grooves

Many micro-optical systems use optical fibers for input and output. The alignment of fibers to waveguides, laser diodes and other components is greatly facilitated by V-grooves, V-shaped structures etched into silicon as shown in Figure 11.16. When fabricated with dimensions compatible with the standard 125 μm diameter fiber, the long channels may be used as mechanical guides to position fibers with high accuracy.

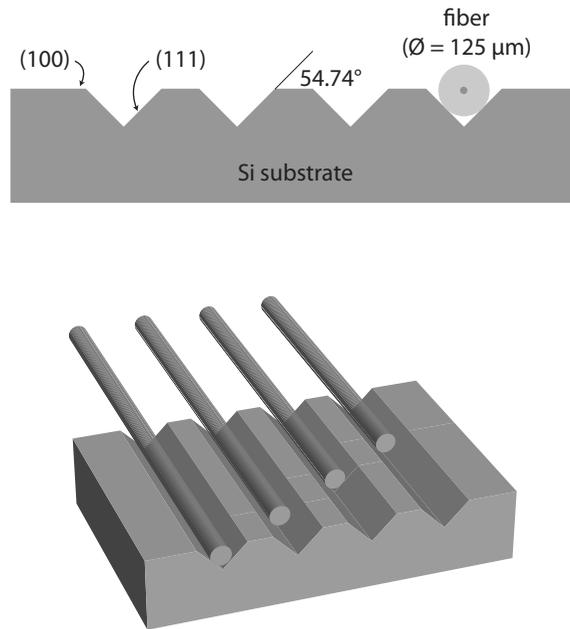


Figure 11.16 V-grooves are fabricated in silicon using an anisotropic wet etch which stops on the (111) crystalline planes; the cross-section *top* is shown for a standard Si wafer with a (100) surface. The width of the grooves is designed so as to position optical fibers (with a standard diameter of 125 μm) at a specified position *bottom*.

V-grooves are generally fabricated using anisotropic wet etching of silicon; the etchants employed have a high selectivity between the (100) and (111) planes of silicon, such that the etch stops on the (111) planes (Bean, 1978). Most popular for this application is KOH, in a temperature-controlled bath, which exhibits a 200:1 selectivity between the two crystal planes. Also used is EDP (ethylenediamine pyrocatechol, $\text{NH}_2(\text{CH}_2)_2\text{NH}_2$, CH_4OH_2 , and H_2O), less popular due to its toxicity and lower selectivity (40:1). V-groove structures have also been fabricated in other materials, notably InP (Bönsch et al., 1998), but silicon remains by far the most popular (Menz et al., 2001, Chapter 6.2.3).

As shown in the figure, the relative orientation of the (111) and (100) crystal planes results in a V-groove angle of 54.74° with respect to the surface, if a (100) silicon substrate is used. For (110) substrates (which are relatively exotic), the anisotropic KOH etch would yield deep vertical sidewalls with an angle of 90° , in certain surface directions. The standard V-groove is designed for positioning 125 μm fibers, generally with a pitch of 250 μm , such that the fiber core is located at the position of the wafer surface. V-grooves with etched clamps which aid in holding the fiber in position have also been developed (Strandman and Bäcklund, 1998). Linear (one-dimensional) V-groove arrays are commercially available, and two-dimensional (stacked) arrays are specialty products. Due to the high quality of the etched facet, V-grooves are also used as mirrors

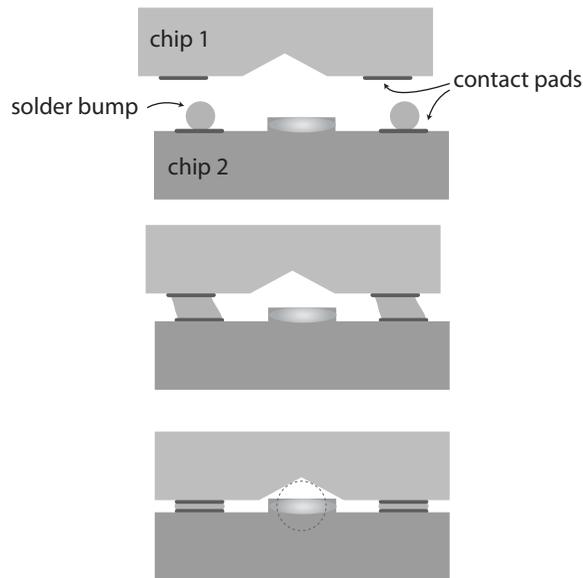


Figure 11.17 The flip-chip process uses surface tension of molten solder to pull two face-to-face chips into alignment. This example shows a V-groove etched chip (chip 1) and another with a waveguide or laser ridge (chip 2). The process *top to bottom* involves deposition of solder balls onto the contact pads; raising the temperature so the molten solder pulls the chips together; resulting in alignment, with the profile of a fiber shown with dotted lines.

for coupling, for example, light from a VCSEL into a fiber, also aligned by a V-groove (Hwang et al., 2006).

11.6.4 Flip-chip bonding

Mounting, alignment and electrical contacting of chips onto a substrate, or onto each other, may be advantageously accomplished by flip-chip bonding, a technique originally developed for electronics packaging (Miller, 1969a). Schematically illustrated in Figure 11.17, the approach uses the surface tension developed by molten solder to pull two chips mounted face-to-face (hence the designation “flip-chip”) into alignment, simultaneously generating electrical, and rudimentary thermal, contact between the two. The approach has proven very useful for hybrid assembly of optoelectronic systems (Wale and Goodwin, 1992).

For a flip-chip process, contact pads are defined photolithographically on the surfaces of two chips to be joined; Au, TiPtAu or CrCuAu are typical metallization materials. Pb/Sn, Au/Sb or In solder balls are then deposited on one set of contacts, and the other chip flipped into rough alignment onto the surface. When the temperature of the substrates is raised above the melting temperature of the solder (typically about 350°), the balls melt and surface tension pulls the two chips into alignment. Positioning accuracies are defined by the accuracy with which the contact pads are defined, and can be better than $1\ \mu\text{m}$. The good electrical contact, coupled with good positioning accuracy,

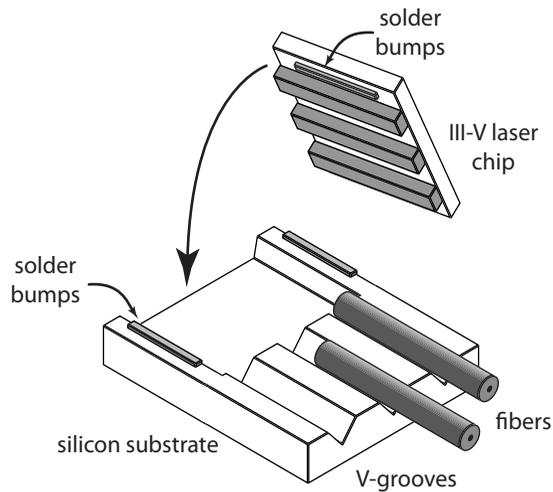


Figure 11.18 The combination of passive alignment techniques, such as V-grooves and flip-chip as shown here, eases the assembly of complex micro-optical systems. The example shows the alignment of an edge-emitting laser array to a linear fiber array.

has also made the technique suitable for coupling optics to electronics, such as VCSEL arrays mounted on CMOS substrates (Krishnamoorthy et al., 1999) or for the assembly and wiring of large LED arrays (Onoe et al., 2009).

11.6.5 Silicon optical bench

We have seen several alignment and assembly techniques which facilitate the assembly of optical microsystems. The advanced state of its etch technology has resulted in silicon becoming an attractive substrate for mounting and aligning micro-optical components, especially since electronics and the optical components on which silicon photonics is based, particularly waveguides, also represent well established Si processes. Using silicon as a mounting substrate for micro-optical components has led to the designation silicon optical bench (SiOB), an allusion to the rather macroscopic optical tables on which optical setups are generally assembled.

The silicon optical bench may include the assembly features we discussed above; in the schematic example of Figure 11.18, a linear array of fibers is coupled to a laser diode array using V-grooves and flip-chip mounting techniques. A SiOB may include more than just alignment structures, however; especially when combined with the technologies of PLCs, the silicon substrate may also include waveguides, electronics, wiring, passive optical structures, detectors and possibly other active optical devices such as thermal modulators (Hashimoto et al., 1998). The PLC and SiOB concepts are similar, where the former tends to focus on integrated optics and the latter on free-space optics, although combinations of the two have of course been demonstrated (Okano et al., 1998).

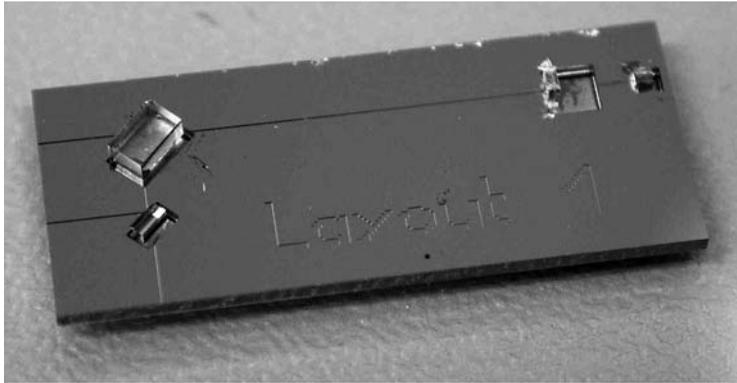


Figure 11.19 A silicon optical bench structure designed for opto-chemical analysis. The 20 mm long silicon chip has a beam splitter *upper left*, another polarizing beam splitter *lower left*, a quarter wave plate *right* and a mirror *far right*. The positioning structures were etched 25 μm into the surface. Photo courtesy of Florian Krogmann.

Application examples

As a result of its flexibility, the silicon optical bench approach has been used for a wide variety of micro-optical applications. The SiOB structure is adapted to the particular demands of any optical implementation, so there is no “standard” silicon optical bench, but rather a host of custom-designed configurations. SiOBs have been configured for mounting lasers with ball lenses (employing deep V-grooves and integrated resistors for signal damping) (Akashi et al., 2004), as a platform for an entire Fourier transform spectrometer, including moving micromirrors, actuators, beam splitters and fiber mounts (Yu et al., 2006) and as a platform for an optical coherence tomography (OCT) system with integrated two-axis micromirror and a mounted GRIN lens (Xu et al., 2008).

An example of an all-optical silicon optical bench is shown in Figure 11.19. The 20 mm long, 525 μm thick Si chip was designed to facilitate the alignment of optical components for a chemical analysis system. As can be seen in the figure, the optical bench contains two beam splitters, a quarter wave plate and a mirror. The alignment features were etched 25 μm into the Si surface using deep RIE processes and the components fixed using UV-curable optical adhesives. This optical system was designed for chemical analysis in liquid, and detected polarization rotation due to an analyte found in the channel between the large beam splitter and the quarter wave plate.

A final example, shown in Figure 11.20, shows that three-dimensional structures are also possible using the SiOB approach. In this case, two input and two output optical fibers are aligned to a 4 mm square Si chip on which tunable optical filters had been fabricated; this filter will be discussed in Section 12.4.2 and is shown in detail in Figure 12.14. As seen in the photo, two pairs of silicon alignment chips, with openings etched through the substrate through which a fiber may be threaded, position the fibers at the front and back sides of the filter chip. This entire structure is then aligned on a further silicon substrate. While not terribly practical, this example demonstrates the structural flexibility attainable with advanced processing technology.

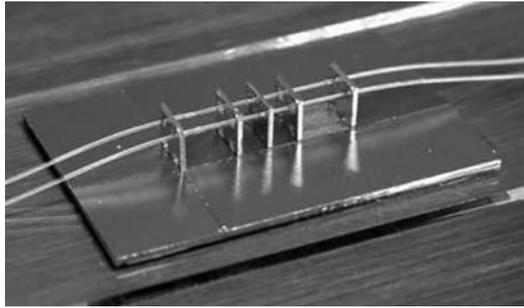


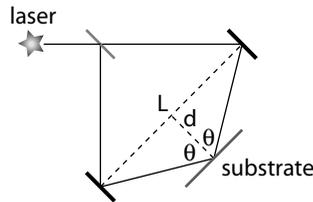
Figure 11.20 A silicon optical bench used for alignment of optical fibers at 90° onto the surface of a silicon tunable optical filter chip (shown in detail in Figure 12.14). Two vertical chips, 4 mm on a side, are used on either side of the central filter chip to position and align the fibers. Photo courtesy of Dennis Hohlfeld.

Problems

Due to the nature of the topics discussed in this chapter, some of the more qualitative problems may require that the student consult the literature to expand her horizons a bit.

1. Derive Equation 11.1.
2. Why are standard photoresist processes unsuitable for fabrication of refractive microlenses?
3. Explain the difference between chemical and physical etching. Is dry etching always a physical etch? Are there processes which are both chemical and physical?
4. Given a photoresist with $n_L = 1.6$, determine t_{PR} for reflow lenses with an aperture diameter of $90\ \mu\text{m}$ and a focal length of $0.4\ \text{mm}$.
5. A cylindrical photoresist post with a diameter of $110\ \mu\text{m}$ and a height of $1.5\ \mu\text{m}$ is subject to reflow. If $n_L = 1.58$ for the photoresist, what is the focal length of the resultant lens?
6. Given a $200\ \mu\text{m}$ diameter photoresist reflow microlens fabricated on BK7 with $R = 750\ \mu\text{m}$, estimate the change in focal length for visible wavelengths between the photoresist and glass lenses if the continuous lens relief is etched 1:1 into the underlying substrate.
7. The same photoresist reflow lens as in the previous example is to be used for IR wavelengths, using a silicon substrate. If the etch rate of Si is exactly twice that of photoresist, estimate the differences in focal length between the photoresist and silicon lenses.
8. Design a continuous-relief refractive lens with a focal length of $500\ \mu\text{m}$ and a diameter of $100\ \mu\text{m}$ for operation at the dark-red wavelength of $656\ \text{nm}$. First use photoresist ($n_{L(PR)} = 1.58$) on BK7 and then BK7 ($n_{L(BK7)} = 1.51432$ at $\lambda = 656\ \text{nm}$) itself. Calculate the radius of curvature and the sag height for both cases.
9. Which of the two lenses in the previous problem would be suitable for use at the green wavelength of $486\ \text{nm}$ and why?

10. Assuming that the BK7 lens designed in Problem 8 for use at 656 nm is instead used at $\lambda = 486$ nm, calculate the longitudinal chromatic aberration. Compare your value for CA with that for a Fresnel lens with the same focal length.
11. For a Fresnel-like lens fabricated in BK7, for which $f/\# = 1.2$ at $\lambda_0 = 633$ nm, would a 10 zone lens be definable using standard optical lithography? How about a 30 zone variant?
12. Given the interference lithography setup shown in the figure, propose typical dimensions for the spacing between the mirrors, a useful exposure wavelength and then plot the exposure period as a function of mirror/substrate spacing d .



13. For a binary optical lens, if $N = 4$, how many phase levels are there and what is the expected η ?
14. Why does the surface of a KOH-etched V-groove make an angle of exactly 54.74° with the (100) surface?
15. Calculate the required mask width of a V-groove etch to place the core center of a $125 \mu\text{m}$ fiber exactly at the height of the wafer surface.
16. Consider a stripe laser monolithically integrated with a passive (unpumped) waveguide made using the same waveguide layers. Is there a problem?
17. What are the sources of optical loss which reduce the coupling efficiency in a hybridly-assembled optical system?

Part III

Neoteric optics

12 Tunable micro-optics

*If the Monkees were doing this show, they'd already be tuned up.*¹

Frank Zappa, 1970

Most optical systems benefit from tunability. The ability to controllably change the focal length, magnification, or orientation of a single component or optical assembly is essential for many applications and is typically accomplished macroscopically using mechanical movement of parts of the system with respect to each other. If we think of focussing a microscope, zooming a camera lens or rotating a mirror for beam alignment, tunability is usually accomplished by twisting a dial, pushing a knob or shifting a fixture.

In the micro-world, a host of effects provides a number of novel means to tune micro-optical components, such that tunable micro-optics has developed into a dynamic research area (Friese et al., 2007; Krogmann et al., 2007). On the one hand, new materials, such as liquids, distensible membranes, liquid crystals and deformable structures, may be used for the fabrication of micro-optical components, and on the other, phenomena such as surface tension, pressure, mechanical deformation and the application of electric or magnetic fields may be employed to tune the optical characteristics. Most of these do not apply to macroscopic systems, so that we have a uniquely micro-optical topic at hand.

We consider in this chapter some of the techniques and devices which have been used to realize tunable micro-optical components and systems. A sizable fraction of the concepts employ liquids, either as droplet lenses or in microfluidic systems, and other soft-matter elements, such that we study a number of these in detail. Liquid-crystal-based components and tunable diffractive structures form the second half of the discussion, and we conclude with a look at the spectrum of what else has been shown to be possible. Extensive references to the literature are provided, since there are as yet no suitable textbooks on the topic.

¹ Frank Zappa (1940–1993), American rock musician, at a concert in The Netherlands in 1970.



Figure 12.1 Surface tension generates a spherical water droplet which acts as an excellent microlens. Note the magnified structure of the leaf in the middle droplet.

12.1 Liquid microlenses

As the optically-attentive reader may have observed by looking at a wet window after a rainstorm, water droplets can make excellent lenses; a natural example is shown in Figure 12.1. The shape of liquid droplet is primarily defined by surface tension; for diameters below about 1 mm, gravitational effects are negligible and the form of the resulting lens is defined by the wetting properties of the liquid on the substrate on which it is placed (De Gennes, 1985). The natural shape of such a droplet is a sphere, and this physical principle was the basis for the self-assembly microlenses of Section 11.2.

We saw in the last chapter that liquid droplets of controlled shape and size are very useful as refractive micro-optical elements; we now change the shape of the spherical droplet dynamically and make it tunable.

12.1.1 Electrowetting

The application of an electric field to a liquid droplet on a surface can change the wetting behavior of the liquid and thus the shape of the liquid surface; this phenomenon is known as electrowetting (Quilliet and Berge, 2001) and forms the basis for tunable liquid micro-optics (Krogmann et al., 2008a). The effect is neatly illustrated in Figure 12.2: a voltage applied between the droplet and the substrate causes the contact angle to increase and thus the radius of curvature of the hemisphere to decrease.

The radius of curvature of the liquid droplet on a surface, which results in a certain focal length, is defined by the contact angle of the liquid, θ_0 [rad], as defined in Figure 12.3. The contact angle for a given configuration is given by the Young equation² (Mugele and Baret, 2005),

$$\theta_0 = \frac{\sigma_{sv} - \sigma_{sl}}{\sigma_{lv}} \quad (12.1)$$

² Named after the same Thomas Young whom we met in Section 5.2.1.

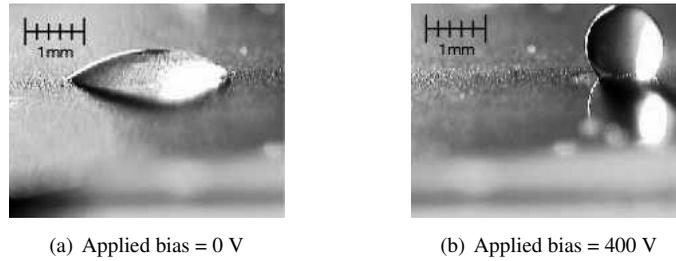


Figure 12.2 Variation of contact angle due to an applied electric field; the liquid droplets are immersed in a second (immiscible) liquid, which is electrically contacted. Photos courtesy of Thorsten Faber.

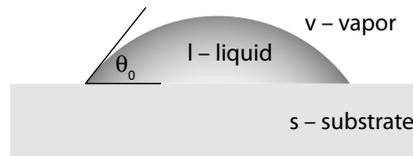


Figure 12.3 Definition of the contact angle θ_0 of a liquid droplet on a surface. The liquid, substrate and vapor regions are defined for determination of the interfacial energies, σ_{sv} , σ_{sl} and σ_{lv} .

where the σ [J/m^2] correspond to the interfacial energies and the subscripts s , v , and l to the substrate, vapor (the medium surrounding the droplet) and liquid, respectively. This expression is based on a constant pressure drop across the liquid/vapor interface and may also be interpreted as surface tension pulling away from the contact line (the point at which liquid/vapor/substrate meet).

If energy is added to this system by, for example, applying an electric field, the electrocapillarity effect results in a change in the energy balance at the interface, leading ultimately to a change in the contact angle. First demonstrated by Gabriel Lippmann³ in 1875 (Lippmann, 1875) for a droplet of Hg in an electrolyte, he showed that the contact angle changes with applied field in a manner characterized by his eponymous equation,

$$\cos \theta_V = \cos \theta_0 + \frac{\epsilon_m}{2t_h \sigma_{lv}} (V - V_{zc})^2; \quad (12.2)$$

in this expression, θ_V [rad], the contact angle which results due to an applied bias V above the potential difference of zero charge V_{zc} [V], is given as a function of the zero-bias contact angle θ_0 , the permittivity of the material (Hg in this case) ϵ_m [F/m], the

³ Jonas Ferdinand Gabriel Lippmann (1845–1921), born in Belgium to French parents, received the Nobel prize in Physics in 1908 for his development of color photography based on a complex interference phenomenon. His PhD thesis, published as (Lippmann, 1875), established his scientific reputation, launching his career as a professor in Paris, where he remained until his death on the ocean liner *France* in the mid-Atlantic.

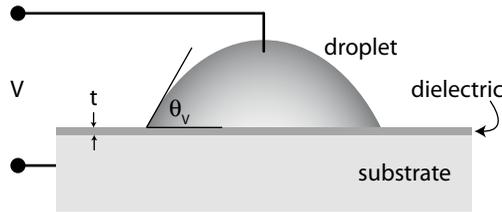


Figure 12.4 Electrowetting-on-dielectrics (EWOD): the droplet is separated from the substrate by an insulating dielectric, of thickness t and permittivity ϵ_d , and a voltage is applied between liquid and substrate to result in a bias-dependent contact angle θ_V . In practice, the droplet is electrically contacted by a liquid electrolyte surrounding it.

thickness of the induced charge layer on the droplet surface t_h [m] and the surface tension (or interfacial energy) between the liquid and the (vapor) ambient σ_{lv} [J/m²].

Electrowetting on dielectrics

Equation 12.2 is only valid for applied voltages below those for which electrolysis sets in, about 100 mV. To circumvent this limitation, and allow a significantly larger change in contact angle with bias, a new configuration using a dielectric layer between the droplet and the lower electrode was proposed in the 1990s (Berge, 1993). Resulting in so-called electrowetting-on-dielectrics (EWOD), and shown schematically in Figure 12.4, this configuration leads to a voltage-dependent contact angle which varies as

$$\cos \theta_V = \cos \theta_0 + \frac{\epsilon_d}{2t\sigma_{lv}} V^2 = \cos \theta_0 + \eta \quad (12.3)$$

where in this case t [m] represents the thickness of the dielectric layer with permittivity ϵ_d and V the voltage between droplet and substrate. The parameter η [] is often referred to as the electrowetting parameter and represents the strength of the electrostatic energy with respect to the surface tension.

In practice, the (electrically conducting) liquid droplet in an EWOD system is immersed in an (electrically insulating) electrolyte, the two liquids chosen to be immiscible and with closely matched densities. The droplet is electrically contacted through the electrolyte and the EWOD effect is then based on a charge buildup at the interfaces. A complementary effect is dielectrophoresis, in which an insulating droplet is immersed in a conducting electrolyte or two non-conducting liquids with differing dielectric constants are employed.

Since electrowetting results in a bias-dependent contact angle, a voltage may be applied to one side of a liquid droplet (using a structured electrode array on the substrate), deforming it and causing it to move. For this reason, electrowetting has been extensively applied to microfluidic systems (Srinivasan et al., 2004; Kedzierski et al., 2009). Optically, in addition to the microlenses which we will discuss presently, electrowetting-based displays have received considerable attention (Hayes and Feenstra, 2003).

12.1.2 Tunable liquid lenses

Using the optical properties of a spherical liquid droplet, EWOD has proven to be a very useful means to realize tunable liquid lenses (Berge and Peseux, 2000). By electrically changing the contact angle, the radius of curvature of the lens can be changed and the focal length is then given by

$$f(V) = \frac{D}{2(n_L - n_A) \sin \theta_V} \quad (12.4)$$

for aperture diameter D , refractive indices of the lens liquid and surrounding ambient (typically the electrolyte), n_L and n_A , respectively, and the voltage variation of contact angle, θ_V , given by Equation 12.3. The liquid systems used generally employ a water-based lens and an oil-based ambient.

Miniaturized systems

A number of miniaturized EWOD-based tunable lens systems have been demonstrated; however, these are not necessarily microsystems. The lens structures usually include some means to position the lens on the substrate, either using a mechanical recess or a inhomogeneous dielectric to generate a non-uniform electric field (Chang et al., 2007). A number of systems are available commercially; typical specifications include an aperture diameter of 7.75 mm, an operating voltage range of 30 – 40 V and a tuning range of -5 to +13 dpt⁴; response times are in the range of 100 ms. Applications are in sensing or as tunable lenses for cell phone cameras (Wippermann et al., 2007). An alternative system uses a 3 mm diameter liquid lens mounted in a tube, with a focal length range of -100 to +50 dpt, and a maximum voltage bias of 170 V (Kuiper and Hendriks, 2004). With a response time of 10 ms, this latter system was conceived for a miniaturized zoom lens system (Kuiper et al., 2007).

12.1.3 MEMS-based liquid lenses

Using MEMS fabrication and assembly technologies, micro-optical liquid lens configurations have been developed using designs and structures suitable for fabrication in one- and two-dimensional arrays. The lateral positioning of the lenses may be accomplished using a number of micropatterning techniques, including the definition of circularly-symmetric electrode arrays (Krupenkin et al., 2003; Liu et al., 2008) and the definition of physical alignment features such as V-grooves (Krogmann et al., 2006).

The latter approach is shown schematically in Figure 12.5. A silicon substrate has a recess etched completely through the wafer using a KOH liquid etchant and the silicon is oxidized to generate a high quality SiO₂ dielectric layer, 100 to 300 nm thick; this layer is subsequently coated with a thin hydrophobic coating, such as Teflon⁵, to increase the

⁴ Diopters. Recall that the diopter is defined as the reciprocal of the focal length in meters; see Section 7.1.3.

⁵ Teflon is a DuPont brand name for polytetrafluoroethylene (PTFE), a fluoropolymer with a wide range of applications due to the the large contact angle that water and oil-based compounds form on its surface; it is known to culinarily-interested micro-opticians as the surface on non-stick cookware.

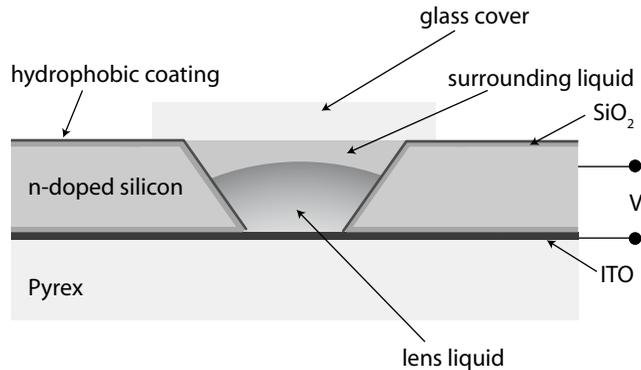


Figure 12.5 A MEMS-based tunable liquid lens based on a KOH-etched V-groove bonded to a Pyrex electrode substrate. An SiO_2 dielectric layer separates the lens from the electrically-contacted Si and the lens liquid is electrically contacted by the ITO electrodes on the lower substrate. Diagram courtesy of Florian Krogmann.

zero-bias contact angle and reduce sticking of the liquids to the sidewalls. The silicon chip is then bonded onto a Pyrex glass substrate on which ITO electrodes have been defined. Once filled with lens and surrounding liquids, the cavity is sealed with a cover glass, forming a compact stand-alone tunable microlens system, as seen in Figure 12.6.

The lens system is filled with two liquids, one which acts as a lens and the other which surrounds it. A wide variety of liquids may be employed, if they fulfill the requirements for refractive index, conductivity, immiscibility and density matching. Many of these are proprietary products, often sold for use in lasers or as index-matching fluids for microscopy; a refractive index range of $1.3 \leq n \leq 1.7$, dispersion values of $20 \leq \nu_d \leq 100$ and widely variable viscosities are available commercially. The lens liquid employed in the structure shown here is water-based, with dissolved salts providing a refractive index around 1.5 and a density of 2.1 g/cm^3 ; the surrounding liquid is a perfluorocarbon with $n \approx 1.293$. The total liquid volume is about 100 nl.

For this configuration, the electrical contact to the lens is via the ITO electrodes on the glass substrate; the EWOD capacitor is then formed by the SiO_2 layer on the conducting silicon. With aperture diameters varying between 300 and $800 \mu\text{m}$, the lenses were operated with ac voltages at 1 kHz. An initial contact angle (defined by the Teflon layer) of 96° led to a minimum focal length of 2.3 mm; application of 45 V to the system led to $f \rightarrow \infty$. The lens performance is diffraction limited and maximum power consumption (for focus at infinity) is 0.3 mW.

12.1.4 Other liquid systems

A variety of further concepts using electrowetting for micro-optics has been demonstrated, including many variations on the setups we have just discussed. The focussing time of liquid lenses has been decreased, for example, by using an induced oscillation in the lens (López and Hirsá, 2008); by sweeping the focus at 100 Hz, the focal point is traversed every few milliseconds, such that a timed imaging system is required.

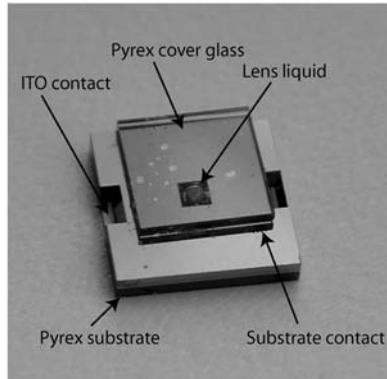


Figure 12.6 Photograph of the V-groove-based liquid lens system shown schematically in Figure 12.5; the lens itself is seen in the cavity toward the center. The entire assembly has a size of $8 \times 8 \times 1.5$ mm. Photo courtesy of Florian Krogmann.

Dielectrophoresis

Systems based on dielectrophoresis, in contrast to electrowetting, use two liquids with differing dielectric constants. Whereas electrowetting is an interface phenomenon, such that the applied electric field changes the contact angle, dielectrophoresis is a bulk effect, implying that the application of a non-uniform electric field results in a movement of the liquid droplet due to induced polarization in the dielectric fluids (Gascoyne et al., 2004). Typical dielectrophoretic systems use concentric rings of electrodes to generate the required field gradients, and while requiring relatively high voltages, power consumption is low since there is no current flow. Despite the fact that heating effects in these systems can be detrimental, dielectrophoretic microlens arrays have been successfully demonstrated using two immiscible liquids and a suitably patterned electrode array (Ren and Wu, 2008).

Movable lenses

In addition to changing its focus, the change in contact angle which results from electrowetting can also be used to move a microlens; the effect is the same as that used for moving droplets in electrowetting-based microfluidic systems. Using patterned electrode arrays and specialized surface structures, liquid microlenses (Krogmann et al., 2008c) have been positioned on a two-dimensional transparent substrate with an accuracy of $70 \mu\text{m}$, and the lenses were subsequently tunable in focal length in the range of 0.6 to 1.2 mm; an example of such a system is shown in Figure 12.7.

Liquid microlenses may also be moved by purely optical means (Krogmann et al., 2008b), employing opto-electrowetting. In this approach, an optical field projected onto a photoconductive substrate is used to generate a non-uniform electric field which, depending on the bias conditions, may be used to push or pull a lens into position. Alternatively, suspended nanoparticles at the liquid/ambient interface may absorb incident optical energy and lead to a temperature change, and thus also serve to allow con-

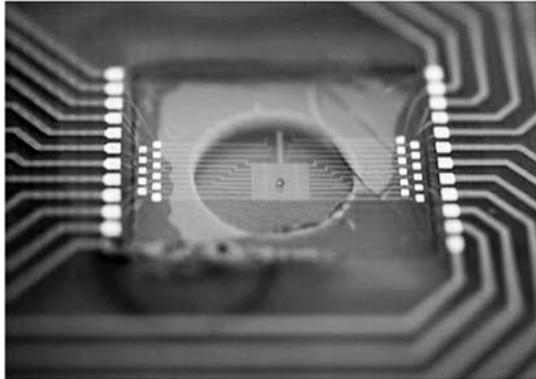


Figure 12.7 A movable liquid microlens (the small dot at the center) placed on a two-dimensional electrode array. By applying suitable voltage sequences on the electrodes, the liquid lens could be positioned and subsequently tuned in focal length. The electrode array has dimensions 4×4 mm. Photo courtesy of Ryaz Pasha Shaik.

trolled movement using light as an actuator (Liu et al., 2006). These techniques, while flexible, are at best experimental at present.

12.2 Membrane microlenses

A conceptually different variation on liquid micro-optics is that of membrane-based microfluidic lenses. As for the liquid lenses to which we were just introduced, the liquid itself performs the optical function, but the optical surface for the microlenses we consider now is defined by a flexible, highly distensible membrane, whose radius of curvature is typically changed by varying the pressure on the liquid (Werber and Zappe, 2008).

12.2.1 Membrane lens concept

The basic structure of pressure-actuated membrane microlenses is shown in Figure 12.8. In the example of the figure, a microfluidic channel is defined in silicon and the resultant cavity sealed on one side by Pyrex and on the other using a highly-distensible PDMS membrane. The channel is filled with an optical liquid, preferably water-based substances to which the PDMS is impermeable, with a suitable refractive index; similar liquids to those used for the electrowetting lenses can often be used.

Since the liquid is in a sealed cavity, applying hydrostatic pressure results in a distension of the membrane, which is designed to be highly flexible (Hoshino and Shimoyama, 2003). PDMS-based membranes can withstand an elongation of up to 400%, such that a complete hemisphere can be generated on the substrate surface. The distended PDMS film attains a (nearly) spherical profile at the center of the aperture, whose radius of curvature may be tuned by pressure; the result is thus a pressure-tunable microlens.

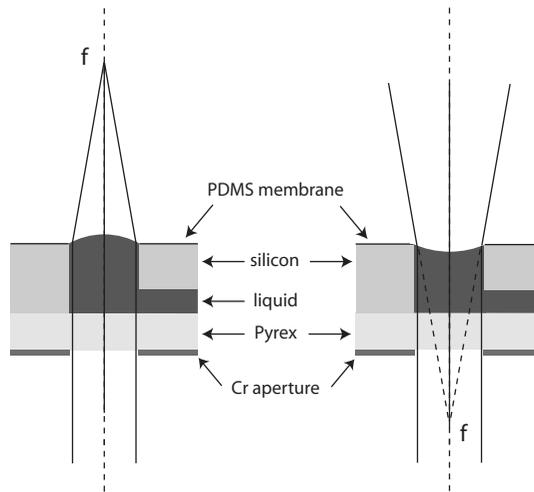


Figure 12.8 Schematic cross-sections through a membrane liquid microlens. A microfluidic channel etched into Si and sealed using a Pyrex substrate on one side and a PDMS membrane on the other is filled with an optical liquid. Pressure applied to the liquid causes the membrane to distend, forming a convex *left* or concave *right* lens, for positive and negative pressures, respectively. Diagram courtesy of Armin Werber.

The active region of the lens (typically the central 80% of the diameter) is defined by an opaque aperture stop (Cr in the figure), since at the outer edges the profile of the membrane departs from that of a sphere.

Since the membrane is relatively thin, typically 50 μm or less, its optical effect is negligible and the refractive power of the lens is due only to the liquid, whose surface profile is given by the distended PDMS film. As seen in Figure 12.8, both convex and concave profiles may be generated, using positive (with respect to the ambient) or negative pressures, respectively.

12.2.2 Miniaturized membrane lenses

Unlike the purely liquid lens of Section 12.1, the tunable membrane lens also works (to an extent) in macroscopic implementations; membrane lenses with aperture diameters up to 2.7 cm have been demonstrated (Sugiura and Morita, 1993). Smaller versions, with $5 \leq D \leq 10$ mm, configured as plano-convex, biconvex, biconcave, or cylindrical structures have also been fabricated, and used in miniaturized zoom and other imaging systems (Zhang et al., 2004c). Fabrication also typically uses PDMS membranes, occasionally formed using SU-8 as a mold.

Actuation of these macroscopic lenses is either done externally, using pressure controllers, or on-chip by means of piezo-actuators (Draheim et al., 2009) or expanding hydrogels (Dong et al., 2006). One envisaged applications area for these lenses is for ophthalmological correction, where defocus and astigmatism correction of $\pm 20\text{dpt}$ (± 50 mm in focal length) is required, with an aperture of 2.54 cm (Marks et al., 2009).

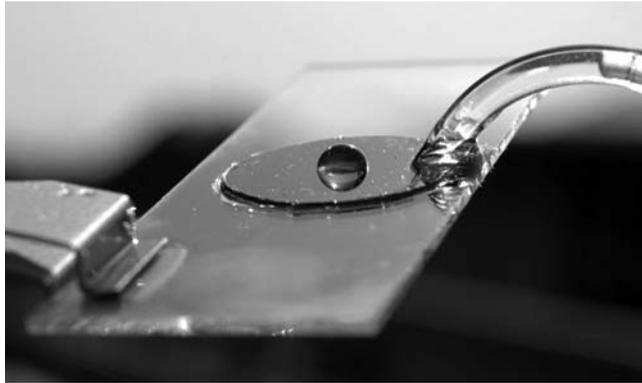


Figure 12.9 Profile view of a distended membrane micro lens, with a diameter of $4\,000\ \mu\text{m}$, on a silicon disk bonded to a glass substrate. In this test structure, the pneumatic actuation is applied through the tube on the right; tweezers are seen on the left. Photo courtesy of Armin Werber.

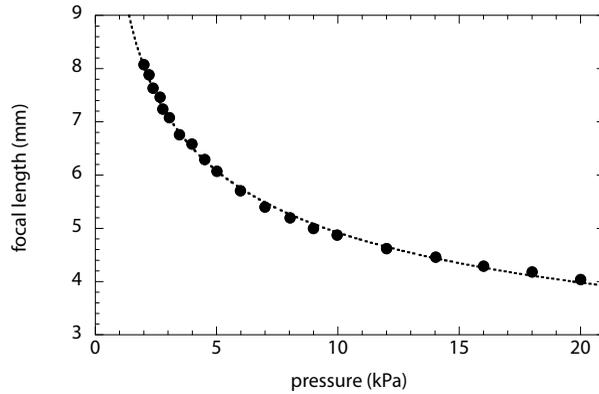
Whereas these lenses with centimeter diameters are feasible, gravitational effects result in a deformation of the membrane and thus the lens. As a result, most of these devices can not be operated in all orientations or require very thick membranes and thus high operating pressures to offset gravitational deformation.

12.2.3 Micro membrane lenses

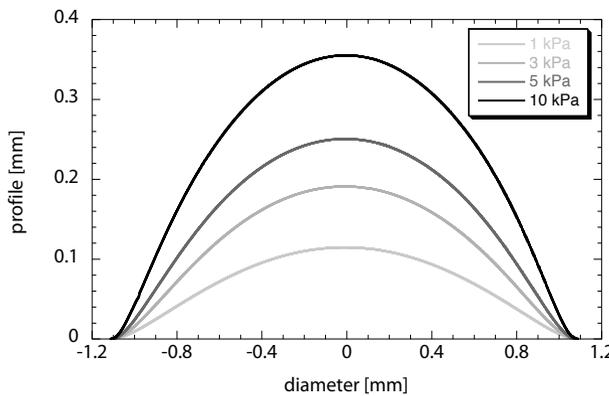
Pneumatic membrane microlenses typically have aperture diameters between several millimeters down to $400\ \mu\text{m}$, and use a PDMS membrane with a thickness on the order of $50\ \mu\text{m}$ (Werber and Zappe, 2005). These lenses are generally fabricated on a microfluidic substrate assembled from silicon and glass (Agarwal et al., 2004) or molded PDMS channels bonded to glass (Chronis et al., 2003); an example is shown in Figure 12.9.

Microlenses of this type have been operated using a liquid with $n = 1.6057$, such that a pressure-induced variation of the radius of curvature in the range $200\ \mu\text{m} \leq R \leq \infty$ leads to a usable focal length range of $1 \leq f \leq 18\ \text{mm}$. The smallest radius is reached for a pressure of $54\ \text{kPa}$, at which point the lens has a sag height of $40\ \mu\text{m}$. Typical characteristics of tunable membrane microlenses are shown in Figure 12.10. The variation of focal length with pressure, shown in Figure 12.10(a), shows the expected decrease of f with increasing pressure (varying roughly as $(\text{pressure})^{-1/3}$), in the range 4 to $8\ \text{mm}$. The change in membrane profile as a function of pressure is given in Figure 12.10(b) for a $2\ \text{mm}$ aperture lens; we see a deviation from an ideal profile at the very edges of the lens, due to the attachment of the membrane onto the substrate. Finally, the MTF (recall Section 7.2.3) for two pressures (and thus two radii of curvature), measured for a two-microlens combination (one of fixed focal length, the other the tunable membrane lens) is shown in Figure 12.10(c); we see that the resolution of the lens varies with focal length, but at values close to the diffraction limit for lenses of this size.

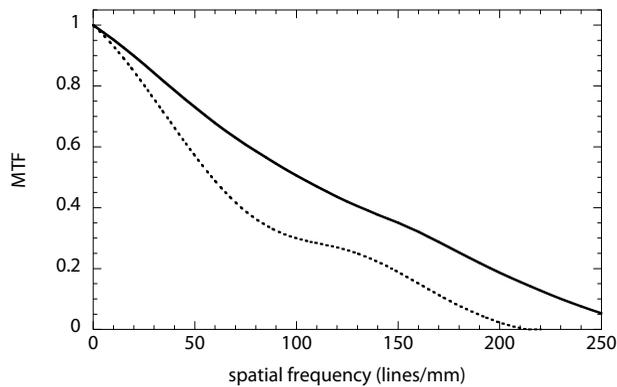
Numerous variations on this concept have been demonstrated. Plano-convex and biconvex lenses (Zhang et al., 2003, 2004a), as well as meniscus lenses (Zhang et al.,



(a) Focal length as a function of applied pressure; the circles are measurement values and the dotted line a fitted curve of the form $(\text{pressure})^{-1/3}$.



(b) Measured profiles at different pressures, varying from 1 kPa *gray* to 10 kPa *black*.



(c) MTF for 2 kPa *dotted* and 5 kPa *solid* applied pressure

Figure 12.10 Characteristics of pneumatic membrane microlenses; shown are the focal length as a function of pressure for a 1 mm aperture lens; membrane profile as a function of pressure for a 2 mm aperture lens; and the MTF for two pressures of a dual microlens (tunable and fixed) system. Data courtesy of Khaled Aljasem and Wei Zhang.

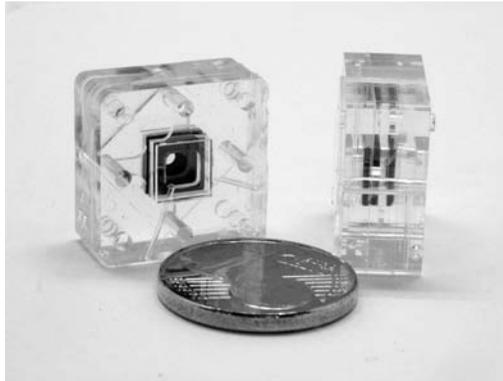


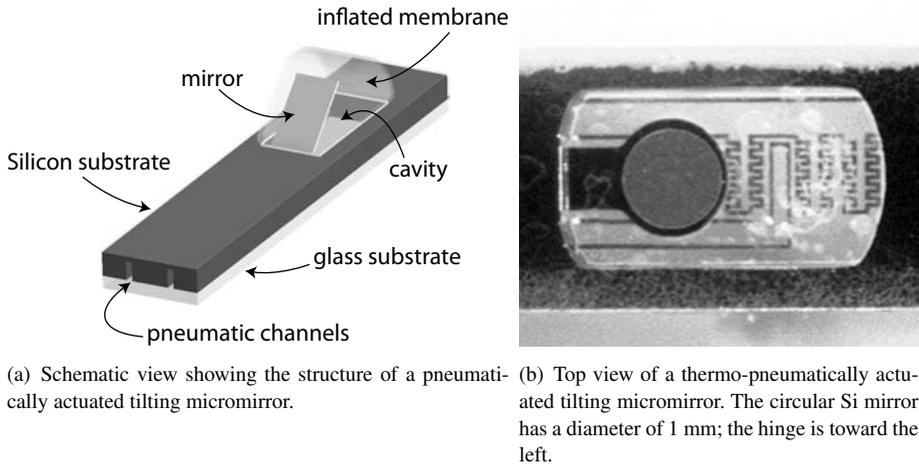
Figure 12.11 A multi-chamber tunable achromatic lens, using three pneumatic chambers separated by flexible membranes. The lens chambers themselves are made from silicon, seen as the darker chips at the center; the pneumatic access channels, connectors and alignment features are fabricated in the clear polymer dies, bonded together to form the complete system. The outer die has a side-length of 14.8 mm. Photo courtesy of Daniel Mader.

2004b) have been made using the membrane approach. With diameters of 20 mm, these relatively large components could nevertheless be tuned from concave to convex profiles, with minimal focal lengths of -6.3 mm and $+14.3$ mm, respectively. In addition, rudimentary zoom systems employing two back-to-back membrane lenses with a zoom ratio of 2 have been fabricated (Zhang et al., 2004d).

Multi-chamber membrane lenses

The membrane lens concept can be significantly extended to include multi-chamber tunable lenses. Using microsystem assembly technologies, a succession of liquid chambers, separated by flexible membranes, may be combined into a single multi-element lens stack, allowing the fabrication of tunable doublets or triplets (Reichelt and Zappe, 2007); as a result, tunable achromatic lenses (using the same principles as outlined in Section 7.5.2) may be made using this approach (Mader et al., 2009).

A tunable achromat, such as that shown in Figure 12.11, is typically configured as a stack of fluidic cavities, separated by flexible membranes. Each cavity is filled with a different optical fluid, for which the refractive index and Abbe number has been chosen to allow achromaticity over a certain wavelength range; a good selection of liquids is commercially available. The pressure applied to each chamber may be regulated separately, such that the curvatures of the internal and external membranes may be optimized to achieve a desired focal length, while maintaining achromatic behavior and, frequently, compensating for spherical aberration. A related structure using one tunable and one fixed lens as a doublet has also been developed (Jeong et al., 2004). These designs are still laboratory prototypes, but demonstrate the utility of this tunable microlens technology.



(a) Schematic view showing the structure of a pneumatically actuated tilting micromirror.

(b) Top view of a thermo-pneumatically actuated tilting micromirror. The circular Si mirror has a diameter of 1 mm; the hinge is toward the left.

Figure 12.12 Pneumatically-actuated tilting micromirrors; the silicon-based mirror is mounted on an elastomer membrane and can tilt up to 80° . In the photo on the right, the meandering heater structure in the cavity can be seen through the clear elastomer membrane. Diagram and photo courtesy of Armin Werber.

12.2.4 Pneumatic micro-optics

The pressure-induced distension of flexible elastomer membranes can be used as a means of actuation for a wider range of tunable micro-optical components (Werber and Zappe, 2006b). Using silicon micromachining, a micromirror attached to a hinge can be fabricated on a membrane, as shown schematically in Figure 12.12(a), resulting in a pneumatically-actuated tilting micromirror. A gold layer is deposited on the mirror to enhance reflectance and chip dimensions are $7.2 \times 2.2 \times 1 \text{ mm}^3$, with a reflecting mirror area of $1 \times 1.3 \text{ mm}^2$.

For this mode of operation, the microfluidic cavities are filled with air and applying pressure then again causes the membrane to distend, tilting the micromirror. With externally-applied pressures up to 65 kPa, tilt angles of up to 75° could be attained using this structure, considerably higher than the angles achievable with standard micromachined and electrostatically-actuated micromirrors; recall the discussion of Section 6.4.2. Actuation of these micromirrors may also be accomplished on-chip, using thermo-pneumatic principles: integrated heaters inside a sealed, gas-filled cavity (as seen in the photo of Figure 12.12(b)), are used to generate a pressure increase. Tilt angles up to 12.5° were reached using thermo-pneumatic actuators, but the approach generated chip temperatures of up to 300°C and required high electrical input powers.

An alternative micromirror configuration using this actuation approach is that of a piston mirror, undergoing purely vertical movement. A hexagonal micromachined silicon mirror, with a side-length of $825 \mu\text{m}$, was fabricated on a distensible PDMS membrane, which, when inflated, resulted in vertical motion of the mirror (Werber and Zappe, 2006a). Shown in the photos of Figure 12.13, pressure actuation for these mirrors was

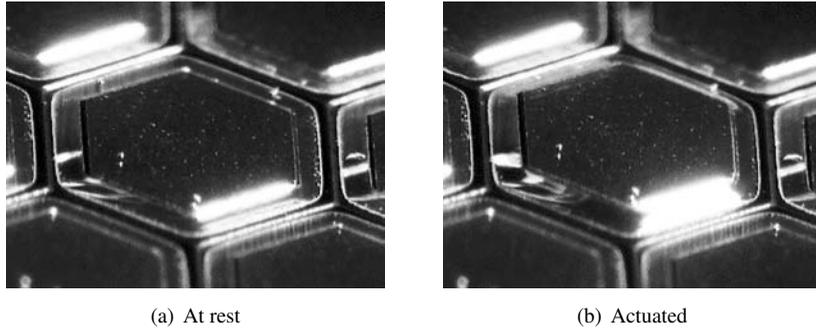


Figure 12.13 Thermo-pneumatically-actuated piston micromirrors exhibiting purely vertical motion; the mirror at the right is actuated. The hexagonal silicon mirror surface has a side-length of 825 μm . Photos courtesy of Armin Werber.

by means of on-chip heaters, regulated by integrated thermistors for temperature monitoring.

From the ideal gas equation and geometrical considerations, the vertical movement, Δs , of this mirror as a function of internal temperature change may be found from

$$\Delta s = \frac{2V_0\gamma}{3\sqrt{3}a_c^2}\Delta T \quad (12.5)$$

in which a_c [m] represents the side length of the hexagonal layout of the actuator, V_0 [m^3] the initial (unactuated) volume of the cavity, γ [K^{-1}] the thermal coefficient of expansion of air (with a value of about $\gamma = 0.0036$ [K^{-1}]) and ΔT [K] the temperature change of the air in the cavity. Assuming on-chip actuation using resistive heaters, with total electrical resistance of R_H [Ω] and thermal resistance of the entire system R_{th} [K/W], the stroke can be expressed as a function of applied voltage to the heaters, V [V], as

$$\Delta s = \frac{2V_0\gamma}{3\sqrt{3}a_c^2} \frac{R_{th}}{R_H} V^2. \quad (12.6)$$

We thus expect a quadratic variation of mirror stroke motion with applied electrical bias, and this behavior has been verified experimentally. Vertical movement of up to 385 μm using an actuation voltage of 32 V on Pt heaters has been demonstrated, with a parasitic angular movement below 0.1° . Although this type of performance illustrates the capabilities of tunable pneumatic micro-optics, problems such as the slow response time (with time constants in the range of seconds) and permeability of the membrane still need to be addressed.

12.3 LCD-based tunable micro-optics

Tunable optics has benefitted from the use of liquid crystal technology, which we covered in Section 10.6.2, particularly for the realization of reconfigurable microlenses (Sato, 1979; Masuda et al., 1997) and switchable gratings (Yeralan et al., 2002).

12.3.1 Liquid crystal microlenses

Liquid crystal microlenses have been demonstrated using a nematic cell configuration, in which a circular transparent electrode (using ITO) with a hole at the center is used to generate an inhomogeneous electric field in the cell; the counter electrode is planar, on the second cell surface. Using liquid crystals with positive dielectric anisotropy, the homogeneous liquid crystal distribution becomes homeotropic under the electrode generating a quadratic refractive index profile in space, much like that of a GRIN lens. Without an applied bias, the director of the liquid crystals is parallel to the substrate, such that the extraordinary refractive index, n_e , is seen; when a bias is applied, the director turns perpendicular to the substrate near the edge of the aperture, providing a lower n , leading to a uniform decrease of refractive index toward the lens edge, and thus a concave lens function (Nose and Sato, 1989).

Typical liquid crystal lens sizes have apertures ranging from 0.5 to several millimeters and are generally less than 200 μm thick; the voltages required are in the range of 3 to 4 V. The maximum numerical aperture of this type of lens has been found to be achievable for an aperture/thickness ratio of about 3 (Masuda et al., 1996). Convex lenses may be realized using either liquid crystal materials with negative dielectric anisotropy or by suitably choosing the applied bias regime, such that switchable concave/convex structures have been designed (Nose et al., 1991); using dual-frequency liquid crystals, fast response times, 4 – 6 ms, are achievable (Fan et al., 2005). An advantage of using liquid crystals for realizing microlenses is that non-uniform, non-circularly-symmetric structures may be defined and tuned, as are non-spherical patterns useful for correction of aberrations (Honma et al., 1999). A remaining disadvantage is that polarized light must be employed, limiting the scope of applications.

In addition to the patterned electrode described above, numerous other fabrication approaches are useful for liquid crystal microlenses, including the use of polymer-dispersed liquid crystal droplets, surface-relief structures defined by microcontact printing (Cheng et al., 2006) or the use of photo-induced alignment of liquid crystals in the cell (Kinoshita et al., 2007).

12.3.2 Other tunable liquid crystal devices

A number of other tunable micro-optical components using liquid crystals as the active medium have been proposed and demonstrated. A tunable Fresnel zone plate, with 80 zones in a 1 cm aperture, has been fabricated using lithographic patterning of the liquid crystal regions in the annular regions (Ren et al., 2003). The diffraction efficiency of

the lens, with a focal length of 50 cm, could be tuned by applying a bias to the liquid crystals, albeit not over a wide range.

Using a means to generate a uniformly varying electric field, liquid crystal cells have also been conceived as spatially-varying phase shifters, allowing correction of lower-order aberrations (Naumov et al., 1998), a concept which has been used to compensate for spherical aberration in a Blu-ray disc reader system (Knittel et al., 2005).

12.4 Tunable diffractive micro-optics

As with the refractive lenses and micromirrors we saw above, tuning of diffractive optical structures is also of considerable interest. Again, several novel approaches for achieving tunability have been developed for tuning, among other structures, gratings, Bragg mirrors and photonic crystals.

12.4.1 Gratings

Gratings, which we saw in Section 8.3, and Bragg mirrors or filters, to which we were introduced in Section 5.3.5, are periodic diffractive structures whose primary characteristic parameter is their period. These structures can be “tuned” by changing the period, which may be accomplished by one of several means.

Micromechanical tuning

Micromechanical tuning of grating period or shape has a long history, and forms the basis for a number of useful applications, especially displays and spectrometers (Solgaard, 2009, Chapters 10 and 11). These tunable devices rely on, for example, vertical movement of individual grating corrugations (typically by $\lambda/2$) to allow discrete switching between various periods (Solgaard et al., 2003) or analog positioning of individual periodic elements to allow a fully programmable optical transfer function (Senturia, 2002). More advanced structures can approximate and vary the blaze of a tunable grating using multi-step movable grating elements (Wang et al., 2009).

In addition, fiber Bragg gratings have long been tuned using mechanical (Xiong et al., 1999) or thermal (Eggleton et al., 2000) means. The variability of fiber Bragg grating period with external parameters, such as strain or temperature, forms the basis for a large family of fiber-based sensors (Rao, 1997), as we saw at the end of Section 9.7.2, thus representing passive “tuning” used for different applications.

Stretchable gratings

We saw for membrane lenses above that silicone elastomers are a category of materials with high mechanical flexibility; they may also be used for the fabrication of stretchable gratings. Using standard or gray-scale lithography to generate rectangular or triangular profiles, respectively, grating structures with periods ranging from 4.5 to 32 μm have been fabricated, with profile depths around 0.5 μm (Simonov et al., 2005). These grat-

ings could be mechanically stretched using micromanipulators or calipers, resulting in a period change of $\Delta\Lambda = 18\%$, with no reduction in diffraction efficiency.

An attractive alternative for tuning stretchable gratings with more promise for integration in a microsystem is the use of electroactive polymers (EAP), sometimes known as artificial muscles. Due to the very large strains that these materials can achieve in the presence of an electric field (up to 380 %), a correspondingly large change in grating period is expected. The effect has been demonstrated using an acrylic elastomer grating, whose initial period of 1 μm could be increased by 32 %, albeit with an applied voltage of 4.5 kV (Aschwanden and Stemmer, 2006). To illustrate the utility of the effect, the diffracted light from a white light source was scanned from the blue to the red using this stretchable grating, suggesting its suitability as a basis for a video display. EAP materials have also been used to compress a circularly-symmetric grating operated in transmission (Aschwanden et al., 2007).

12.4.2 Bragg filters

Tunable micro-optical filters have been demonstrated using a number of approaches, including micromechanically movable Fabry-Perot filters (Milne et al., 2009). Non-mechanical filters may be realized using diffractive Bragg mirrors which are wavelength-tunable through a controlled change in temperature or by a change in the mirror thickness.

As we saw in the discussion of interference filters in Section 5.3.5, the layer structure used for the VCSEL (a resonant cavity surrounded by two Bragg mirrors, as shown in Figures 5.26 and 10.17(a)) may be used as a wavelength-selective filter. The Bragg mirrors have a relatively broad reflection spectrum (the stop-band) and the width of the central resonant cavity (typically an integral number of half-wavelengths thick) defines a sharp transmission peak, usually in the center of the stop-band, as we saw in Figure 10.18. This type of a structure requires a layer sequence using materials transparent at the operating wavelength and providing an adequate refractive index step, Δn , to generate sufficient Bragg reflection.

Temperature tuning

We saw briefly in Section 3.2.3 that the refractive index of a material changes with temperature. As a result, temperature tunable Bragg mirrors and filters may be made using layers which have an appropriately large thermal response. For a cavity of integer half-wavelength thickness, $t = m\lambda/2$, of material with refractive index n_c , the transmission wavelength λ_T is given by

$$\lambda_T = \frac{2n_c t}{m} \quad (12.7)$$

for index $m = 1, 2, 3, \dots$. The tunability of transmission wavelength is then a function solely of the optical thickness of the cavity; the Bragg mirrors themselves change little with temperature, and their effect on the change in reflection wavelength is mini-

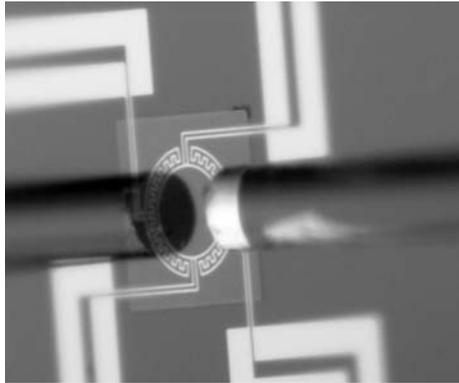


Figure 12.14 An optical fiber *right* coupling into a silicon-based thermo-optic filter, the square membrane at the center, which is heated using the circular electrode structure. An image of the input fiber specularly reflected from the surface is seen at the left. Photo courtesy of Dennis Hohlfeld.

mal. Thus a thermally-tunable filter requires a cavity material with a large thermo-optic effect.

Using silicon-based materials (amorphous silicon and its oxides and nitrides), a thermally tunable diffractive Bragg filter, as is shown in the photograph of Figure 12.14, may thus be conceived (Hohlfeld et al., 2003; Hohlfeld and Zappe, 2004, 2007). Using the resonant Bragg filter structure, amorphous silicon is employed for the central $\lambda/2$ cavity, and SiO_2 and Si_3N_4 for the Bragg mirrors on either side.

To tune the filter with temperature, the optical thickness of the resonant cavity is changed with temperature. Amorphous silicon has a change in refractive index as a function of temperature (a thermo-optic effect) of magnitude $3.25 \times 10^{-4} \text{ K}^{-1}$ but a thermal coefficient of expansion of $2.6 \times 10^{-6} \text{ K}^{-1}$, so that the change in index dominates over any physical length change with temperature. Optical path length change is thus a refractive index change. The thermo-optic coefficients of SiO_2 and Si_3N_4 are in the range 10^{-7} K^{-1} , so that the Bragg mirrors are not expected to change significantly with temperature, as we mentioned above.

Using these materials and considering their nominal refractive indices, $\lambda/4$ layers for the Bragg mirrors used thicknesses of $t_{\text{oxide}} = 267 \text{ nm}$ for SiO_2 (the L layer) and $t_{\text{nitride}} = 202 \text{ nm}$ for Si_3N_4 (the H layer); the amorphous silicon cavity required a thickness of 208 nm. 12 HL layer pairs are deposited for both the top and bottom mirrors. Using microsystems fabrication techniques, this multi-layer system, with a total thickness of about $4 \mu\text{m}$, was fabricated as a suspended membrane and thus with a low thermal mass.

Using this structure, a maximum temperature change of 450°C was achieved using electrical power input from integrated resistive heaters. This temperature variation led to a filter wavelength tuning range of up to 40 nm, with a spectral width of 0.3 nm. Due to the small thermal mass of the filter structure, tuning over this range could be accom-

plished with millisecond time constants, a decided advantage of this microsystems-based structure.

Polymer swelling

An alternative approach for realizing tunable polymer-based Bragg filters is through a swelling of the structure. Many polymers swell in the presence of water or solvents in the ambient atmosphere; this effect has been used to fabricate tunable Bragg gratings as well as one and two-dimensional grating structures (Mönch et al., 2006a,b).

Planar Bragg filters, with the same type of resonant cavity structure as used above, were fabricated using alternating layers of PS and PMMA deposited on Pyrex. A half-wavelength cavity was surrounded by 16 layer pairs forming the Bragg mirrors, where the high index PS layers were 94.3 nm thick, and the low-index PMMA layers 100.6 nm thick, both corresponding to $\lambda/4$ for operation at $\lambda = 600$ nm; the entire filter could be fabricated using spin coating and subsequent UV cross-linking. Thus structure has the same type of transmission characteristic as for the vertical cavity laser (qualitatively as seen in Figure 10.18, but less sharp due to the lower quality of the mirrors), with a peak reflectance of about 0.6.

If organic solvents, such as acetone or toluene, are present in the atmosphere above the filter, these are absorbed by the polymer layers which swell in response; the layer thicknesses thus increase, and the transmission wavelength is red-shifted. A wavelength shift of 600 to 1200 nm was demonstrated using this approach, corresponding to a very large tuning range. The concomitant disadvantage of this tuning mechanism is the relatively long time constants involved (on the order of minutes), which currently limits the regimes of application. Nevertheless, the mechanism illustrates one further means by which wavelength-selective mirrors may be fabricated using novel MEMS processes, and suggests possible means for making these tunable through mechanical distortion.

12.4.3 Photonic crystals

Photonic crystals are the subject of Chapter 14.2, where we will consider their structure and behavior in detail. Summarizing their essential properties, photonic crystals are (occasionally complex) periodic structures which exhibit an optical bandgap; this bandgap may also be tuned, opening new opportunities for the application of these devices. One important structural family of photonic crystals is the two-dimensional slab, which is the configuration we consider here.

A number of approaches for tuning these structures has been proposed; most of these are based on modeling, as experimental results are rare. Simulation has shown that semiconductor-based photonic crystals may be tuned by the application of strain to a suspended Si stripe, in which the periodic structure has been defined (Rajic et al., 2003). The application of 3% strain can result in the shift of the optical bandgap by 73%, a significant figure. Simulations have also been done for diamond, GaAs and Si_3N_4 as basis materials.

Alternatively, polymer-based photonic crystals may also be realized using stretchable structures, not unlike the gratings we just considered. Stretchable PDMS sheets,

patterned into three-dimensional photonic crystals using colloidal assembly of microspheres, have been deformed by application of mechanical strain, which has led to a marked color change in the diffracted light; a wavelength peak shift of 589 to 563 nm was measured (Fudouzi and Sawada, 2006). Alternatively, using flexible PDMS as a base material, stretching a two-dimensional photonic crystal by 10% led to a shift in the refraction angle of an incident optical field by up to 75° ; the concept was developed for the application of mechanical deformation using MEMS-based actuators (Park and Lee, 2004). We will consider fluidic-tunable photonic crystals in Section 13.6.1 and other tunability concepts in Section 14.2.5.

12.5 A tunable micromanagerie

We have only scratched the surface of micro-optical tunability in this chapter; there is a host of further mechanisms and micro-optical devices whose characteristics may be tuned. We take a glance at some of these in this final section, to give the reader an idea of what is possible.

12.5.1 Electro-optical lenses

The electro-optical effects we studied in Section 10.5.2 may be used to realize tunable lenses (Shibaguchi and Funato, 1992). Using electro-optically active materials such as LiNbO_3 , LiTaO_3 or lead-lanthanum zirconate-titanate (PLZT) ceramics, tunable microlenses have been demonstrated, as have tunable Fresnel zone plates. In these materials, n varies parallel to an applied electric field, due to the electro-optical effect, resulting in a change of the optical behavior. The centimeter-sized chips, however, require bias voltages of up to 600 V, not immediately practicable.

12.5.2 Hydrogel micro-optics

Hydrogels are colloidal gels with a high affinity to water; they swell very strongly in a humid atmosphere, absorbing water efficiently, such that they are most popularly used in diapers. We saw in Section 12.4.1 above that swelling of polymers can be used for tunable gratings; similarly, swelling hydrogels can be used for tunable microlenses, either as the lens itself or as an actuation mechanism.

Using poly(N-isopropylacrylamide), a hydrogel which responds to a variety of stimuli, tunable microlenses have been demonstrated, using the hydrogel as the lens body itself (Kim et al., 2004). These “lenses”, however, had diameters of 1 to 2 μm , such that optical performance is seriously limited; this may be considered to have been more of a conceptual demonstration, since these tunable microlenses are not a useful size.

More practical is the use of hydrogels for actuating deformable or movable microlenses. A hydrogel has been used to tune the focal length of a water-based liquid microlens, with an aperture diameter of 6 mm, over a focal length range of $0 \leq f \leq 60$ mm (Dong et al., 2006). The lens was defined, as for the liquid lenses

of Section 12.1, by surface tension between water and an oil ambient; surrounding the circular lens, the hydrogel compressed the contact line, thereby decreasing the radius of curvature. This process can be relatively fast, with tuning time constants on the order of 20 μs .

12.5.3 Tunable lasers

A final example of tunable micro-optical components is the tunable laser. Whereas tunable semiconductor laser diodes, especially DFB or DBR structures as we saw in Section 10.3.3, have a long history and are well established (Coldren et al., 2009), new approaches for tuning microlasers using MEMS technologies are being developed. One popular example, for which we analyzed one implementation in 8.5.1, is the external cavity laser using micromechanically movable grating mirrors; numerous variations on this theme have been demonstrated.

Another novel approach for tuning the emission wavelength of a laser diode is by means of a stretchable PDMS grating reflector in a DFB dye laser (Li et al., 2006b). The Bragg grating was embedded in a liquid core waveguide filled with a Rhodamine dye and was optically pumped using a Nd:YAG laser; the tuning range, limited by the gain spectrum of the dye, was 60 nm, in the orange and red portions of the visible spectrum. This opto-fluidic device will be examined in more detail in Chapter 13.

13 Fluidic micro-optics

*Mais vous ne vous portez point bien, vous n'avez point dormi; le chocolat vous remettra: mais vous n'avez point de chocolatière, j'y ai pensé mille fois; comment ferez-vous?*¹

Madame de Sévigné, 1671

Liquids play an important role in both society and in many technologies. Although the viscosity, optical absorption and scattering properties of chocolate make this particular liquid unsuitable for application in optics, the use of other fluids has resulted in fluidic micro-optics developing into an innovative new discipline. The use of fluids for optics is not new; macro-optical examples include rotating pools of Mercury used as reflecting telescope mirrors (Borra, 1982) and we have already seen micro-optical examples in the previous chapters.

Fluidic micro-optics, for which the term optofluidics has attained common usage, is based on the combination of microfluidics and micro-optics, and has evolved in the last decade. Microfluidics, as we mentioned in Chapter 12, is a branch of microsystems engineering which employs microfabricated fluidic channels, micropumps and microvalves to realize complete fluidic systems (Srinivasan et al., 2004; Kedzierski et al., 2009). Many microfluidic systems have been developed for biochemical analysis, in which optical components already play a role; laser diodes, LEDs, photodetectors and other micro-optical devices have been integrated with fluidic structures to yield micro-total-analysis (so called μ TAS) systems.

Optofluidics, however, concerns the use of fluidic effects and systems for the realization of the optical component itself (Psaltis et al., 2006; Monat et al., 2007). We consider a few examples of these devices in this chapter, including lasers, lateral microlenses, waveguides and filters, all based on fluidic effects.

¹ “You do not feel well, you have not slept; chocolate will help you recover: but, you do not have a chocolatière, and I thought about it a thousand times: what will you do?” Marie de Rabutin-Chantal, marquise de Sévigné (1626–1696) was a French aristocrat whose witty letters today provide a glimpse into the pleasures and ailments of nobility in 17th century France. These chocolate musings are from Letter CXVI written by Madame de Sévigné to Madame de Grignan (her daughter) in February, 1671 (Sévigné, Marquise de, 1823), p. 320

13.1 Optofluidics

The combination of micro-optics and microfluidics has seen rapid development in the past ten years. A variety of components and systems has been demonstrated, some of which we have seen previously. The liquid lenses of Sections 12.1 and 12.2 are an important branch of fluidic micro-optics, which could just have well been considered in this chapter; these are a special case of microfluidic tunable optical devices (Levy and Shamai, 2008).

Alternatively, the integrated optical Mach-Zehnder chemical sensor we discussed in Section 9.7.2 was an example of an evanescent wave sensor operated in liquid; bio-sensing using opto-fluidic concepts is another applications area subject to significant development (Erickson et al., 2008) and particularly surface-plasmon resonance sensors are beginning to emerge from the laboratory and are becoming industrially viable (Homola et al., 1999; Karlsson, 2004). The use of optofluidics for biochemical sensing and analysis is a particularly active research area, in which optical trapping of cells and particles in liquids, micro-flow cells, optically-driven pumps and capillary electrophoresis are used in miniaturized systems (Hunt and Wilkinson, 2008).

Apart from these applications, optofluidics has also been developed as a technology useful for conceiving new types of micro-optical components. We have mentioned liquid-based microlenses, but will see that reconfigurable liquid waveguides, optofluidic lasers and fluidic filters, displays and resonance-based sensors have also been demonstrated. The following sections provide a brief overview of this rapidly developing field.

13.2 Fluidic waveguides

Some of the earliest work in optofluidics revolved about the the use of fluid streams as liquid waveguides. Rather than filling a microfluidic channel with a fluid, which can also act as a waveguide if the refractive index of the liquid is higher than that of the material defining the channel, liquid/liquid (so-called L^2) waveguides can also be defined (Wolfe et al., 2004). As shown in Figure 13.1, L^2 waveguides are based on two liquids which flow with low Reynolds² numbers (5–500), implying laminar flow with no turbulence³; as a result, the two liquids have no or little intermixing and the interfaces between them are very smooth.

The microfluidic channels, as seen schematically in the figure, are defined using PDMS (with $n = 1.40$), with a width of 300 μm and a height of 100 μm . The chan-

² Osborne Reynolds (1842–1912) was one of the first professors of engineering the United Kingdom, at the University of Manchester. He made extensive fundamental contributions to the science of fluid dynamics, but also espoused a theory that all of space is not filled with ether but rather infinitesimally small granular particles; he is remembered more for the former.

³ The Reynolds number is defined as the ratio of inertial to viscous forces in fluid flow; low numbers imply laminar, high numbers turbulent flow. The exact value of the transition depends on the geometry of the system, and varies between about 10^3 to 10^6 .

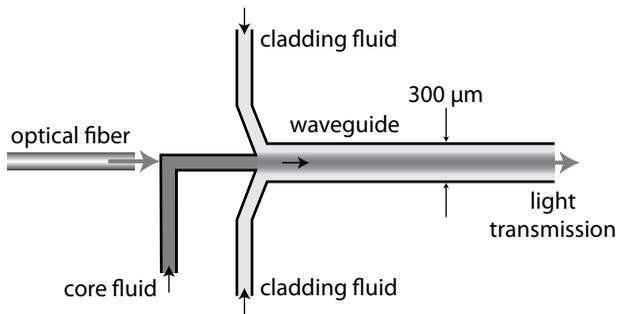


Figure 13.1 Schematic sketch of a fluidic waveguide. Core and cladding fluids are combined in the waveguide section and move to the right in laminar flow; light is coupled into the waveguide using the optical fiber at the left and is guided in the core fluid. Diagram based on (Wolfe et al., 2004).

nel layout is designed to allow the injection of two fluids into a central waveguide. CaCl_2 ($n = 1.445$) is injected at the center, whereas DI (de-ionized) water ($n = 1.335$) flows on either side of the central stream; the high-index liquid thus forms a waveguide core, with a water cladding. Typical flow rates are $175 \mu\text{l}/\text{min}$, and changing the flow conditions allows a dynamic change of waveguide width and direction. Single-mode waveguides, with a liquid core width of about $10 \mu\text{m}$, can thus be defined. In addition, varying the flow rates can change the direction of the liquid core, so that the system can be reconfigured two-dimensionally. This approach has been used to realize an optical switch, allowing dynamic fluidic (and thus optical) switching of the waveguide into one of three output channels of an optofluidic Y-coupler.

In addition to the use of L^2 waveguides as passive and switchable waveguiding structures, they also form the basis for one family of optofluidic lasers, as we shall see below.

13.3 Optofluidic lasers

One of the most active areas of optofluidics is in the conception and fabrication of fluidic lasers, generally microfabricated dye lasers based on the same principles as their macroscopic counterparts. These employ a fluorescent dye which is typically optically pumped using an external light source. The optofluidic variants use microfluidic components and systems to realize miniaturized versions of this type of laser; these structures benefit from the use of small pumps which nevertheless lead to high pump speeds, and the resultant high flow speeds of the dye result in reduced photo-bleaching.

13.3.1 Fabry-Perot optofluidic dye lasers

The earliest example of a microfluidic dye laser employed a vertical Fabry-Perot cavity fabricated using glass and SU-8 materials and the techniques of microsystems assembly

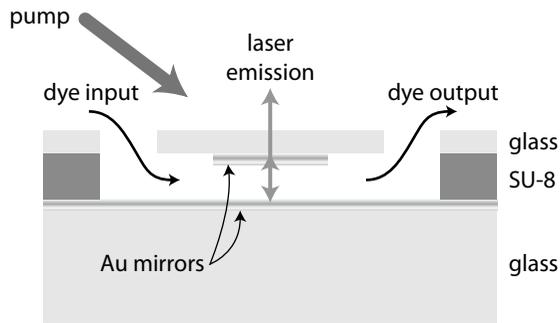


Figure 13.2 Schematic cross-section of an optofluidic Fabry-Perot dye laser. The resonant cavity is formed by two Au mirrors deposited in the glass surfaces defining the upper and lower sides of the fluidic channel. Optical pumping is by a Nd:YAG laser and emission is vertically from the substrate. Diagram based on (Helbo et al., 2003).

(Helbo et al., 2003); the structure is shown schematically in Figure 13.2. The vertical laser cavity, defined by the thickness of the SU-8 spacer layer, is $15\ \mu\text{m}$ long and is delimited by Au-coated facets.

The microfluidic channel is filled with rhodamine 6G dye⁴, typically diluted to a concentration of 10^{-2} mol/l in ethanol; this dye fluoresces in the yellow for optical excitation in the green. As with most of the other optofluidic lasers we discuss here, this device is optically pumped, typically using a frequency-doubled Nd:YAG laser emitting at 532 nm. Above a threshold pump power density of about $34\ \text{mW}/\text{cm}^2$, the optofluidic structure lases at 570 nm with a spectral linewidth of 5.7 nm. Due to the need for an external pump laser, this microfabricated laser is not completely integrated, but an array of these could conceivably be pumped using a single external light source.

Numerous variations on this structure have been demonstrated. Using the same dye and optical pumping scheme, a longer lateral cavity, which emits from the substrate edges, has been defined using femtosecond laser direct writing in glass (Cheng et al., 2004). Monomode emission has been obtained by integrating a 130th order Bragg grating into the cavity of a lateral in-plane structure, consisting of $16\ \mu\text{m}$ wide, $10\ \mu\text{m}$ deep channels (Balslev and Kristensen, 2005). Optical pumping using 5 ns pulses repeated at 10 kHz resulted in single-mode lasing at 577 nm. An alternative means to realize such a structure has been to generate the grating using nano-imprint lithography in cyclo-olefin copolymer (COC)⁵ (Nilsson et al., 2005).

A further step in system miniaturization of an optofluidic Fabry-Perot laser is the use of a closed-loop fluidic circuit for the dye (Galas et al., 2005). Using soft lithography to define SU-8 and PDMS fluidic channels on a $5 \times 3\ \text{mm}^2$ glass chip, the same dye and optical pump configuration as described above was employed. A flow rate of up

⁴ Rhodamine represents a family of fluorescent dyes popularly used as a gain medium in dye lasers and in many biophotonic applications, particularly fluorescence microscopy.

⁵ COC is a biocompatible, thermo-plastic polymer, frequently used in biophotonic applications, a popular proprietary version of which is sold under the brand name TOPAS®.

to 300 nl/min (5 mm/s) could be achieved using the integrated pumps, and it was seen that a pump rate greater than about 1.5 mm/s avoids photobleaching in this miniaturized system. In addition, by changing the dye concentration in the range 10^{-2} to 10^{-3} mol/l, the emission wavelength could be tuned, from 569 to 561 nm, respectively.

A final variation of optofluidic Fabry-Perot lasers has demonstrated how multiple-wavelength emission is possible when two dyes are mixed in the active region (Kou et al., 2006). Using a mixture of 10^{-3} mol/l rhodamine 6G and 10^{-2} mol/l rhodamine 101 in a polymer-based vertical cavity defined by Au-coated fiber facets spaced 140 μm apart, simultaneous emission at 559 and 597 nm was observed. Using continuously varying concentrations of dyes pumped through the cavity, the relative intensities of emission could conceivably be continuously varied.

13.3.2 Waveguide optofluidic dye lasers

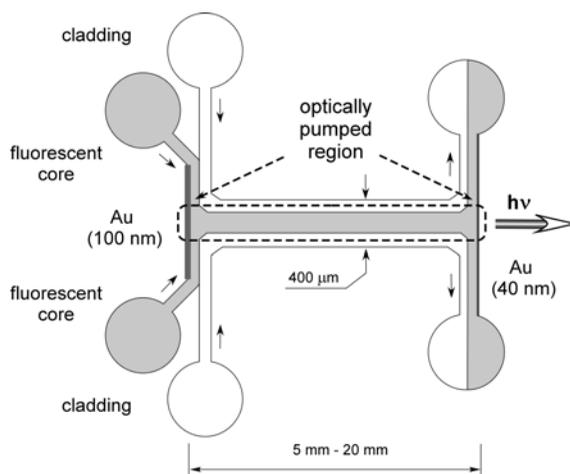
The L^2 waveguides described in Section 13.2 may also form the basis for another family of optofluidic dye lasers. Again using liquids with a higher refractive index than the surrounding “cladding” liquid (DI water), in this case various fluorescent dyes diluted in ethylene glycol, these are combined using laminar flow in a PDMS-based fluidic channel. Pumping using a tungsten halogen lamp results in planar (in the plane of the substrate) light emission, with a broad spectrum centered at various wavelengths depending on the dye used, and is thus a fluorescent light source useful for optical stimulation in microfluidic systems (Vezenov et al., 2005b). Rhodamine 6G dye results in emission around 582 nm, fluorescein at 547 nm, and coumarin 343 at 497 nm; optical conversion efficiency of this structure is about 0.01 %.

By defining a resonant cavity, this configuration can be used to achieve optofluidic lasing (Vezenov et al., 2005a). Using the structure shown in Figure 13.3(a), the 10 mm long resonant cavity in the PDMS waveguide with a cross-section of $100 \times 400 \mu\text{m}^2$, is defined by the Au coated mirrors at either side. Employing a core liquid of rhodamine 640 dye (2 mM dye, diluted in methanol) and a cladding material of pure methanol, the laser was again externally pumped using a frequency-doubled Nd:YAG laser (532 nm) resulting in laser emission at $\lambda = 625$ nm, as seen in Figure 13.3(b).

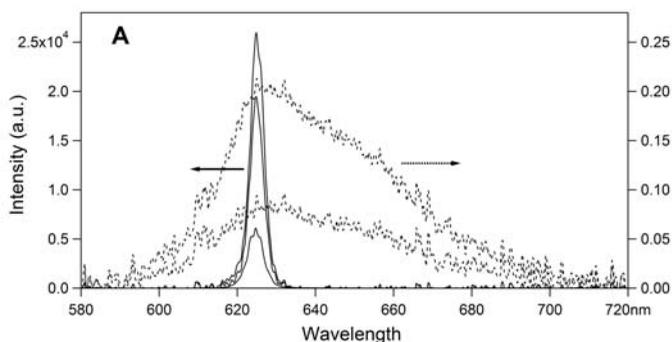
By altering the flow conditions in the liquid waveguide, the laser characteristics can likewise be changed. Variation of the refractive index step between the core and cladding, for example, can be used to optimize the lasing transition as well as the angular width of the emission. In addition, varying the composition of the dye liquid, in particular using a continuous mixture of different components, can allow continuous tuning the the emission wavelength; a range of about 615 to 635 nm was demonstrated using this approach.

13.3.3 DFB optofluidic dye lasers

A further alternative optofluidic laser device employs a DFB structure for achieving feedback and resonance in the laser cavity, as seen schematically in Figure 13.4(a). Using molding, a 15th order Bragg grating has been incorporated into a PDMS-based



(a) Schematic top view of the optofluidic laser. The 10 mm long laser cavity, found between the Au mirrors, is pumped using a frequency-doubled Nd:YAG laser and emission is in the plane of the substrate as shown.



(b) Optical output below *dashed* and above *solid* laser threshold; note the difference in scales.

Figure 13.3 Schematic sketch and output characteristics of an L^2 waveguide-based laser. Figures from (Vezenov et al., 2005a); copyright © 2005, American Chemical Society; reprinted with permission. Figures courtesy of George Whitesides, Harvard University.

microfluidic channel, resulting in a DFB structure (Li et al., 2006b); the very high order allows the use of relatively large periods ($3.08 \mu\text{m}$), easily fabricated using soft lithography. Again when filled with a rhodamine 6G dye (diluted to 1 mM), the 532 nm pump ($150 \text{ kW}/\text{cm}^2$, using 6 ns pulses repeated at 10 Hz) resulted in narrow-band laser emission at 567 nm, with a linewidth of $\Delta\lambda = 0.21 \text{ nm}$.

Perhaps more interesting, however, is the realization of a *tunable* DFB laser using this approach (Li et al., 2006c). The same fluidic channel structure, with a cross-section of $2 \times 3 \mu\text{m}^2$, was fabricated using the 15th order Bragg grating with a $\pi/2$ phase shift in

the center to force single longitudinal mode operation. Due to the low refractive index step between liquid and channel sidewalls (< 0.003), only single lateral and transverse modes are supported. The dyes used were either rhodamine 6G (for emission around 580 nm) and rhodamine 101 (for emission around 625 nm), diluted in methanol or ethylene glycol.

Tunability of this laser was attained by mechanically stretching the entire flexible PDMS cavity using two translation stages attached at either end. The entire 1 cm long cavity was stretched by 400 to 500 μm , resulting in a change in grating period of $\Delta\Lambda$ between 25 and 28 nm. This level of $\Delta\Lambda$ gave rise to an emission wavelength tuning range of almost 30 nm, around 580 or 625 nm. The approach is completely general, and, using suitable dyes, tunable lasers emitting between 320 and 1200 nm are conceivable. Lower-order (first, second and third order) gratings have been used in this design to strongly reduce the lasing threshold pump energy (Song et al., 2009b); the 5 mm long grating sections had periods of 204, 408, and 612 nm, respectively, and led to the optical emission characteristics shown in Figure 13.4(b). Threshold pump energy was 78 nJ per pulse for these structures.

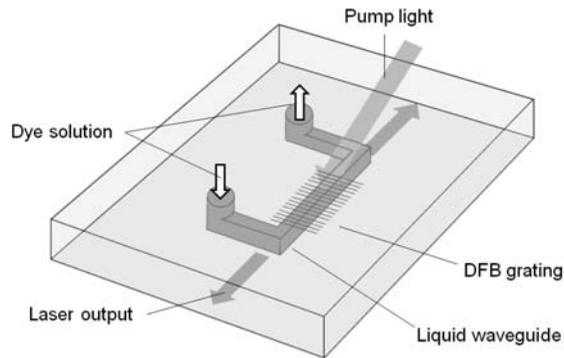
As an extension of this work, an array of five waveguides with varying Λ has been fabricated in parallel on the same substrate; pumped by a single external pulse, simultaneous multiple-wavelength emission could be demonstrated, suggesting great flexibility in the emission characteristics of opto-fluidic laser arrays.

13.3.4 Vertically-emitting optofluidic dye lasers

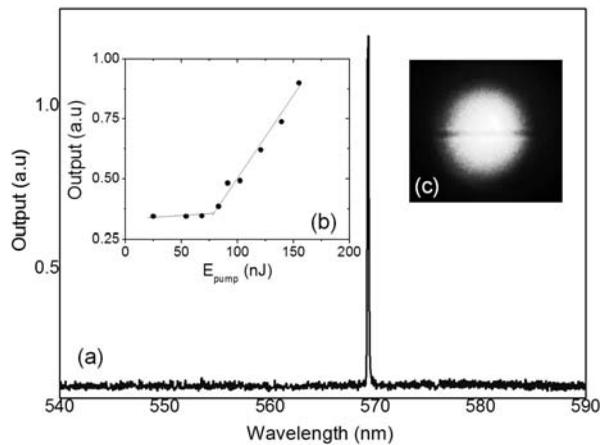
Finally we consider vertically-emitting optofluidic dye lasers, which use a circular grating defined either in perfluoropolyether (PFPE), a polymer material stiffer than PDMS and thus suitable for the definition of high-resolution features (Chen et al., 2009), or SU-8 (Song et al., 2009c). The cavity of the former was formed using the PFPE substrate, with the grating patterned using soft lithography, and a PDMS capping layer; the gain medium was again a rhodamine 6G dye. The circular grating, with a period of $\Lambda = 410$ nm was designed to achieve in-plane feedback by reflection in the $m = 2$ order, but generate coupling out of plane, leading to vertical emission, using the $m = 1$ order of the grating. Pumping at 532 nm led to the emission of a vertical laser mode at $\lambda = 580$ nm, for pump energy densities above 6 $\mu\text{J}/\text{mm}^2$. This structure may be promising for the realization of two-dimensional arrays of vertically-emitting lasers, the exact emission wavelength tunable by changing the geometry of the grating, but requiring only a single pump laser. The low energy threshold suggests that pumping using LEDs may be possible.

13.4 Fluidic microlenses

We have considered the most popular forms of fluidic microlenses in Sections 12.1 and 12.2. There are, however, a number of more exotic optofluidic microlens concepts, in particular those which employ the hydrodynamics of fluid flow. The examples of liquid



(a) Schematic representation of the DFB laser, fabricated using PDMS; the grating section is 5 mm long. Optical emission is in the plane of the substrate.



(b) Emission spectrum *a*; PI characteristic *b*; and optical near-field intensity profile *c* of the opto-fluidic DFB laser. Emission is at 569.1 nm for a threshold pulse energy of 78 nJ. Figure from (Song et al., 2009b); copyright © 2009, American Institute of Physics; reprinted with permission.

Figure 13.4 An opto-fluidic DFB laser based on a dye-filled microfluidic cavity and a molded grating. Figures courtesy of Demetri Psaltis, EPFL.

microlenses we saw in Chapter 12 used static liquid formations, either as droplets or in membrane-defined cavities; tuning was accomplished by relatively slow variations in the liquid profile.

One can, however, also use hydrodynamics to form a lens, by carefully controlling the flow conditions; these are achieved using flow at low Reynolds numbers such that viscous forces dominate the fluid behavior and intermixing of two liquids is only by diffusion, not due to turbulence. An in-plane microlens (implying that the optical axis is in

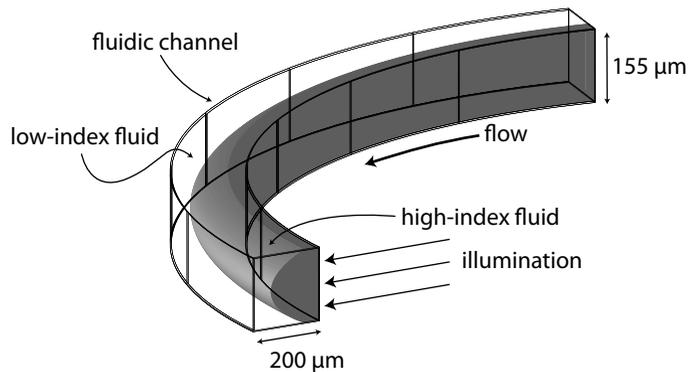


Figure 13.5 Schematic representation of a lateral microlens based on the laminar flow of two liquids (low index H_2O ($n = 1.335$) and high index CaCl_2 ($n = 1.445$)) in a curved fluidic channel; illumination is from the right as shown. Diagram based on (Mao et al., 2007).

the plane of the substrate) has been configured using two liquids with varying refractive index, undergoing laminar flow in a microfluidic channel (Tang et al., 2008). The width of this channel broadens at the position of the lens, such that the liquid with higher n expands two-dimensionally, attaining a spherical profile when seen from above. By varying the flow conditions and optimizing the shape of the expansion chamber (Song et al., 2009a), the radius of curvature may be dynamically tuned; variations in focal length in the range $6 \leq f \leq 12$ mm have been demonstrated.

An alternative approach relies on laminar flow of two liquids in a curved microfluidic channel with a rectangular profile (Mao et al., 2007). Due to the hydrodynamics of fluid flow in this structure, the boundary region between the two liquids (which again have differing indices of refraction) can take on a convex profile, as shown schematically in Figure 13.5; focussing is thus in the direction normal to the plane of the substrate, with the optical axis in that plane, as before. Using H_2O ($n = 1.335$) and CaCl_2 ($n = 1.445$) as the two fluids, it was seen that focal length decreases linearly with flow rate, demonstrated in the range from 0 to 200 $\mu\text{l}/\text{min}$.

13.5 Filters and displays

Optofluidic systems have been employed for filtering optical fields, effects which may ultimately be useful for displays; these represent some of the most commercially advanced optofluidic components. In both filters and displays, the spectrum or intensity of a transmitted optical signal may be variably attenuated using fluidic components.

13.5.1 Filters

The ring resonators to which we were introduced in Section 9.5.4, which form the basis for the disk lasers of Section 10.3.5, may also be operated in liquid. Small changes in the refractive index of the liquid surrounding the resonator changes the effective waveguide

index and thus the resonant frequency. As a result, the resonator may either be used as a sensor, responding to ambient refractive index changes, or, alternatively, the external liquid index may be varied to tune the resonant wavelength. The concept may thus be used as a liquid tunable optical filter.

This functionality has been demonstrated using a SiO₂-based ring resonator integrated with a microfluidic chip (Armani et al., 2005). It was seen that the resonant frequency could be tuned by up to 2 nm, and that both coupling between resonators and waveguides as well as their losses are a function of the ambient liquid refractive index. Conceptually similar is the use of a fluid as the cladding of a solid waveguide (Zhu et al., 2005); by varying the ambient liquid refractive index to a value close to that of the solid waveguide core, the confinement of the mode decreases, losses increase and the transmitted optical field is attenuated.

13.5.2 Displays

There has been considerable academic and industrial effort expended on the development of displays using electrowetting, and this application may be the most advanced device implementation using optofluidics (Hayes and Feenstra, 2003). The concept is based on two liquids, an opaque or colored oil in a transparent water-based compound, deposited in a cell (one pixel) on a white reflecting substrate. Without bias, the opaque oil results in little or no reflection, and thus generating a black pixel; using electrowetting, the opaque fluid may be moved away from the substrate, leading to a white pixel, upon the application of a voltage. Using a 15 μm oil film in a $250 \times 250 \mu\text{m}^2$ pixel, reflectivity changes of 35% and a contrast level of 15 have been demonstrated using this approach. Color displays may be realized using color filters positioned over this arrangement.

This optofluidic electrowetting-based technology is now quite advanced and pixels with dimensions down to $50 \times 150 \mu\text{m}^2$ with low driving voltages have been demonstrated (Zhou et al., 2009). An example of a display using $300 \times 900 \mu\text{m}^2$ pixels fabricated using an SU-8 process with In₂O₃ : SnO₂ electrodes is shown in Figure 13.6 (Lao et al., 2008). Using a water/oil-based system, a bias of 9 V is required to switch between black and white pixels. For the “on” state, yielding a white pixel, 64% of the area is clear, and an integrated Al reflector below the pixel leads to transmission of up to 90% of the incident light. More advanced systems use electromechanical pressure to pump pigments into and out of a pixel, resulting in bright color displays based on this optofluidic technology (Heikenfeld et al., 2009); these devices are being commercialized as a form of “electronic paper”.

13.6 Other optofluidic concepts

Many other novel devices using the concepts of optofluidics populate the scientific literature; the field is seeing rapid development and the creativity of micro-opticians seems boundless.

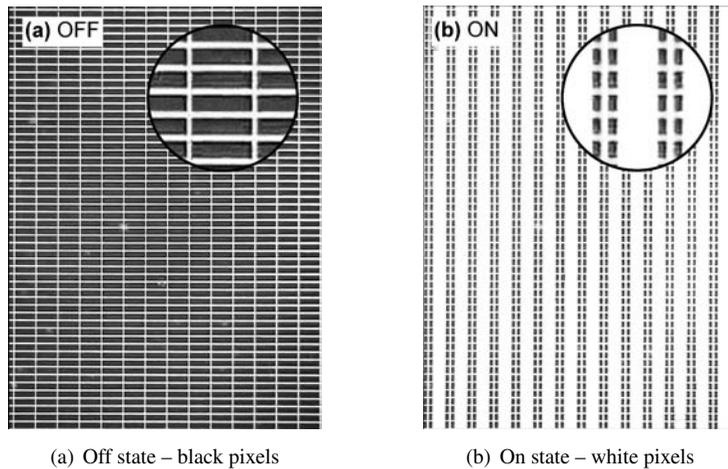


Figure 13.6 An electrowetting display based on a water/oil system with $300 \times 900 \mu\text{m}^2$ pixels; shown are the “on” and “off” states yielding black and white pixels, respectively. Figures from (Lao et al., 2008); copyright © 2008, IEEE; reprinted with permission. Figures courtesy of Jason Heikenfeld, University of Cincinnati.

13.6.1 Fluidic-tunable photonic crystals

Photonic crystals, one of the subjects of Chapter 14, may be tuned, as we saw in Section 12.4.3. Since, as we saw in a few examples above, microfluidic systems may be used to implement fluid flow of variable refractive index through a given system, this approach may be used to tune photonic crystals. For example, a SOI-based photonic crystal, consisting of 140 nm holes repeated with a period of 434 nm, has been combined with a microfluidic network fabricated in a silicon elastomer (Erickson et al., 2006). Using a mixture of two miscible fluids with different refractive indices, H_2O and CaCl_2 , a row of holes in the photonic crystal could be selectively filled with a fluid of variable index by placing the fluidic channel over a single row in the periodic structure. As a result, the transmission properties of the photonic crystal could be dynamically changed, changing the transmission wavelength from 1 491 to 1 502 nm.

Along the same lines, it has also been proposed that certain holes in a two-dimensional photonic crystal network could be filled and emptied using an microfluidic network (Intonti et al., 2006). By matching the index of the surrounding material, the effect of a hole can be negated, such that two-dimensional photonic crystal circuits may be “written” using this approach.

13.6.2 Microscopy

A further novel optofluidic implementation is the use of optofluidics for the realization of a compact liquid microscope (Heng et al., 2006). Using an array of holes as optical apertures, the device functions essentially as a large array of scanning near-field optical

microscopes. The biological sample, a nematode in the demonstration presented, flowed past the aperture array and its structure was imaged sequentially. A resolution of about $0.5\ \mu\text{m}$ was achieved using this novel imaging arrangement.

13.6.3 Optofluidic tweezers

Finally, optical tweezers, the use of light to trap and manipulate micrometer-sized objects ranging from nanospheres to cells, have been implemented using miniaturized optofluidic concepts. A particle in liquid suspension, for example, has been optically trapped in a gap between two fibers (Domachuk et al., 2005). By manipulating the position of the particle, which may be a spherical bead and thus act as a micro-ball lens, the optical coupling between the fibers may be modulated. As a result, an optofluidic reconfigurable optical coupling system may be realized, allowing a novel means of optical modulation.

14 Nano-optics

*Le biologiste passe, la grenouille reste.*¹

Jean Rostand

As we saw in the historical introduction, optics is a field with a long and rich history, through which dozens of generations of scientists and engineers have passed, each contributing a little more to the understanding of and creative uses for light. As scientific and engineering developments have made the micro and nano worlds more accessible, new aspects of optics have come to light; nano-optics is thus a field whose physics and applications have only begun to be appreciated but has considerable promise for novel and still unknown science and engineering.

We conclude this text with a look at several aspects of micro-optics which employ nanometer-scale structures to achieve new forms of functionality. The nano-optics field profits from an extensive overlap of developments in basic physics, particularly quantum optics, classical and micro-optics and advanced fabrication technology. In contrast to mainstream nanotechnology, the prefix “nano” for optics does not necessarily always imply the use of structures on nanometer size scales, for which smaller is always better; the designation is frequently employed for optical systems with feature sizes smaller than the wavelength, sometimes thus also referred to as sub-wavelength optics. We consider here micro- and nano-optical concepts with size scales ranging from about a micrometer to a few nanometers; we will focus on devices rather than on fundamental physical effects, of which there is nevertheless a rich variety.

14.1 Nanophotonics

Nanophotonics, which represents the convergence of micro-optics, photonics and nanotechnology, is expected to yield significant technical and scientific, as well as economic, impact. We begin here with a smattering of topics on which researchers have focussed in this area to provide a flavor for the field, before addressing a few of these in detail in the following sections.

¹ “The biologist passes on, the frog stays.” Jean Rostand (1894-1977), French herpetologist, in *Inquiétudes d'un biologiste* (Paris, Gallimard, 1967), p. 66

14.1.1 Physics and devices

Work in nanophotonics is taking place on a broad front, addressing fundamental physical questions and developing applications. Advances in nanoscale physics, for example, are facilitated by the highly concentrated optical fields achievable in nanophotonic systems, as are new capabilities in near-field measurements and photon localization (Shen et al., 2000). An important area in nanophotonics is that of near-field imaging, and the interactions of photons with electrons on the nanoscale has provided numerous new tools for high-resolution microscopy (Novotny and Hecht, 2006).

Engineering work has focussed on areas in which nanophotonic concepts allow new types of functionality for small-scale optical systems. We have seen some of the resultant concepts previously, such as the use of quantum dots to enhance the performance of LEDs in Section 10.2.5, microring resonators used as filters or lasers, in Sections 9.5.4 and 10.3.5 respectively. These structures are seeing further development, using the technologies of nanophotonics to realize numerous applications, including, for example, wavelength-insensitive optical switches (Vlasov et al., 2008).

Microring resonators have also been used to demonstrate optical actuation of micro-opto-mechanical structures. Using a microdisk resonator placed close to a suspended linear waveguide, the intense evanescent field due to the guided modes has been used to attract the two structures to one another (Eichenfield et al., 2007). Alternatively, two closely-spaced suspended ring resonators can exhibit attractive and repulsive behavior, and thus be bent toward or away from each other, the required forces again based solely on the optical fields propagating in the waveguides (Wiederhecker et al., 2009). This cavity-enhanced optical dipole force can result in nanometers to micrometers of movement for milliwatts of guided optical power.

Self-assembly techniques, which play an important role in the fabrication of many types of nanophotonic devices and systems, have been employed to realize nanoscale refractive lenses, with diameters ranging from 3 μm to 50 nm (Lee et al., 2009). These sub-wavelength structures, fabricated using organic compounds, have been shown to have extremely short focal lengths when operated in the near-field, and may be used for imaging beyond the diffraction limit in a near-field configuration.

A final area currently of considerable interest is that of optical antennas (Muhlschlegel et al., 2005). These typically consist of metallic dipoles fabricated on an insulating substrate. With lengths ranging from 190 to 400 nm and widths of 45 nm, these gold structures are resonant at optical frequencies and are used to excite surface plasmons (the subject of Section 14.3), concentrate optical fields or to generate a supercontinuum spectrum².

² A supercontinuum spectrum is one with a large spectral bandwidth but high spatial coherence, and may be imagined as laser emission with a broad spectrum. Supercontinuum spectra are typically generated using laser pulses propagated through a non-linear medium, often a fiber; see Section 14.2.6.

14.1.2 Technological impact

The combination of developments in optics and nanotechnology will likely contribute to advances in a number of technological areas (MONA, 2008; Jenkins, 2008). In particular, it is envisaged that quantum dots and wires, high index-contrast structures (notably photonic crystals) and nanotubes will have a significant impact in a variety of applications.

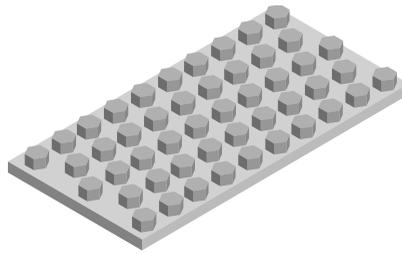
Of the technologies which stand to benefit from advances in nanophotonics, displays will probably represent the most significant market share; flat panel displays are expected to benefit greatly from development and incorporation of carbon nanotubes. Also of economic significance is the area of photovoltaics, in which quantum dots will play an increasingly large role. In addition, nanowires and photonic crystals for enhanced extraction of light from LEDs will enhance the capabilities of solid state lighting; quantum dots as well as plasmonic detectors and lenses will contribute to improved performance of imaging systems; and new generations of optical sensors will rely on sensitized quantum dots and nanowires. Finally, optical interconnect and telecommunications, often the driving forces behind developments in micro-optics and photonics, will benefit from new types of electronic/photonic integration, made possible by advances in nanophotonics.

14.2 Photonic crystals

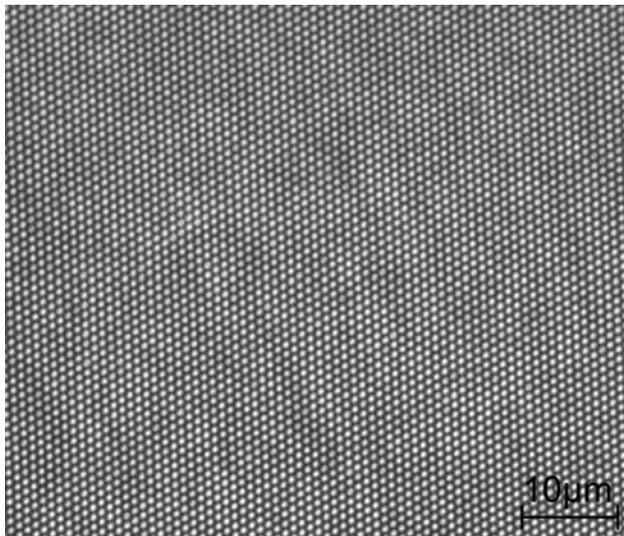
Photonic crystals represent one of the earliest developments in micro-optics, which paved the way for the increasingly refined structures of nano-optics. A photonic crystal is a spatially repeating structure with periods in the sub-micrometer regime; due to the periodic nature of its features, the photonic crystal can exhibit a “photonic bandgap”, analogous to the electronic bandgap found in crystalline semiconductors due to the periodic arrangement of atoms. As a result, the transmission of photons in certain energy ranges is inhibited in photonic crystals, a feature which leads to numerous very attractive applications and has stimulated a large body of research in the past fifteen years (Joannopoulos et al., 1995; Inoue and Ohtaka, 2004).

14.2.1 Photonic crystal structures

Photonic crystals exist as one-dimensional (1D), two-dimensional (2D) and three-dimensional (3D) structures, depending on the number of spatial directions in which a periodic variation of the refractive index is realized. A grating, such as that shown in Figure 8.22, or a planar Bragg mirror, as we analyzed in Section 5.3.5, for example, could represent 1D photonic crystals. Due to the popularity of the topic in the photonic community, the technical literature is replete with references to photonic crystals even when the structures are simple gratings; although there is no clear distinction between “grating” and “1D photonic crystal”, most photonic crystal structures (which exhibit an optical bandgap) have a relatively large refractive index step between the periods,



(a) Schematic representation of a 2D photonic crystal, consisting of a two-dimensional arrangement of hexagonal posts.



(b) A 2D hexagonal photonic crystal fabricated in photoresist using two-beam interference lithography and dual exposure. The period is $1.14\ \mu\text{m}$. Photo courtesy of Christoph Schlägl.

Figure 14.1 2D photonic crystals consisting of hexagons in a hexagonal arrangement, shown schematically and fabricated in photoresist.

usually 2 or more; this criterion distinguishes photonic crystals from other periodic structures, which do not usually exhibit bandgap behavior.

2D and 3D structures

A 2D photonic crystal arrangement is that such as shown in Figure 14.1; the schematic of Figure 14.1(a) shows a two-dimensional arrangement of hexagonal posts, realized in photoresist in Figure 14.1(b). An alternative 2D concept would be an array of holes in a slab, as we will see presently.

A complete 3D photonic crystal, finally, such as that shown in Figure 14.2, has periodically varying features in all three spatial dimensions and is that which most closely

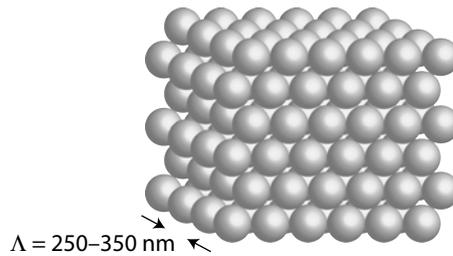


Figure 14.2 Schematic representation of a 3D photonic crystal consisting of an arrangement of stacked spheres. If the spheres have a higher refractive index than the space between them, the structure corresponds to that of an opal.

represents the semiconductor crystal analogue. In the arrangement shown, the spheres have a higher refractive index than the surrounding space, representing the structure of an opal³. The inverse structure (consider the spheres to be holes in a material with higher n), is thus an inverse-opal, and also possibly a form of photonic crystal. Not all arrangements of spheres (or other shaped features) lead to photonic bandgap behavior and usually a diamond lattice structure is desirable.

For use at visible optical wavelengths, the 1D, 2D and 3D structures typically have periods, a [m], in the range 250 to 350 nm, such that feature sizes are in the range of 100 nm, and frequently smaller. In addition, the required accuracy and repeatability (in position and size) of the periodic features, over at least ten unit cells, is typically measured in the tens of nanometers, explaining why photonic crystals are generally considered “nano-optical” structures. The required Δn , as we mentioned above, is generally in the range 2 to 3.5, so that only a limited range of material combinations is suitable to realize photonic bandgap behavior.

Slab structures

Although considerable effort has gone into the fabrication of 3D photonic crystals, and this arrangement is still an enviable goal for many researchers, as we shall see below, 2D slab structures have proven not only to be more easily fabricated but also to have very useful optical properties (Johnson et al., 1999). The 2D slab photonic crystal, shown schematically in Figure 14.3, has a two-dimensional periodic pattern defined in a planar membrane. As a result, light is confined in the third dimension due to total internal reflection in the membrane (which essentially represents a slab waveguide) and photonic bandgap effects are restricted to light propagation in plane of the membrane.

The slab structure has the advantage that it may be much more easily fabricated than true 3D photonic crystals and we will see below that it forms the basis for a wide variety of photonic crystal devices.

³ An opal is a semiprecious gemstone which has a microscopic structure (hexagonally close-packed silica spheres with a diameter of about 250 nm) of the form shown in Figure 14.2; due to interference effects resulting from the 3D periodic patterns, the stone has a pleasing iridescent quality. It is, however, porous and thus relatively delicate; washing up the dishes while wearing an opal ring is not to be recommended.

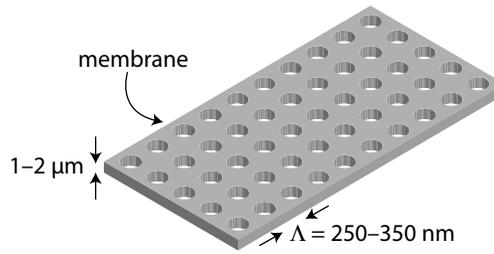


Figure 14.3 Schematic representation of a 2D slab photonic crystal. The structure is periodic in two dimensions and relies on confinement due to the refractive index step in the third (vertical) dimension. Not to scale.

14.2.2 Theoretical analysis

When an electromagnetic field with wavelength λ is incident on a photonic crystal, the periodic refractive index generates forward and backward propagating fields, which interfere with each other and result in complex propagation behavior. At certain λ , the resulting interference effects give rise to a range of prohibited incident wavelengths, generating a photonic bandgap. As we might expect, these effects scale with size, so that the effects are usually a function of λ/a ; optical frequencies thus require sub-micrometer periods.

A theoretical analysis of light propagation through a photonic crystal can become quite involved, particularly for 3D structures (Fan et al., 2006). The fundamental calculation requires vectorial solution of Maxwell's equations in a system with periodically repeating boundary conditions (Joannopoulos et al., 1995). The approach closely parallels the calculation of the electronic bandgap due to the effects of a periodic crystalline lattice, no doubt familiar to students with a rudimentary background in solid-state physics; see for example (Kittel, 2005, Chapter 7) or (Brennan, 1999, Chapter 8).

Simple model

To illustrate the origin of the bandgap, we can analyze a 1D system using an analysis of coupled modes, inspired by the approach taken in (Solgaard, 2009, Chapter 14). For propagation of an electromagnetic field⁴ through a 1D Bragg grating, for example, the periodic variation of Δn gives rise to forward and backward propagating waves; this effect has been analyzed for and forms the functional basis of optical multilayers (Section 5.3), waveguide grating couplers (Section 9.4.2), DBR lasers (Section 10.3.3), and VCSELs (Section 10.3.4).

Using the coupled mode approach in one dimension (Zappe, 2004, Chapter 4.4.2), we can consider the effect of a periodic corrugation with period Λ , which leads to the Bragg (resonance) condition β_0 ,

$$\beta_0 = \frac{m\pi}{\Lambda}, \quad (14.1)$$

⁴ We consider only the electric field here; most rigorous photonic crystal calculations employ the magnetic field, since the mathematics is somewhat simplified in that case.

on the propagation of electromagnetic fields with unperturbed propagation constant β ; the index m represents the mode order. If we define the difference between the unperturbed propagation constant and the Bragg condition as

$$\Delta\beta = \beta - \beta_0 = \beta - \frac{m\pi}{\Lambda}, \quad (14.2)$$

then the structure will support forward and backward propagating fields, E_F and E_B , in the z direction,

$$E_F = E_{F0} e^{+j\Delta\beta z} \quad (14.3)$$

and

$$E_B = E_{B0} e^{-j\Delta\beta z}, \quad (14.4)$$

respectively.

The total electric field, E_z , is then found from the sum

$$E_z = E_F e^{+j\beta_0 z} + E_B e^{-j\beta_0 z} \quad (14.5)$$

where the forward and backward propagating fields are coupled by the relationships

$$\frac{dE_F}{dz} = \kappa_m E_B e^{j2\Delta\beta z} \quad (14.6)$$

$$\frac{dE_B}{dz} = \kappa_m E_F e^{-j2\Delta\beta z}. \quad (14.7)$$

In these two previous expressions, κ_m [m^{-1}] represents the coupling coefficient, which is a function of the mode order m as well as the amplitude and form of the periodic refractive index corrugation, Δn .

Solving Equations 14.6 and 14.7 yields forward and backward propagating waves, where the propagation constant for the perturbed coupled waves, β_{1D} [m^{-1}], is given by (Solgaard, 2009, Chapter 6.6)

$$\beta_{1D} = \beta - \Delta\beta \pm j\sqrt{\kappa_m^2 - \Delta\beta^2}. \quad (14.8)$$

Combining Equations 14.2 and 14.8, we obtain

$$\beta_{1D} = \beta - \left(\beta - \frac{m\pi}{\Lambda}\right) \pm j\sqrt{\kappa_m^2 - \left(\beta - \frac{m\pi}{\Lambda}\right)^2} \quad (14.9)$$

which may be solved for β to yield

$$\beta = \frac{m\pi}{\Lambda} \pm \sqrt{\kappa_m^2 + \left(\beta_{1D} - \frac{m\pi}{\Lambda}\right)^2}. \quad (14.10)$$

This last relationship may be normalized by Λ/π , yielding

$$\beta \frac{\Lambda}{\pi} = 1 \pm \sqrt{\left(\kappa_m \frac{\Lambda}{\pi}\right)^2 + \left(\beta_{1D} \frac{\Lambda}{\pi} - 1\right)^2} \quad (14.11)$$

or

$$\beta' = 1 \pm \sqrt{\kappa'^2 + (\beta'_{1D} - 1)^2} \quad (14.12)$$

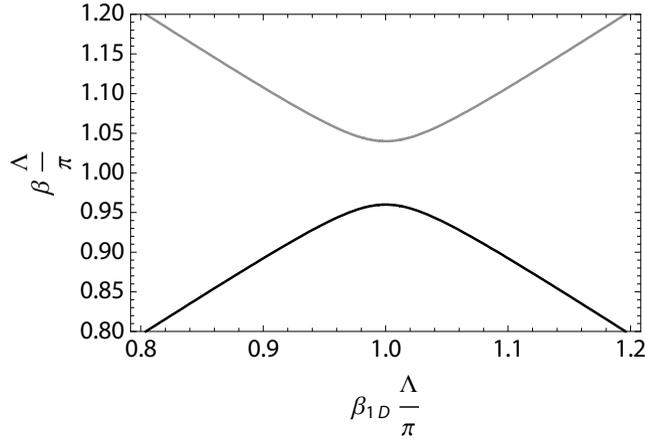


Figure 14.4 Plot of the bandgap which results from propagation of an electromagnetic wave through a 1D photonic crystal, as calculated by the coupled-wave approach given by Equation 14.11, for $m = 1$.

which is only a function of the dimensionless parameters

$$\beta' = \beta \left(\frac{\Lambda}{\pi} \right) \quad (14.13)$$

$$\beta'_{1D} = \beta_{1D} \left(\frac{\Lambda}{\pi} \right) \quad (14.14)$$

$$\kappa' = \kappa_m \left(\frac{\Lambda}{\pi} \right). \quad (14.15)$$

We recall that the propagation constant β , that of the incident electromagnetic field, is defined as

$$\beta = \frac{2\pi}{\lambda} = \frac{2\pi n}{\lambda_0} = \frac{2\pi n\nu}{c} = \frac{2\pi n}{hc} E \quad (14.16)$$

for photon energy E and material refractive index n . Since β_{1D} represents the propagation constant for the perturbed waves propagating in the 1D photonic crystal, Equation 14.11 relates the propagation of the coupled, perturbed waves to the energy of the incident field. Plotting β as a function of β_{1D} will then allow us to see how propagation through the structure varies with incident photon energy.

The resultant relationship is plotted for the normalized case in Figure 14.4. We see from the characteristic that, far from the resonance condition, β_{1D} and β are linearly related. Close to resonance (the Bragg condition, for which $\beta \approx \beta_0$, so that $\Delta\beta$ is close to zero), however, a gap in the allowed energies appears. For a range of β (and thus E) values, there are no solutions for β_{1D} and thus propagation through the structure is not allowed. This range corresponds to a photonic bandgap.

Rigorous modeling

This rudimentary model is illustrative, but the band structure of real photonic crystals is generally calculated using commercially-available software, which also yields characteristics for 2D and 3D structures of arbitrary shape. An example of the results of such a calculation is shown in Figure 14.5, in which the photon energy (proportional to a/λ , for photonic crystal period a and incident photon wavelength λ) for TE and TM polarizations is plotted as a function of wave-vector for a number of crystalline directions. The first four bands are shown, and we see that there is a pronounced bandgap between the first and second bands for TE, and between the second and third for TM. In higher bands (not shown), further bandgaps may exist.

14.2.3 Fabrication techniques

Whereas the theory of photonic crystals was developed over twenty years ago, experimental realization took some time, since the required feature sizes and their tolerances are in the nanometer range. A number of techniques have emerged in the ensuing years, allowing these high-resolution optical structures to be fabricated in a wide variety of configurations. We only look at a few here; the literature teems with further creative examples.

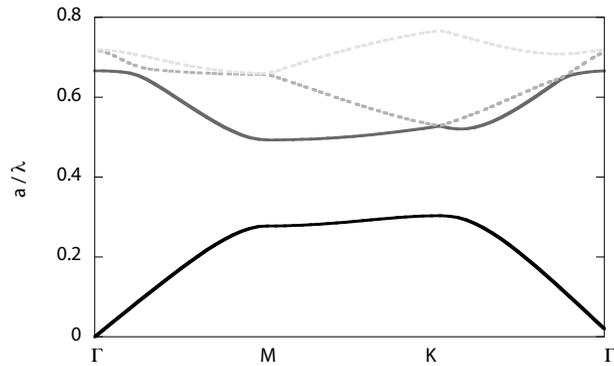
2D techniques

Two-dimensional photonic crystals are (relatively) easy to fabricate, since standard high-resolution photolithography, as well as etch and deposition technologies, may be employed. Due to the required feature sizes, electron beam lithography is usually required. This approach has been widely used to define the 2D slab structures shown in Figure 14.3; the slab, which can be an epitaxially grown on a III-V semiconductor-based heterostructure, a nitride/oxide multilayer stack or a silicon-on-insulator (SOI) structure, is patterned using e-beam resist (usually PMMA) and subsequent etching. The lower supporting layers are then etched away, leaving a suspended 2D structure (Charlton and Parker, 1998).

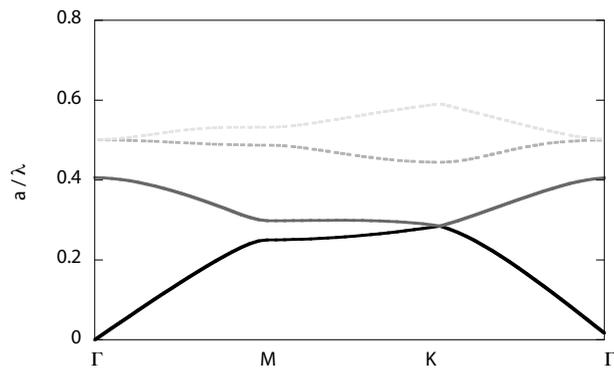
An alternative approach uses e-beam to define the 2D pattern in PMMA and then transfers this into silicon. The silicon is then used as a mold to define the pattern in a soft polymer (PDMS) which is subsequently used again as a mold for transfer into another PMMA layer (Pisignano et al., 2004). This room-temperature molding process has been used to define 160 nm holes with a 550 nm period and has the advantage that, once the high-resolution pattern has been defined, the mold can be re-used to generate the photonic crystal repeatedly.

3D techniques

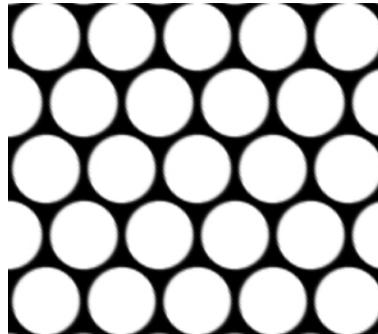
The great aspiration of many microtechnology-oriented nano-opticians is to develop an efficient means for fabrication of 3D photonic crystal structures. Although a wide variety of techniques has been demonstrated, many of these are either slow, expensive, lacking in flexibility or all of the above. Whereas complex three-dimensional structures have been assembled by stacking orthogonal semiconductor stripes using wafer fusion



(a) Energy bands for TE.



(b) Energy bands for TM.



(c) 2D photonic crystal structure for which the band structure was simulated; the black regions are Si, the white regions air.

Figure 14.5 Calculated band structure for a 2D photonic crystal consisting of a hexagonal array of circular air holes in silicon with radius/period = 0.45. The first four bands are shown: *bottom to top* solid black, solid gray, dashed grey, dashed light gray. The abscissa represents the wave vector in different crystalline directions and the ordinate (a/λ) is proportional to the photon frequency or energy. Calculations courtesy of Yaxiu Sun.

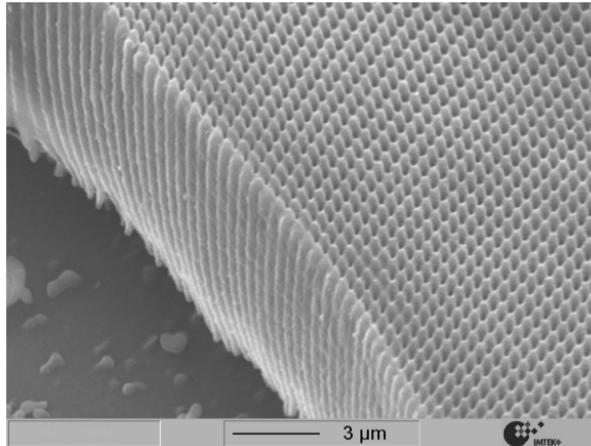


Figure 14.6 SEM photograph of a 2D photonic crystal structure defined in photoresist using three-beam interference lithography.

at 30 nm accuracy (Noda et al., 2000), leading to so-called “woodpile” structures, the most promising techniques involve some form of self-assembly.

Many of these approaches involve stacking of micro- and nanospheres (Ye et al., 2006). As an example, hydrogel microspheres, with diameters of 848 or 964 nm, were suspended in liquid, where these pack closely together due to capillary forces as the liquid dries (Tsuji and Kawaguchi, 2005). The result is a 3D structure with periods of 1 390 and 1 250 nm, respectively. Silica microspheres have also been arranged using holographically-generated optical traps (Roichman and Grier, 2005), such that not only periodic but also patterned structures may be defined, allowing the definition of waveguides and defects. Self-assembly controlled by varying the pH of the surrounding liquids, and taking advantage of forces generated by hydrophilic and hydrophobic surfaces, has been employed to generate a wide variety of complex structures and is also applicable to photonic crystals (Onoe et al., 2007).

Inverse opal structures (the configuration of Figure 14.2, but for which the spheres represent holes in a surrounding material with higher refractive index) have been generated by stacking silica spheres using self-assembly, filling the interstitial space with a liquid precursor, hardening this and dissolving the spheres, leading to voids in the structure. Other approaches use laser direct writing (Deubel et al., 2006) or atomic layer deposition techniques in which the inverse opal is grown layer by layer using 2-photon polymerization, concomitantly allowing the definition of 3D waveguides (Rinne et al., 2008).

Interference lithography, which we encountered in Section 11.3.4, can be expanded to allow the exposure of 3D patterns (Kang et al., 2008). A three-dimensional face-centered cubic (fcc) pattern has been generated with a four-beam configuration using short-pulse emission from a frequency-tripled Nd:YAG laser emitting at 355 nm (Campbell et al., 2000). Photosensitive SU-8 polymer was used as a photoresist, since layers of 10 to 60 μm thickness can be defined and exposed. The approach may also be used to

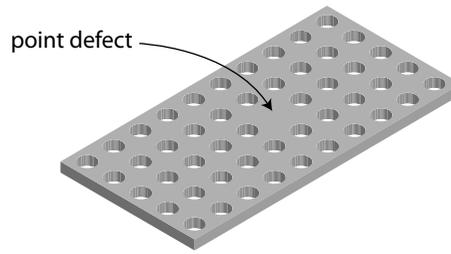


Figure 14.7 Schematic representation of a 2D slab photonic crystal with a defect; a cell is missing from the center row. Such a structure may act as a micro-optical cavity, confining photons to the defect region.

define inverse opals, by filling the exposed and developed voids with TiO_2 and removing the photoresist. An example of a photonic crystal structure defined using interference lithography is shown in Figure 14.6.

14.2.4 Photonic crystal devices

Although a wealth of fundamental scientific work has been performed on and with photonic crystals, the advanced state of fabrication technology has stimulated an equally rich spectrum of applications (Noda, 2006; Benisty et al., 2006). Although some of these employ a homogeneous photonic crystal, most devices based on this technology require the incorporation of defects: the physics of photonic crystals really becomes interesting when a cell, or row of cells, in a periodic structure is modified or removed.

A photonic crystal point defect is shown schematically in Figure 14.7: one cell near the center of the 2D slab structure has been removed. Such a point defect, as we will see in more detail below, may correspond to an optical microcavity, since photons of a certain energy range injected into or generated inside this volume cannot escape through the periodic crystal, due to the photonic bandgap. The controlled incorporation of defects into 2D and 3D photonic crystal structures is one of the challenges in fabrication, but essential for a wide variety of applications.

Waveguides

The generation of a line defect in a 2D slab photonic crystal, as shown in Figure 14.8, results in a waveguide. By removing one or more rows of cells⁵, light can propagate along the line defect, guided on either side by the periodic photonic crystal structure (Krauss, 2003). Photons with energies within the photonic bandgap are (horizontally) captured inside the waveguide, since the bandgap does not allow propagation within the plane of the slab. Vertical confinement is due to total internal reflection at the top and bottom surfaces, as for a classic slab waveguide.

⁵ Such waveguides are frequently designated W_n , where n represents the number of rows removed. Figure 14.8 is thus a W_1 waveguide; removing 2 parallel rows results in a W_2 waveguide, and so forth.

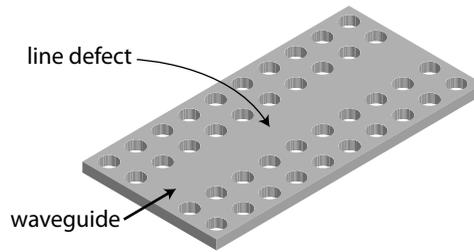


Figure 14.8 Schematic representation of a 2D slab photonic crystal waveguide based on a line defect generated by removing one row of cells. This structure corresponds to a W1 waveguide.

Such photonic crystal waveguides tend to have higher optical losses than their classical counterparts, and losses tend to decrease as the waveguides become wider. Whereas W1 waveguides have losses in the range 100–200 dB/cm, these decrease to about 10–20 dB/cm for W3 and 2 dB/cm for W7 structures (Benisty et al., 2006; Kotlyar et al., 2004). A significant source of optical loss can be out-of-plane scattering, due to inhomogeneities in the structure; since guiding vertically in the 2D slab structure is by TIR, guiding is not as absolute as it is in the plane of the waveguide.

The great advantage of photonic crystal-based waveguides, however, is that they can be subject to very tight curves without these leading to excessive losses. Since the photonic bandgap does not permit lateral propagation outside the waveguide, it may even make a 90° bend without losing any photons laterally. In a 2D GaAs-based photonic crystal slab waveguide operating at $\lambda = 1.55 \mu\text{m}$, as shown in Figure 14.9, 60° bends with a radius of curvature less than $1 \mu\text{m}$ were shown to have 100% transmission through the bend (Chow et al., 2001); compare this behavior with that seen for classical waveguides, as we discussed in Section 9.5.1. As a result, photonic crystal waveguides and waveguide systems can be very compact, since the large radii of curvature and resulting long transition lengths required for curves in classic waveguide structures are reduced to almost zero.

Light emitters

Photonic crystals are also popularly employed for enhancing the emission and out-coupling efficiency of micro-optical light emitters. A single point defect in a 2D slab structure, fabricated in a membrane consisting of a multi-quantum well structure using InGaAsP, was used as a laser cavity, resulting in a defect mode laser (Painter et al., 1999). The defect itself had a volume of 2.5 half-wavelengths (about $0.03 \mu\text{m}^3$) and was optically pumped; lasing at $\lambda \approx 1.5 \mu\text{m}$ was observed, and such a microcavity laser may lead to ultra-low threshold and low-noise optical sources.

Incorporating a photonic crystal, designed so that the modes which propagate in it are optimized for emission of photons, onto a thin LED structure can strongly enhance the photon extraction efficiency (Wierer et al., 2009). For an InGaN/GaN-based LED, emitting at $\lambda = 450 \text{ nm}$, $\eta \approx 73\%$ was demonstrated using this approach. Alternatively, nano-imprint lithography was used to define a photonic crystal pattern into a Cr layer on top of a green GaP LED emitting at $\lambda = 525 \text{ nm}$ (Kim et al., 2007); the 180 nm

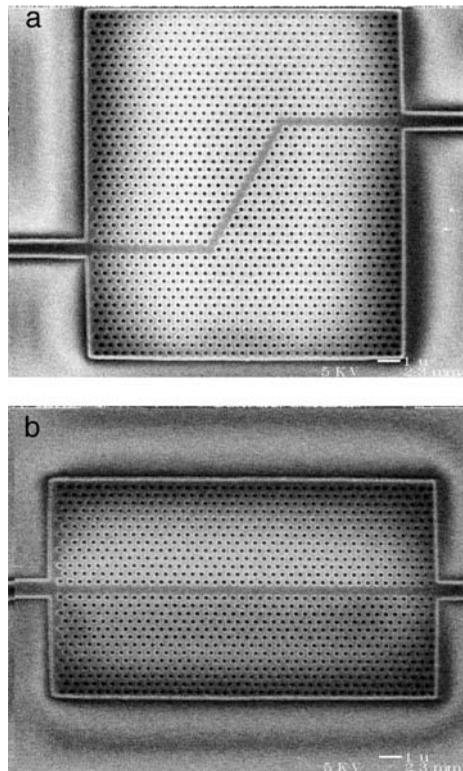


Figure 14.9 Straight *bottom* and bent *top* photonic crystal waveguides, fabricated in GaAs sandwiched between Al_xO_y and SiO_2 . The W1 waveguides were defined by removing one row of cells; the upper waveguide has 60° bends. Photos from (Chow et al., 2001); copyright © 2001 Optical Society of America; reprinted with permission. Photo courtesy of Shawn-Yu Lin, Rensselaer Polytechnic Institute.

diameter holes with a 295 nm period served to increase emission efficiency by a factor nine. The ultimate LED structure is likely one which is fabricated completely from a photonic crystal substrate, such that, since their propagation is not allowed, all photons must escape from the material.

As an alternative application, a photonic crystal structure has been used to maintain single-mode operation of 1.3 μm VCSELs (Leisher et al., 2006). As seen in Figure 14.10, the periodic pattern was etched into the top VCSEL facet using a focussed ion beam, leading to single-mode operation with a side-mode suppression ration (SMSR) of over 37 dB at output powers exceeding 1 mW. The design of the photonic crystal may be adapted to the demands of the laser characteristic, adjusting the hole/period ratio to optimize single-mode emission for different wavelengths.

Other applications

The incorporation of defects or groups of defects into the photonic crystal is one means to realize optical microcavities in two or three dimensions. Through the use of optimized

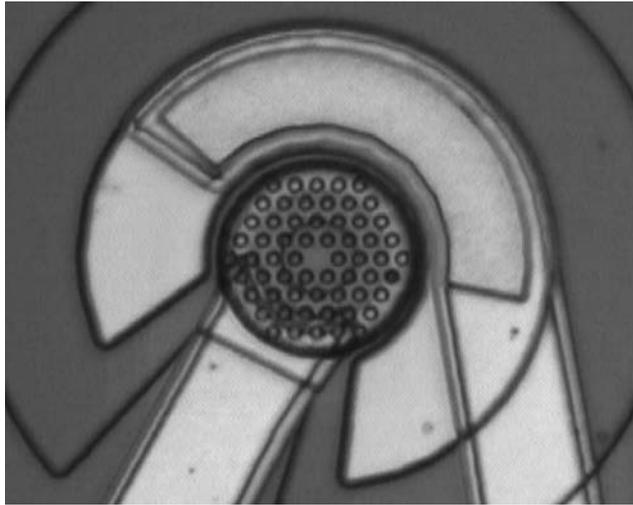


Figure 14.10 Microscope photograph of a photonic crystal (with a central defect) etched onto the surface facet of a VCSEL; the period is $3.7\ \mu\text{m}$ and the cell diameter $1.85\ \mu\text{m}$. Photo from (Leisher et al., 2006); copyright © 2006 IEEE; reprinted with permission.

designs, very high Q values, in the range of those achievable with microdisks, have been demonstrated (Song et al., 2005). Using 2D photonic crystal slab structures, it has been found that the formation of defect cavities with non-abrupt edges, using subtle changes in cell position and size, reduces losses and strongly increase the quality factor. An example of the changes in resonance behavior due to small and precisely-defined shifts of cells around a defect in a photonic crystal are shown in Figure 14.11 (Noda, 2006). Q values of $6 \cdot 10^5$ have been demonstrated and theoretical expectations indicate that values up to $2 \cdot 10^7$ may be possible using this approach.

Photonic crystal waveguides have interesting dispersive properties, making them useful for various means for modulation of light. Optical delay lines have been conceived as has the generation of “slow light”, propagation at low group velocity with low dispersion (Corcoran et al., 2009). A low group velocity allows significantly increased efficiency of non-linear effects, such that green light emission ($\lambda = 520\ \text{nm}$) has been generated in a silicon photonic crystal waveguide using third-harmonic generation. The structure employed was an $80\ \mu\text{m}$ long W1 waveguide fabricated on a $200\ \text{nm}$ thick membrane.

The photonic bandgap suggests that photonic crystals may make excellent mirrors for a range of wavelengths, much like an advanced version of the Bragg mirrors of Section 5.3.5, on which the functionality of DFB lasers and VCSELs rely. Scanning micromirrors using photonic crystal reflectors, for example, have been fabricated using standard silicon technology; the $675\ \text{nm}$ diameter holes with $820\ \text{nm}$ period were etched into SiO_2 on Si (Jung et al., 2009). The advantage of using a photonic crystal in this application is that the resultant mirrors have low polarization sensitivity and small angular dependence of reflectance, which is not the case for standard Bragg mirrors.

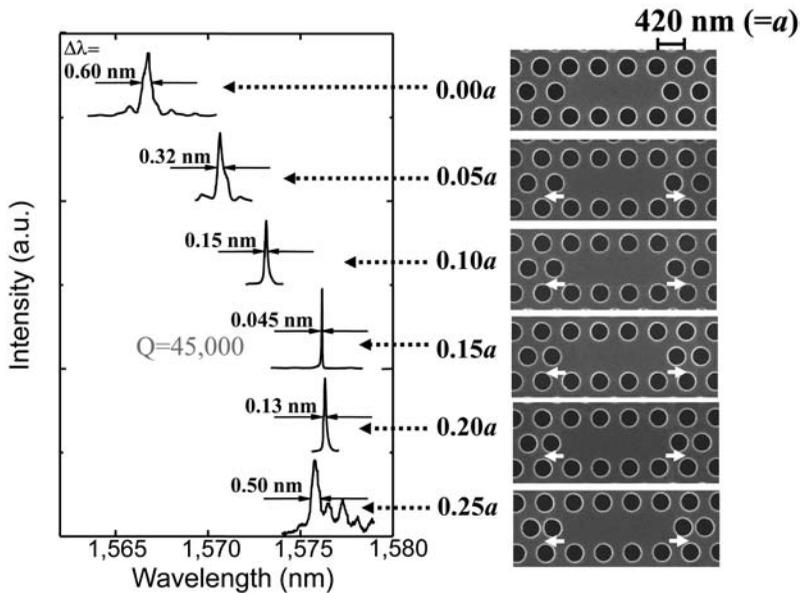


Figure 14.11 Resonance spectra of variably-sized photonic crystal based resonant cavities, formed by very small shifts of the cells bordering the cavity; $a = 420$ nm is the period, so the shifts are in steps of only 21 nm. Note the change in resonant frequency. Figure from (Noda, 2006); copyright © 2006 IEEE; reprinted with permission. Figure courtesy of Susumu Noda, Kyoto University.

Photonic crystal mirrors may be used to generate a waveguide-based Fabry-Perot cavity, into which light is coupled using waveguide proximity couplers (Dotan et al., 2009). The reflectors are generated using a varying hole radius for the photonic crystal (115 to 25 nm), and the cavity itself may be considered to be a large defect in the waveguide structure. This SOI-based configuration functions as an add/drop filter around $\lambda = 1.55$ μm .

14.2.5 Tunable photonic crystals

Since photonic crystals have highly wavelength-dependent characteristics, it becomes immediately interesting to make them tunable. We have already considered some of the possible tuning concepts in earlier chapters: mechanical deformation in Section 12.4.3 and the use of fluidics in Section 13.6.1. One important motivating application for many of the concepts developed is that of a full-color display (Fudouzi and Xia, 2003).

Numerous other approaches have been proposed and demonstrated, many taking advantage of the self-assembly processes used for generation of 3D photonic crystals. Superparamagnetic colloidal nanoclusters, for example, have been suspended in solution and subsequently aligned using a magnetic field (Kim et al., 2009). Chains of various lengths and configurations could be obtained using this approach, and the resultant structures fixed lithographically. These structures have tunable optical properties, such

that multi-color images (at visible wavelengths) can be fabricated on flexible substrates using this technique.

The swelling processes outlined in Section 12.4.1 also play a role in many tuning concepts. For instance, a photonic “gel” fabricated using a block co-polymer which swells in the presence of chloroform (NH_4Cl) has been used to make a widely-tunable photonic crystal, whose absorption peak shifts from 364 to 1 627 nm (Kang et al., 2007).

Alternatively, 3D arrays of silica nanospheres have been self-assembled using a sol-gel process; the interstitial space is then filled with polyferrocenylsilane, a metallopolymer, which swells in solvent, thus increasing the spacing between the spheres and therefore the period of the photonic crystal (Arsenault et al., 2007). The entire assembly is fabricated in an electrochemical cell, representing a compact, self-contained display unit which can generate a wide variety of colors. In related work, a vapor-pressure tunable photonic crystal has been generated using a colloidal polymer gel and metallopolymer spheres; the structure swells in dichloromethane, resulting in reproducible and repeatable tuning (Arsenault et al., 2005).

14.2.6 Photonic crystal fibers

A final variant on the current theme is photonic crystal fiber (PCF). Whereas standard optical fibers, of the type we considered in Section 9.6, use a small refractive index step for radial guiding, the PCF employs a periodic structure in the axial direction to essentially create a photonic bandgap, thereby guiding photons in the azimuthal direction (Russell, 2006).

There are numerous periodic patterns used for guiding in PCF; a nice sampler of some of the different structures implemented is shown in Figure 14.12. Due to the confinement which derives from the periodic structure outside the core, guiding takes place even for $n_g = 1$, since total internal reflection does not play a role in these structures. There is a wide variety of other PCF configurations, including solid core, multiple cores and complex lattice-like cladding layers. The fabrication of these structures is hardly trivial, and is based on drawing stacks of solid and hollow glass capillaries at high temperature.

The optical losses in PCF are still higher than those in state-of-the-art standard fibers, with best values in the range of 0.28 dB/km at $\lambda = 1.55 \mu\text{m}$; hollow-core fiber has values closer to 1.2 dB/km. PCF does have a number of essential features which make it an attractive light guide for numerous applications, however: single-mode operation is possible over a wider wavelength range; the damage threshold is high; and bend losses are low.

In addition, the high electric field concentrations possible in PCF imply that the efficiency of non-linear effects is higher than in solid-body fiber. As a result, PCF is often used for the generation of supercontinuum spectra which, for visible wavelengths, can span the range 450 to 800 nm with a spectral power density of 4.5 mW/nm. Finally, hollow-core PCF structures are attractive for novel types of sensors, particularly since gas flow can be directed into the fiber and thus strongly overlap the optical field; Raman gas cells have been made using this approach.

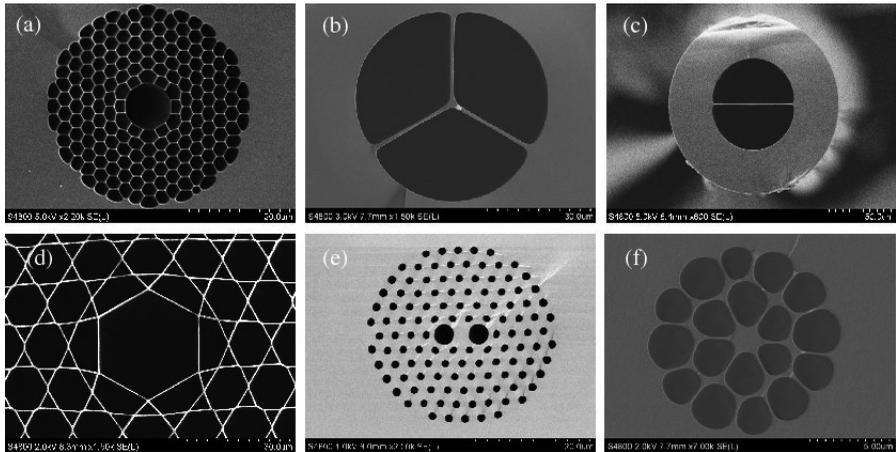


Figure 14.12 Several examples of photonic crystal fiber structures: (a) hollow core photonic band gap PCF (core diameter 10 μm); (b) “Mercedes” fibre (core diameter 2 μm); (c) nanoweb (≈ 500 nm wide) fibre; (d) kagomé-lattice hollow core PCF (core diameter 28 μm); (e) solid core birefringent PCF (core diameter 3 μm); (f) highly nonlinear solid core PCF made from a lead silicate glass (core diameter 1.5 μm). Photos courtesy of Philip Russell, Max Planck Institute for the Science of Light; copyright © Max Planck Institute for the Science of Light.

14.3 Plasmonics

The quantized collective excitation of electrons, or quantized plasma oscillations, in a material is termed a plasmon, a quasiparticle well known to students of advanced solid state physics. Within the past decade, it has been found that plasmons may have interesting interactions with photons, leading to a rich variety of phenomena and a number of useful devices derived from these. The resultant technology is termed plasmonics, and represents one of the most active research areas at the interface of nanophysics and optics.

The observation of “abnormal” transmission of light through an array of sub-wavelength sized holes was one of the original experiments which led to the current developments in plasmonics (Ebbesen et al., 1998). It was observed that the transmission efficiency (normalized relative to the open aperture area) through 150 nm holes in a 200 nm thick Ag film was greater than unity; this effect was eventually determined to be due to the coupling of surface plasmons to the optical field, thereby enhancing the transmission beyond that which would be classically expected.

14.3.1 Plasmons

The use of plasmons in optics is not new. Surface plasmon resonance sensors, which we mentioned in Sections 9.7.2 and 13.1, have a long history and represent successful laboratory instruments due to their high sensitivity. What is new in plasmonics is that nanoscale structures are employed to generate new physical effects and, at the same time, suggest a means to strongly miniaturize plasmonic systems.

As mentioned above, a plasmon⁶ is a collective excitation of electrons, and is typically found at the surface of metals. It consists of charges which interact with an electromagnetic field and may be highly localized or propagate along the surface (Yin et al., 2005), and may be imaged using fluorescence techniques (Ditlbacher et al., 2002). The electromagnetic field is TM polarized and is tightly concentrated at the surface, decaying exponentially in the surface normal direction. Plasmons may be generated on the surface of metallic films, typically insulator/metal/insulator sandwich structures, or on nanoparticles in a variety of shapes. The material requirement to support a plasmon is that the real dielectric constants of the metal and insulator are related simply as

$$\epsilon_{metal} = -2\epsilon_{insulator} \quad (14.17)$$

such that the dielectric constants must be of opposite sign.

The plasmon resonance frequency is a function of ϵ_{metal} and $\epsilon_{insulator}$ and, for nanoparticles, is strongly shape-dependent. Nevertheless, there is a considerable variety of nanostructures which supports plasmons: Ag nanospheres, with radii of 400 to 600 nm; Au nanospheres, with radii of 600 to 800 nm; and nanorods with lengths in the range 500 to 2 000 nm have all been shown to be suitable (Lal et al., 2007). We refer the reader to review articles in the literature for a more detailed look at the physics of surface plasmons (Barnes et al., 2003; Maier and Atwater, 2005; Barnes, 2006).

14.3.2 Plasmonic waveguides

Since surface plasmons can propagate along a metallic surface, extremely compact plasmonic waveguide structures can be fabricated to guide an optical field along the surface. Plasmonic waveguides can have dimensions much less than λ but provide high confinement, particularly in the vertical direction. As a result, these devices are considered promising candidates for highly compact optical data transmission systems.

Plasmonic waveguide structures are typically thin metal stripes (thicknesses of about 70 nm are common), generally using Ag, Au, Cu or Al, deposited on an insulating substrate such as silica. The propagation losses tend to be high, due predominantly to absorption in the metal itself. Loss is a strong function of waveguide width, which is typically in the range $50 \text{ nm} \leq W \leq 5 \text{ }\mu\text{m}$; narrower waveguides have significantly higher loss. As a result, 200 nm wide waveguides have been shown to have propagation lengths of several micrometers whereas micrometer-wide structures can lead to propagation over centimeter length scales (Zia et al., 2006); losses are lowest in silver waveguides.

Coupling light in and out of a plasmonic waveguide requires that the light be provided with non-zero momentum in the waveguide direction. As a result, coupling is generally achieved using prisms, scattering from surface features, such as nanoparticle arrays, or periodic corrugations on the waveguide, such as grating couplers. Using a curved grating structure and a tapered waveguide, a plasmonic lens has been demonstrated

⁶ Strictly, a surface plasmon polariton.

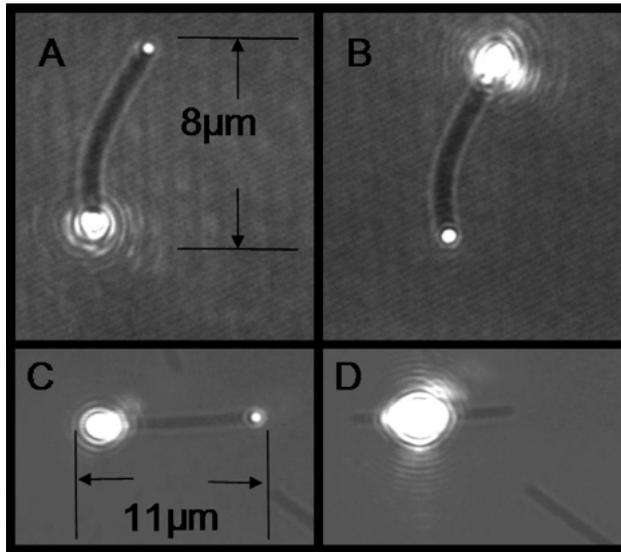


Figure 14.13 A silver nanowire with a diameter of about 100 nm showing optical waveguiding at $\lambda = 830$ nm. Excitation is from the bottom *A*; the top *B*; and the left *C*. Excitation at the center *D* does not lead to a guided mode since plasmons are not excited in this configuration. Photos from (Sanders et al., 2006); copyright © 2006 American Chemical Society; reprinted with permission. Photo courtesy of Mark Reed, Yale University.

which focusses externally coupled light into a waveguide, achieving a focal spot size of less than 100 nm (Vedantam et al., 2009).

Nanowires

Optical guided-wave propagation using plasmons is also possible using nanowires. These metallic structures, typically fabricated in Au or Ag, are shaped as classical wires (or fibers), with cross-sections on the order of 50×200 nm and lengths up to 250 μm , for use at wavelengths in the range of $\lambda = 1.55$ μm (Oulton et al., 2008). Alternative technologies use etched V-shaped grooves (Ozbay, 2006) or custom-grown single-crystal semiconductor whiskers (Yan et al., 2009) to realize nanowires.

Optical losses in nanowires tend to be lower than those for photonic crystal waveguides, with typical values of 2.4 dB/cm for 500 nm diameter structures (Bogaerts et al., 2004). In addition, due to the tight confinement, nanowires may have very tight curves without exhibiting the concomitant excessive loss as seen in classical waveguides or optical fibers (Barrelet et al., 2004), but not necessarily the 90° bends possible with photonic crystals. Nevertheless, silver nanowires, with a diameter of 100 nm and lengths ranging from 3 to 20 μm have been shown to allow wire bends with radii of curvature down to 4 μm (Sanders et al., 2006). Examples showing optical transmission through such Ag nanowires are shown in Figure 14.13, in which the incident and transmitted optical fields are clearly seen. Patterned nanowires have also been used to realize couplers, splitters and waveguide interferometers (Bozhevolnyi et al., 2006).

Nanowire structures may be assembled into more complex systems using microcontact printing, electric fields or fluidic assembly. It is expected that these concepts will prove to have useful applications for lasers, non-linear optics, sensors, and solar cells (Yan et al., 2009).

14.3.3 Plasmonic light sources

Plasmonics is playing an increasingly relevant role in the improvement and development of traditional and novel light sources. Nanostructures rely on plasmonic effects primarily to enhance the efficiency of light emission.

The incorporation of nanostructured Ag films into the quantum wells of LEDs results in an enhancement of the density of states and thus the emission, due to a coupling of the optical field with surface plasmons (Okamoto et al., 2005); an emission enhancement of a factor 32 was observed. OLEDs have benefitted from a nano-patterned dielectric layer over the cathode (Liu et al., 2005) and VCSELs have used nano-apertures to enhance the near-field due to plasmons, where these have been employed for near-field microscopy (Hashizume and Koyama, 2004).

Stimulated emission of plasmons has also been achieved, leading to a plasmon-based laser, or spaser (Noginov et al., 2009). Nanospheres consisting of a dye-doped silica shell surrounding a Au core, with a diameter of 44 nm, were optically pumped at $\lambda = 488$ nm; optical feedback was generated by an oscillating surface plasmon mode, until the absorption losses in the metal were compensated to allow lasing at $\lambda = 531$ nm. This structure corresponds to what is currently the smallest nano-laser structure successfully demonstrated.

14.3.4 Plasmonic sensors

As we mentioned at the outset of this section, surface plasmon resonance sensors are well established and have been used in a laboratory environment for quite some time (Homola et al., 1999); these are generally based on thin metallic films deposited on glass prisms. Advances in nanotechnology have led to advanced forms of plasmonics for sensors, taking advantage of the concentration of optical fields possible with nanoparticles.

The absorption peak of surface plasmons generated on nanoparticles shifts with a change in the refractive index of the surrounding medium, typically the analyte of interest. The effect is an unspecific one: refractive index changes can be due to many factors. Nevertheless, nanoparticles on a surface lead to two effects which enhance the sensitivity of these configurations: enhanced electric field due to the sharp, concentrated metallic surfaces and excitation of localized surface plasmons, which amplify fluorescence and second harmonic generation (Lal et al., 2007).

The increased concentration of an optical field due to nanoparticles, which essentially act as nano-antennas, also amplifies the scattering of this field due to single molecules; the resultant Raman scattering has a spectrum specific to a molecular species and surface-enhanced Raman spectroscopy (SERS) takes advantage of the subsequent high sensitivities to molecular concentrations achievable with nanostructured surfaces

(Kneipp et al., 1997). In SERS, Ag nanoparticles enhance the Raman shift by factors from 10^6 to 10^{14} over conventional configurations, and a variety of nanoparticle shapes has been determined to be effective for this application (Drachev et al., 2005). Most suitable are generally dimers, nanoparticle pairs separated by a precise gap. Nanoshells and extremely sharp tips have also been employed (Pettinger, 2006).

A related use of plasmon-based field enhancement on surfaces is for the improvement of solar cell efficiency. Gold nanoparticles deposited on the solar cell have been shown to be effective in orienting the light along the surface, strongly increasing absorption at longer (600 to 800 nm) wavelengths (Ferry et al., 2009).

14.4 Metamaterials

Metamaterials are structured media not found in nature, often with interesting customizable propagation characteristics. Generally not homogeneous, metamaterials usually consist of an assembly of sub-wavelength structures, such that they appear macroscopically uniform, on the length scale of the wavelength, but with behavior which results from the interaction of electromagnetic fields with the microscopic structures.

One attractive optical property, also not found in nature, which can be realized using metamaterials is negative index of refraction. We recall that in the early pages of this text (Sections 2.3.2 and 3.1), we defined the refractive index as

$$n = \sqrt{\epsilon\mu}, \quad (14.18)$$

where the dimensionless constants ϵ and μ define the permittivity and permeability of the medium, respectively. Although n was seen to be, in general, a complex value, its magnitude was always tacitly assumed to be positive. If a material exhibits $\epsilon < 0$ and $\mu < 0$, however, it has been shown that $n < 0$ can result, since we take the negative square root (Veselago, 1968). Such negative-index materials can have very attractive optical properties (Pendry, 2000).

Negative index materials may be realized using electrically and magnetically resonant structures and much initial work in this area was performed at microwave and THz frequencies; only recently have the capabilities of nanotechnology allowed the demonstration of negative refraction at optical frequencies (Enkrich et al., 2005; Soukoulis et al., 2007). Negative index structures require electrical and magnetic resonators with resonant frequencies corresponding to the fields of interest; these resonators generate the negative permittivity and permeability. The most successful of these have been split ring resonators, essentially nanometer-sized LC circuits consisting of a capacitor coupled with an inductor, where the small sizes are required to achieve resonance at optical frequencies (Enkrich et al., 2005). An example of such a split ring array, with which negative permeability at NIR and visible wavelengths was achieved, is shown in Figure 14.14.

The required nanostructures to yield negative ϵ and μ can be relatively complex and the effects are highly wavelength dependent (Shalaev, 2007); optical losses are high and most structures are planar. One example is an array of 565×265 nm cells of ten pairs

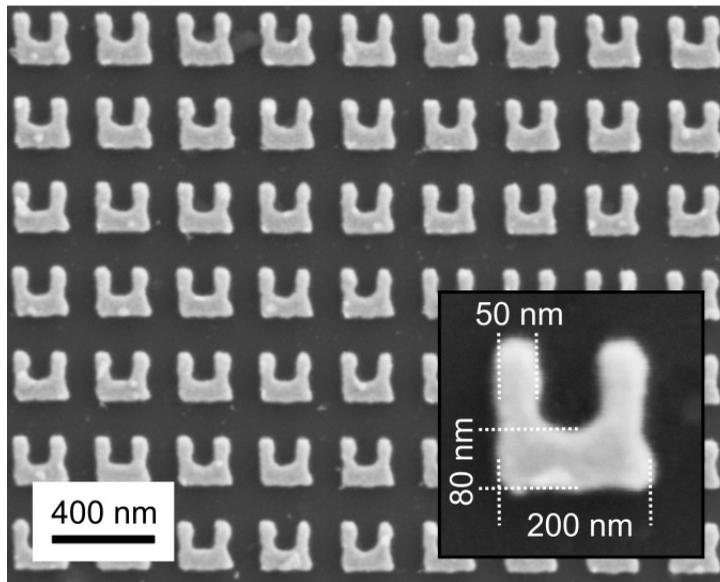


Figure 14.14 SEM photograph of a split ring array which shows negative permeability at optical wavelengths. Photo from (Enkrich et al., 2005); copyright © 2005 by The American Physical Society; reprinted with permission. Photo courtesy of Martin Wegener, Karlsruhe Institute of Technology.

of 30 nm Ag and 50 nm MgF₂ layers, which show $n = 0.63$ for $\lambda = 1\,200$ nm, crossing zero at $\lambda = 1\,475$ nm and reaching $n = -1.23$ for $\lambda = 1\,775$ nm (Dolling et al., 2007). Recent work has demonstrated a 3D metamaterial with a negative refractive index using a “fishnet” structure (Valentine et al., 2008).

Although the optical metamaterials field (sometimes referred to as “transformation optics”), particularly that aspect concerned with anomalous refraction, is still in its infancy, a range of novel applications has been envisaged. As proposed in (Pendry, 2000), “superlenses” based on negative index materials also focus the evanescent field and may result in a perfect focus, beyond the diffraction limit; optical “black holes” perfectly absorb electromagnetic radiation (Genov et al., 2009); and “invisibility cloaks”, which bend light around an object such that it is not visible to an observer, have been demonstrated at near-IR wavelengths (Gabrielli et al., 2009). Aspiring Harry Potters and Hermione Grangers had therefore best study micro-optical engineering to keep abreast of these rapidly evolving developments.

References

- Abbate, G., U. Bernini, E. Ragozzino, and F. Somma (1978). The temperature dependence of the refractive index of water. *Journal of Physics D* 11, 1167–1172.
- Abe, H. (2004). Device technologies for high quality and smaller pixel in CCD and CMOS image sensors. *Technical Digest of the 2004 IEEE International Electron Devices Meeting*, 989–992.
- Adachi, S. (1985). GaAs, AlAs, and $\text{Al}_x\text{Ga}_{1-x}\text{As}$: Material parameters for use in research and device applications. *Journal of Applied Physics* 58, R1–R29.
- Agarwal, M., R. Gunasekaran, P. Coane, and K. Varahramyan (2004). Polymer-based variable focal length microlens system. *Journal of Micromechanics and Microengineering* 14, 1665–1673.
- Agrawal, G. and N. Dutta (1986). *Long-Wavelength Semiconductor Lasers*. New York: Van Nostrand Reinhold.
- Airy, G. (1835). On the diffraction of an object-glass with circular aperture. *Transactions of the Cambridge Philosophical Society* 5, 283–291.
- Akashi, T., S. Higashiyama, H. Takemori, and T. Koizumi (2004). A silicon optical bench incorporating a tantalum-nitride thin-film resistor. *Journal of Micromechanics and Microengineering* 14, 283–289.
- Alferness, R. (1981). Guided-wave devices for optical communication. *IEEE Journal of Quantum Electronics* QE-17(6), 946–959.
- Aljaseem, K., A. Werber, A. Seifert, and H. Zappe (2008). Fiber optic tunable probe for endoscopic optical coherence tomography. *Journal of Optics A: Pure and Applied Optics* 10, 044012.
- Almeida, V., C. Barrios, R. Panepucci, and M. Lipson (2004). All-optical switching on a silicon chip. *Optics Letters* 29(24), 2867–2869.
- Analui, B., D. Guckenberger, D. Kucharski, and A. Narasimha (2006). A fully integrated 20-Gb/s optoelectronic transceiver implemented in a standard 0.13- μm CMOS SOI technology. *IEEE Journal of Solid-State Circuits* 41(12), 2945–2955.
- Anderson, E., D. Olynick, B. Harteneck, E. Veklerov, G. Denbeaux, W. Chao, A. Lucero, L. Johnson, and D. Attwood (2000). Nanofabrication and diffractive optics for high-resolution x-ray applications. *Journal of Vacuum Science & Technology B* 18(6), 2970–2975.
- Armani, A., D. Armani, B. Min, K. Vahala, and S. Spillane (2005). Ultra-high-Q microcavity operation in H_2O and D_2O . *Applied Physics Letters* 87, 151118.

- Arsenault, A., V. Kitaev, I. Manners, G. Ozin, A. Mihi, and H. Míguez (2005). Vapor swellable colloidal photonic crystals with pressure tunability. *Journal of Materials Chemistry* 15, 133–138.
- Arsenault, A., D. Puzzo, I. Manners, and G. Ozin (2007). Photonic-crystal full-colour displays. *Nature Photonics* 1(8), 468–472.
- Aschwenden, M., M. Beck, and A. Stemmer (2007). Diffractive transmission grating tuned by dielectric elastomer actuator. *IEEE Photonics Technology Letters* 19(14), 1090 – 1092.
- Aschwenden, M. and A. Stemmer (2006). Polymeric, electrically tunable diffraction grating based on artificial muscles. *Optics Letters* 31(17), 2610–2612.
- Asfour, J.-M. and A. Poleshchuk (2006). Asphere testing with a Fizeau interferometer based on a combined computer-generated hologram. *Journal of the Optical Society of America A* 23(1), 172–178.
- Astier, I. (2007). *Cuisine Inspirée*. Paris: Agnès Viénot Éditions.
- Auch, W. (1994). Progress in fiber optic gyro development and application. In R. DePaula (Ed.), *Fiber Optic and Laser Sensors XI*, Volume 2070 of *Proceedings of SPIE*, pp. 104–112.
- Austin, M. D., H. Ge, W. Wu, M. Li, Z. Yu, D. Wasserman, S. A. Lyon, and S. Y. Chou (2004). Fabrication of 5 nm linewidth and 14 nm pitch features by nanoimprint lithography. *Applied Physics Letters* 84(26), 5299–5301.
- Awano, H. and N. Ohta (1998). Magneto-optical recording technology toward 100 Gb/in². *IEEE Journal of Selected Topics in Quantum Electronics* 4(5), 815–820.
- Bach, H. (Ed.) (1995). *Low Thermal Expansion Glass Ceramics*. Berlin: Springer-Verlag.
- Bach, H. and N. Neuroth (Eds.) (1995). *The Properties of Optical Glass*. Springer-Verlag.
- Bai, J. and G. Chen (2002). Continuous-wave diode-laser end-pumped Nd:YVO₄/KTP high-power solid-state green laser. *Optics and Laser Technology* 34, 333–336.
- Balslev, S. and A. Kristensen (2005). Microfluidic single-mode laser using high-order Bragg grating and antiguiding segments. *Optics Express* 13(1), 344–351.
- Barbour, N. and G. Schmidt (2001). Inertial sensor technology trends. *IEEE Sensors Journal* 1(4), 332–339.
- Barnes, W. (2006). Surface plasmon–polariton length scales: a route to sub-wavelength optics. *Journal of Optics A* 8, S87–S93.
- Barnes, W. L., A. Dereux, and T. Ebbesen (2003, Aug). Surface plasmon subwavelength optics. *Nature* 424, 824–830.
- Barrekette, E. and R. Christensen (1965). On plane blazed gratings. *IBM Journal of Research and Development March*, 108–117.
- Barrelet, C., A. Greytak, and C. Lieber (2004). Nanowire photonic circuit elements. *Nano Letters* 4(10), 1981–1985.
- Barton, J., E. Skogen, M. Mašanović, S. DenBaars, and L. A. Coldren (2003). A widely tunable high-speed transmitter using an integrated SGDBR laser-semiconductor optical amplifier and Mach-Zehnder modulator. *IEEE Journal of Selected Topics in Quantum Electronics* 9(5), 1113–1117.

- Bass, M. and E. van Stryland (Eds.) (2001). *Fiber Optics Handbook*, New York. McGraw Hill.
- Bastard, G. (1988). *Wave Mechanics Applied to Semiconductor Heterostructures*. Les Ulis: Les Editions de Physique.
- Baur, J., P. Schlotter, and J. Schneider (1998). White light emitting diodes. In R. Helbig (Ed.), *Advances in Solid State Physics*, Volume 37, Berlin, pp. 67–78. Springer-Verlag.
- Bean, K. (1978). Anisotropic etching of silicon. *IEEE Transactions on Electron Devices ED-25*, 1185–1192.
- Behrmann, G. and J. Mait (1998). Hybrid (refractive/diffractive) optics. In H. Herzig (Ed.), *Micro-Optics*, Chapter 10, pp. 259–292. London: Taylor & Francis.
- Benisty, H., J.-M. Lourtioz, A. Chelnokov, S. Combrié, and X. Checoury (2006). Recent advances toward optical devices in semiconductor-based photonic crystals. *Proceedings of the IEEE 94(5)*, 997–1023.
- Bennett, B. and R. Soref (1987). Electrorefraction and electroabsorption in InP, GaAs, GaSb, InAs, and InSb. *IEEE Journal of Quantum Electronics QE-23(12)*, 2159–2166.
- Beranek, M., E. Chan, C. Chen, K. Davido, H. Hager, C.-S. Hong, D. Koshinz, M. Rassaian, H. Soares, R. St Pierre, P. Anthony, M. Cappuzzo, J. Gates, L. Gomez, G. Henein, J. Shmulovich, M. Occhionero, and K. Fennessy (2000). Passive alignment optical subassemblies for military/aerospace fiber-optic transmitter/ receiver modules. *IEEE Transactions on Advanced Packaging 23(3)*, 461–469.
- Berge, B. (1993). Électrocapillarité et mouillage de films isolants par l'eau. *Comptes rendus de l'Académie des sciences 317(2)*, 157–163.
- Berge, B. and J. Peseux (2000). Variable focal lens controlled by an external voltage: an application of electrowetting. *The European Physical Journal E 3*, 159–163.
- Biebuyck, H. and G. Whitesides (1994). Self-organization of organic liquids on patterned self-assembled monolayers of alkanethiolates on gold. *Langmuir 10*, 2790–2793.
- Bifano, T., R. Mali, J. Dorton, J. Perreault, N. Vandelli, M. Horenstein, and D. Castañon (1997). Continuous-membrane surface-micromachined silicon deformable mirror. *Optical Engineering 36(5)*, 1354–1360.
- Binh, L. (2006). Lithium niobate optical modulators: Devices and applications. *Journal of Crystal Growth 288(1)*, 180–187.
- Birnstock, J., J. Blässing, A. Hunze, M. Scheffel, M. Stöbel, K. Heuser, G. Wittmann, J. Wörle, and A. Winnacker (2001). Screen-printed passive matrix displays based on light-emitting polymers. *Applied Physics Letters 78(24)*, 3905–3907.
- Bishop, D., C. Giles, and G. Austin (2002, March). The Lucent LambdaRouter: MEMS Technology of the Future Here Today. *IEEE Communications Magazine*, 75–79.
- Blakemore, J. (1982). Semiconducting and other major properties of Gallium Arsenide. *Journal of Applied Physics 53*, R123–R181.
- Blanc, D., S. Pellissier, K. Saravanamuttu, S. Najafi, and M. Andrews (1999). Self-processing of surface-relief gratings in photosensitive hybrid sol-gel glasses. *Advanced Materials 11(18)*, 1508–1511.

- Bogaerts, W., D. Taillaert, B. Luyssaert, P. Dumon, J. van Campenhout, P. Bienstman, D. van Thourhout, R. Baets, V. Wiaux, and S. Beckx (2004). Basic structures for photonic integrated circuits in silicon-on-insulator. *Optics Express* 12(8), 1583–1591.
- Boisset, G., B. Robertson, and H. Hinton (1995). Design and construction of an active alignment demonstrator for a free-space optical interconnect. *IEEE Photonics Technology Letters* 7(6), 676–679.
- Bomse, D., A. Stanton, and J. Silver (1992). Frequency modulation and wavelength modulation spectroscopies: comparison of experimental methods using a lead-salt diode laser. *Applied Optics* 31(6), 718–731.
- Bönsch, P., D. Wüllner, T. Schimpf, and A. Schlachetzki (1998). Ultrasoother V-grooves in InP by two-step wet chemical etching. *Journal of the Electrochemical Society* 145(4), 1273–1276.
- Born, M. and E. Wolf (1980). *Principles of Optics* (6th ed.). Oxford: Pergamon Press.
- Borra, E. (1982). The liquid-mirror telescope as a viable astronomical tool. *Journal of the Royal Astronomical Society of Canada* 76, 245–256.
- Borrelli, N. (2005). *Microoptics Technology* (2nd ed.). New York: Marcel Dekker.
- Borrelli, N., M. Cotter, and J. Luong (1986). Photochemical method to produce waveguiding in glass. *IEEE Journal of Quantum Electronics* 22(6), 896–901.
- Bourouina, T., H. Fujita, G. Reyne, and M. Motamedi (2005). Optical scanning. In M. Motamedi (Ed.), *MOEMS*, Bellingham, WA, pp. 323–368. SPIE.
- Bouwhuis, G., J. Braat, A. Hijser, J. Pasman, G. van Rosmalen, and K. Schouhamer Immik (1985). *Principles of optical disc systems*. Bristol: Adam Hilger Ltd.
- Bowling, D., S. Sargent, B. Tanner, and J. Ehleringer (2003). Tunable diode laser absorption spectroscopy for stable isotope studies of ecosystem–atmosphere CO₂ exchange. *Agricultural and Forest Meteorology* 118, 1–19.
- Bozhevolnyi, S., V. Volkov, E. Devaux, J.-Y. Laluet, and J.-Y. Ebbesen (2006). Channel plasmon subwavelength waveguide components including interferometers and ring resonators. *Nature* 440, 508–511.
- Brand, O. (2007). Packaging. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 1, Chapter 1.15, pp. 431–463. Amsterdam: Elsevier.
- Brandenburg, A. (1997). Differential refractometry by an integrated-optical Young interferometer. *Sensors & Actuators B* 38–39, 266–271.
- Brennan, K. (1999). *The Physics of Semiconductors*. Cambridge: Cambridge University Press.
- Brenner, K.-H., M. Kufner, S. Kufner, J. Moisel, A. Müller, S. Sinzinger, M. Testorf, J. Götttert, and J. Mohr (1993). Application of three-dimensional micro-optical components formed by lithography, electroforming, and plastic molding. *Applied Optics* 32(32), 6464–6469.
- Brenner, K.-H. and W. Singer (1993). Light propagation through microlenses: a new simulation method. *Applied Optics* 32(26), 4984–4988.
- Brewster, D. (1831). *A Treatise on Optics* (facsimile by Adamant Media Corporation 2001 ed.). London: John Taylor.
- Brinker, C. and G. Scherer (1990). *Sol-Gel Science*. New York: Academic Press.

- Brooker, G. (2003). *Modern Classical Optics*. Oxford: Oxford University Press.
- Buck, J. (2004). *Fundamentals of Optical Fibers*. New York: Wiley.
- Buralli, D. and G. Morris (1991). Design of diffractive singlets for monochromatic imaging. *Applied Optics* 30(16), 2151–2158.
- Burch, J. and J. Tokarski (1968). Production of multiple beam fringes from photographic scatterers. *Journal of Modern Optics* 15(2), 101–111.
- Burroughes, J., D. Bradley, A. Brown, R. Marks, K. Mackay, R. Friend, P. Burns, and A. Holmes (1990). Light-emitting diodes based on conjugated polymers. *Nature* 347, 539–541.
- Burrus, C. and E. Ulmer (1971, August). Efficient small-area GaAs – Ga_{1-x}Al_xAs heterostructure electroluminescent diodes coupled to optical fibers. *Proceedings of the IEEE*, 1263–1264.
- Buus, J. (1982). The effective index method and its application to semiconductor lasers. *IEEE Journal of Quantum Electronics QE-18*(7), 1083–1089.
- Campbell, M., D. Sharp, M. Harrison, R. Denning, and A. Turberfield (2000). Fabrication of photonic crystals for the visible spectrum by holographic lithography. *Nature* 404, 53–56.
- Campbell, S. (2001). *The Science and Engineering of Microelectronic Fabrication* (2nd ed.). New York: Oxford University Press.
- Canham, L. (1990). Silicon quantum wire array fabrication by electrochemical and chemical dissolution of wafers. *Applied Physics Letters* 57(10), 1046–1048.
- Carnal, O., M. Sigel, T. Sleator, H. Takuma, and J. Mlynek (1991). Imaging and focusing of atoms by a Fresnel zone plate. *Physical Review Letters* 67(23), 3231–3235.
- Carroll, J., J. Whiteaway, and D. Plumb (1998). *Distributed Feedback Semiconductor Lasers*. New Jersey, USA: IEEE.
- Cassidy, D. and J. Reid (1982). Atmospheric pressure monitoring of trace gases using tunable diode lasers. *Applied Optics* 21(7), 1185–1190.
- Cerullo, G. and S. D. Silvestri (2003). Ultrafast optical parametric amplifiers. *Review of Scientific Instruments* 74(1), 1–18.
- Chandrasekhar, S. and N. Madhusudana (1980). Liquid crystals. *Annual Review of Materials Science* 10, 133–155.
- Chang, Y., K. Mohseni, and V. Bright (2007). Fabrication of tapered SU-8 structure and effect of sidewall angle for a variable focus microlens using EWOD. *Sensors & Actuators: A* 136, 546–553.
- Chang-Yen, D. and B. Gale (2003). An integrated optical oxygen sensor fabricated using rapid-prototyping techniques. *Lab on a Chip* 3, 297–301.
- Charlton, M. and G. Parker (1998, Jun). Nanofabrication of advanced waveguide structures incorporating a visible photonic band gap. *Journal of Micromechanics and Microengineering* 8, 172–176.
- Chen, C. and F. Tseng (2005). Tunable micro-aspherical lens manipulated by 2D electrostatic forces. *Digest of Technical Papers: The 13th International Conference on Solid-State Sensors, Actuators and Microsystem (Transducers 2005)*, 376–379.
- Chen, R. (1993). Polymer-based photonic integrated circuits. *Optics and Laser Technology* 25(6), 347–365.

- Chen, Y., Z. Li, M. Henry, and A. Scherer (2009). Optofluidic circular grating distributed feedback dye laser. *Applied Physics Letters* 95, 031109.
- Cheng, C., C. Chang, C. Liu, and J. Yeh (2006). A tunable liquid-crystal microlens with hybrid alignment. *Journal of Optics A* 8, S365–S369.
- Cheng, Y., K. Sugioka, and K. Midorikawa (2004). Microfluidic laser embedded in glass by three-dimensional femtosecond laser microprocessing. *Optics Letters* 29(17), 2007–2009.
- Cho, K., E. Lee, W. Joo, E. Jang, T. Kim, S. Lee, S.-J. Kwon, J. Han, B.-K. Kim, B. Choi, and J. Kim (2009). High-performance crosslinked colloidal quantum-dot light-emitting diodes. *Nature Photonics* 3, 341–345.
- Choquette, K., M. Hong, R. Freund, J. Mannaerts, R. Wetzel, and R. Leibenguth (1993). Vertical-cavity surface-emitting laser diodes fabricated by in situ dry etching and molecular beam epitaxial regrowth. *IEEE Photonics Technology Letters* 5(3), 284–287.
- Chou, S., P. Krauss, and P. Renstrom (1996). Imprint lithography with 25-nanometer resolution. *Science* 272, 85–87.
- Chou, S., P. Krauss, W. Zhang, L. Guo, and L. Zhuang (1997). Sub-10 nm imprint lithography and applications. *Journal of Vacuum Science and Technology B* 15(6), 2897–2904.
- Chow, E., S. Lin, J. Wendt, S. Johnson, and J. Joannopoulos (2001). Quantitative analysis of bending efficiency in photonic-crystal waveguide bends at $\lambda = 1.55 \mu\text{m}$ wavelengths. *Optics Letters* 26(5), 286–288.
- Chow, W. and S. Koch (1999). *Semiconductor-Laser Fundamentals*. Berlin: Springer-Verlag.
- Chronis, N., G. L. Liu, K.-H. Jeong, and L. P. Lee (2003). Tunable liquid-filled microlens array integrated with microfluidic network. *Optics Express* 11(19), 2370–2378.
- Chu, T., H. Yamada, S. Ishida, and Y. Arakawa (2005). Compact $1 \times N$ thermo-optic switches based on silicon photonic wire waveguides. *Optics Express* 13(25), 10109–10114.
- Chuang, H.-C., R. Jiménez-Martínez, S. Braun, D. Anderson, and V. Bright (2008). Tunable external cavity diode laser using a micromachined silicon flexure and a volum holographic reflection grating for applications in atomic optics. *Journal of Micro/Nanolithography, MEMS and MOEMS* 7(2), 021010.
- Chung, I., J. Mørk, P. Gilet, and A. Chelnokov (2008). Subwavelength grating-mirror VCSEL with a thin oxide gap. *IEEE Photonics Technology Letters* 20(2), 105–107.
- Ciddor, P. (1996). Refractive index of air: new equations for the visible and near infrared. *Applied Optics* 35(9), 1566–1573.
- Coe-Sullivan, S. (2009). Quantum dot developments. *Nature Photonics* 3, 315–316.
- Cohen, M. and J. Chelikovsky (1988). *Electronic Structure and Optical Properties of Semiconductors*. Berlin: Springer-Verlag.
- Coldren, L. (2000). Monolithic tunable diode lasers. *IEEE Journal of Selected Topics in Quantum Electronics* 6(6), 988–999.

- Coldren, L., G. Fish, Y. Akulova, J. Barton, L. Johansson, and C. Coldren (2009). Tunable semiconductor lasers: a tutorial. *IEEE Journal of Lightwave Technology* 22(1), 193–202.
- Cole, S., R. Fork, D. Lamb, and P. Reardon (2000). Multi-turn all-reflective optical gyroscope. *Optics Express* 7(8), 285–291.
- Colladon, D. (1842). Sur les réflexions d'un rayon de lumière à l'intérieur d'une veine liquide parabolique. *Compte Rendu des Séances de l'Académie des Sciences* 15, 800–802.
- Conant, R. (2003). *Micromachined Mirrors*. Dordrecht: Kluwer Academic Publishers.
- Conradie, E. and D. Moore (2002). SU-8 thick photoresist processing as a functional material for MEMS applications. *Journal of Micromechanics and Microengineering* 12(4), 368–374.
- Corcoran, B., C. Monat, C. Grillet, D. Moss, B. Eggleton, T. White, L. O'Faolin, and T. Krauss (2009). Green light emission in silicon through slow-light enhanced third-harmonic generation in photonic-crystal waveguides. *Nat Photonics* 3, 206–210.
- Cormack, A. (1957). Fourier transforms in cylindrical coordinates. *Acta Crystallographica* 10, 354–358.
- Cornelissen, S. A., P. A. Bierden, T. Bifano, and C. Lam (2009). 4096-element continuous face-sheet MEMS deformable mirror for high-contrast imaging. *Journal of Micro/Nanolithography, MEMS and MOEMS* 8(3), 031308.
- Culshaw, B. and J. Dakin (Eds.) (1997). *Optical Fiber Sensors*, Volume 1. Boston: Artech Houser.
- Dai, E., C. Zhou, P. Xi, and L. Liu (2003). Multifunctional double-layered diffractive optical element. *Optics Letters* 28(7), 1513–1515.
- Daly, D. (2001). *Microlens Arrays*. London: Taylor & Francis.
- Daly, D., R. Stevens, M. Hutley, and N. Davies (1990). The manufacture of microlenses by melting photoresist. *Measurement Science and Technology* 1, 759–766.
- Dammann, H. and K. Görtler (1971). High-efficiency in-line multiple imaging by means of multiple phase holograms. *Optics Communications* 3(5), 312–315.
- Das, A., A. Ganguly, and A. Saha (1998). Multilayered step index microlens of low spherical aberration. In Q. Xin and R. Parks (Eds.), *Current Developments in Optical Elements and Manufacturing*, Volume 3557 of *Proceedings of SPIE*, pp. 168–174.
- Däschner, W., P. Long, R. Stein, C. Wu, and S. H. Lee (1997). Cost-effective mass fabrication of multilevel diffractive optical elements by use of a single optical exposure with a gray-scale mask on high-energy beam-sensitive glass. *Applied Optics* 36(20), 4675–4680.
- Datta, M., Z. Hu, and M. Dagenais (2003). A novel method for fabrication of a hybrid optoelectronic packaging platform utilizing passive-active alignment. *IEEE Photonics Technology Letters* 15(2), 299–301.
- d'Auria, L., J. Huignard, A. Roy, and E. Spitz (1972). Photolithographic fabrication of thin film lenses. *Optics Communications* 5(4), 232–235.
- Davidson, N., R. Duer, A. Friesem, and E. Hasman (1992). Blazed holographic gratings for polychromatic and multidirectional incidence light. *Journal of the Optical Society of America A* 9(7), 1196–1199.

- Davidson, N., A. A. Friesem, and E. Hasman (1993). Analytic design of hybrid diffractive-refractive achromats. *Applied Optics* 32(25), 4770–4774.
- Davies, D. (1988). The CD-ROM medium. *Journal of the American Society for Information Science* 39(1), 34–42.
- Dayal, P. B. and F. Koyama (2007). Polarization control of 0.85 μm vertical-cavity surface-emitting lasers integrated with gold nanorod arrays. *Applied Physics Letters* 91(11), 111107–1 – 111107–3.
- De Gennes, P. (1985). Wetting: statics and dynamics. *Reviews of Modern Physics* 57(3, Part 1), 827–863.
- de Gennes, P. and J. Prost (1995). *The Physics of Liquid Crystals*. Oxford: Oxford University Press.
- de Lange, F., A. Cambi, R. Huijbens, B. de Bakker, W. Rensen, M. Garcia-Parajo, N. van Hulst, and C. Figdor (2001). Cell biology beyond the diffraction limit: near-field scanning optical microscopy. *Journal of Cell Science* 114(23), 4153–4160.
- Dean, J. and S. Ghemawat (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM* 51(1), 107–113.
- Debernardi, P., B. Kogel, K. Zogal, P. Meissner, M. Maute, M. Ortsiefer, G. Bohm, and M. Amann (2008). Modal properties of long-wavelength tunable MEMS-VCSELs with curved mirrors: comparison of experiment and modeling. *IEEE Journal of Quantum Electronics* 44(4), 391 – 399.
- Debernardi, P., J. Ostermann, M. Feneberg, C. Jalics, and R. Michalzik (2005). Reliable polarization control of VCSELs through monolithically integrated surface gratings: a comparative theoretical and experimental study. *IEEE Journal of Selected Topics in Quantum Electronics* 11(1), 107–116.
- Dereniak, E. and G. Boreman (1996). *Infrared detectors and systems*. New York: Wiley.
- Deri, R. and E. Kapon (1991). Low-loss III-V semiconductor optical waveguides. *IEEE Journal of Quantum Electronics* 27(3), 626–640.
- Derickson, D. (1998). *Fiber Optic Test and Measurement*. Upper Saddle River, NJ, USA: Prentice Hall.
- Destouches, N., A. Tishchenko, J. Pommier, S. Reynaud, O. Parriaux, S. Tonchev, and M. Abdou Ahmed (2005). 99% efficiency measured in the -1st order of a resonant grating. *Optics Express* 13(9), 3230–3235.
- Deubel, M., M. Wegener, S. Linden, G. von Freymann, and S. John (2006). 3D-2D-3D photonic crystal heterostructures fabricated by direct laser writing. *Optics Letters* 31(6), 805–807.
- Dewald, D. (2000). Using ZEMAX image analysis and user-defined surfaces for projection lens design and evaluation for Digital Light Processing projection systems. *Optical Engineering* 39(7), 1802–1807.
- Dharmarasu, N. and H. Hillmer (2007). Compound semiconductors. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 1, Chapter 1.02, pp. 25–52. Amsterdam: Elsevier.
- Dickensheets, D. (2008). Requirements of MEMS membrane mirrors for focus adjustment and aberration correction in endoscopic confocal and optical coherence tomography imaging instruments. *Journal of Micro/Nanolithography, MEMS and*

- MOEMS* 7(2), 021008.
- Diehl, L., D. Bour, S. Corzine, J. Zhu, G. Höfler, M. Lončar, M. Troccoli, and F. Capasso (2006). High-power quantum cascade lasers grown by low-pressure metal organic vapor-phase epitaxy operating in continuous wave above 400 K. *Applied Physics Letters* 88, 201115.
- Digonnet, M. (Ed.) (2001). *Rare-Earth-Doped Fiber Lasers and Amplifiers* (2nd ed.). Boca Raton, FL, USA: CRC Press.
- Dirac, P. (1927). The quantum theory of the emission and absorption of radiation. *Proceedings of the Royal Society of London. Series A* 114(767), 243–265.
- Ditlbacher, H., J. Krenn, N. Felidj, B. Lamprecht, G. Schider, M. Salerno, A. Leitner, and F. R. Aussenegg (2002). Fluorescence imaging of surface plasmon fields. *Applied Physics Letters* 80(3), 404–406.
- Doble, N., G. Yoon, L. Chen, P. Bierden, B. Singer, S. Olivier, and D. Williams (2002). Use of a microelectromechanical mirror for adaptive optics in the human eye. *Optics Letters* 27(17), 1537–1539.
- Doehring, T., P. Hartmann, R. Jedamzik, and A. Thomas (2005). Status of ZERODUR mirror blank production at Schott. In H. P. Stahl (Ed.), *Optical Manufacturing and Testing VI*, Volume 5869 of *Proceedings of SPIE*, pp. 586902.1–586902.9.
- Doerr, C. and K. Okamoto (2006). Advances in silica planar lightwave circuits. *IEEE Journal of Lightwave Technology* 24(12), 4763–4789.
- Dolling, G., M. Wegener, and S. Linden (2007). Realization of a three-functional-layer negative-index photonic metamaterial. *Optics Letters* 32(5), 551–553.
- Domachuk, P., M. Cronin-Golomb, B. J. Eggleton, S. Mutzenich, G. Rosengarten, and A. Mitchell (2005). Application of optical trapping to beam manipulation in optofluidics. *Optics Express* 13(19), 7265–7275.
- Domash, L., M. Wu, N. Nemchuk, and E. Ma (2004). Tunable and switchable multiple-cavity thin film filters. *IEEE Journal of Lightwave Technology* 22(1), 126–135.
- Dong, L., A. K. Agarwal, D. J. Beebe, and H. Jiang (2006). Adaptive liquid microlenses activated by stimuli-responsive hydrogels. *Nature* 442, 551–554.
- Dotan, I. E., D. Goldring, and D. Mendlovic (2009). Hybrid add-drop filter based on one-dimensional photonic crystal Fabry-Perot resonator. *Journal of Nanophotonics* 3(1), 039502.
- Dötsch, H., N. Bahlmann, O. Zhuromskyy, M. Hammer, L. Wilkens, R. Gerhardt, P. Hertel, and A. Popkov (2005). Applications of magneto-optical waveguides in integrated optics: review. *Journal of the Optical Society of America B* 22(1), 240–253.
- Drachev, V., M. Thoreson, V. Nashine, E. N. Khaliullin, D. Ben-Amotz, V. J. Davisson, and V. M. Shalaev (2005). Adaptive silver films for surface-enhanced Raman spectroscopy of biomolecules. *Journal of Raman Spectroscopy* 36, 648–656.
- Draheim, J., F. Schneider, R. Kamberger, C. Müller, and U. Wallrabe (2009). Fabrication of a fluidic membrane lens system. *Journal of Micromechanics and Microengineering* 19, 095013.
- Ebbesen, T., H. Lezec, H. Ghaemi, T. Thio, and P. Wolff (1998). Extraordinary optical transmission through sub-wavelength hole arrays. *Nature* 391, 667–669.

- Ebi, G., M. Reckermann, H. Zappe, and J. Söchtig (1999). An integrated optical non-contact torque measurement microsystem. *Optical Engineering* 38(2), 240–245.
- Edison, T. (1881). Electric lamp and the manufacture thereof. US Patent 251,549.
- Edlén, B. (1966). The refractive index of air. *Metrologia* 2(2), 71–80.
- Eggleton, B. J., A. Ahuja, P. S. Westbrook, J. A. Rogers, P. Kuo, T. N. Nielsen, and B. Mikkelsen (2000). Integrated tunable fiber gratings for dispersion management in high-bit rate systems. *IEEE Journal of Lightwave Technology* 18(10), 1418–1432.
- Eichenfield, M., C. P. Michael, R. Perahia, and O. Painter (2007). Actuation of micro-optomechanical systems via cavity-enhanced optical dipole forces. *Nature Photonics* 1(7), 416–422.
- Eimerl, D. (1987). Crystal symmetry and the electro-optic effect. *IEEE Journal of Quantum Electronics* QE-23, 2104–2115.
- Einstein, A. (1905). Zur Elektrodynamik bewegter Körper. *Annalen der Physik* 322(10), 891–921.
- Einstein, A. (1906). Zur Theorie der Lichterzeugung und Lichtabsorption. *Annalen der Physik* 325(6), 199–206.
- Eldada, L. (2004). Optical communication components. *Review of Scientific Instruments* 75(3), 575–593.
- Eldada, L. and L. Shacklette (2000). Advances in polymer integrated optics. *IEEE Journal of Selected Topics in Quantum Electronics* 6(1), 54–68.
- Enkrich, C., M. Wegener, S. Linden, S. Burger, L. Zschiedrich, F. Schmidt, J. F. Zhou, T. Koschny, and C. M. Soukoulis (2005). Magnetic metamaterials at telecommunication and visible frequencies. *Physical Review Letters* 95, 203901.
- Erickson, D., S. Mandal, A. Yang, and B. Cordovez (2008). Nanobiosensors: optofluidic, electrical and mechanical approaches to biomolecular detection at the nanoscale. *Microfluidics and Nanofluidics* 4, 33–52.
- Erickson, D., T. Rockwood, T. Emery, A. Scherer, and D. Psaltis (2006). Nanofluidic tuning of photonic crystal circuits. *Optics Letters* 31(1), 59–61.
- Fabry, C. and A. Pérot (1901). On a new form of interferometer. *The Astrophysical Journal* 13, 265–272.
- Fahr, S., T. Clausnitzer, E.-B. Kley, and A. Tünnermann (2007). Reflective diffractive beam splitter for laser interferometers. *Applied Optics* 46(24), 6092–6095.
- Faist, J., C. Gmachl, F. Capasso, C. Sirtori, D. Sivco, J. Baillargeon, and A. Cho (1997). Distributed feedback quantum cascade lasers. *Applied Physics Letters* 70(20), 2670–2672.
- Fan, H. (1967). Effects of free carriers on the optical properties. In R. Willardson and A. Beer (Eds.), *Semiconductors and Semimetals*, Volume 3, New York, pp. 405–420. Academic Press.
- Fan, S., M. F. Yanik, Z. Wang, S. Sandhu, and M. L. Povinelli (2006). Advances in theory of photonic crystals. *IEEE Journal of Lightwave Technology* 24(12), 4493–4501.
- Fan, Y., H. Ren, X. Liang, H. Wang, and S. Wu (2005). Liquid crystal microlens arrays with switchable positive and negative focal lengths. *IEEE/OSA Journal of Display Technology* 1(1), 151–156.

- Fanderlik, I. (1983). *Optical Properties of Glass*. Number 5 in Glass Science and Technology. Amsterdam: Elsevier.
- Faraday, M. (1846). Experimental researches in electricity. Nineteenth series. *Philosophical Transactions of the Royal Society of London* 136, 1–20.
- Ferry, V., M. Verschuuren, H. Li, R. Schropp, H. Atwater, and A. Polman (2009). Improved red-response in thin film a-Si:H solar cells with soft-imprinted plasmonic back back reflectors. *Applied Physics Letters* 95, 183503.
- Fiala, J., R. Gehrke, N. Weber, P. Bingger, H. Zappe, and A. Seifert (2009). Implantable optical sensor for continuous monitoring of various hemoglobin derivatives and tissue perfusion. In *Proceedings of IEEE Sensors 2009*. IEEE.
- Fick, A. (1988). A contact-lens. 1888 (translation). *Archives of Ophthalmology* 106(10), 1373–1377.
- Fiedler, F. and A. Schlachetzki (1987). Optical parameters of InP-based waveguides. *Solid-State Electronics* 30, 73–83.
- Findakly, T. (1985). Glass waveguides by ion exchange: a review. *Optical Engineering* 24(2), 244–250.
- Fizeau, H. (1850). Versuch, die Fortpflanzungsgeschwindigkeit des Lichts zu bestimmen. *Annalen der Physik* 155(1), 167–169.
- Forrest, S. (2000). Active optoelectronics using thin-film organic semiconductors. *IEEE Journal of Selected Topics in Quantum Electronics* 6(6), 1072–1083.
- Forrest, S. (2004). The path to ubiquitous and low-cost organic electronic appliances on plastic. *Nature* 428, 911–918.
- Fossum, E. (1997). CMOS image sensors: Electronic camera-on-a-chip. *IEEE Transactions on Electron Devices* 44(10), 1689–1698.
- Franco, M., A. Passaro, J. Cardoso, and J. Machado (1999). Finite element analysis of anisotropic optical waveguide with arbitrary index profile. *IEEE Transactions on Magnetics* 35(3), 1546–1549.
- Franz, W. (1958). Einfluß eines elektrischen Feldes auf eine optische Absorptionskante. *Zeitschrift für Naturforschung* 13a, 484–489.
- Friese, C., A. Werber, F. Krogmann, W. Mönch, and H. Zappe (2007). Materials, effects and components for tunable micro-optics. *IEEE Transactions on Electrical and Electronic Engineering* 2(3), 232–248.
- Friese, C. and H. Zappe (2008). Deformable polymer adaptive optical mirrors. *IEEE Journal of Microelectromechanical Systems* 17(1), 11–19.
- Fu, Y. and N. Bryan (2002). Design of hybrid micro-diffractive-refractive optical element with wide field of view for free space optical interconnections. *Optics Express* 10(13), 540–549.
- Fudouzi, H. and T. Sawada (2006). Photonic rubber sheets with tunable color by elastic deformation. *Langmuir* 22, 1365–1368.
- Fudouzi, H. and Y. Xia (2003). Photonic papers and inks: color writing with colorless materials. *Advanced Materials* 15(11), 892–896.
- Fujita, H. (1997). A decade of MEMS and its future. In *Proceedings of the 1997 IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, pp. 1–7.

- Fujita, J., M. Levy, R. Osgood, L. Wilkens, and H. Dötsch (2000). Waveguide optical isolator based on Mach-Zehnder interferometer. *Applied Physics Letters* 76(16), 2158–2160.
- Fujita, T., H. Nishihara, and J. Koyama (1981). Fabrication of micro lenses using electron-beam lithography. *Optics Letters* 6(12), 613–615.
- Fujita, T., H. Nishihara, and J. Koyama (1982). Blazed gratings and Fresnel lenses fabricated by electron-beam lithography. *Optics Letters* 7(12), 578–580.
- Fukazawa, T., F. Ohno, and T. Baba (2004). Very compact arrayed-waveguide-grating demultiplexer using Si photonic wire waveguides. *Japanese Journal of Applied Physics* 43(5B), L673–L675.
- Fukuda, M. (1999). *Optical Semiconductor Devices*. New York: Wiley.
- Fukuda, Y., T. Watanabe, T. Wakimoto, S. Miyaguchi, and M. Tsuchida (2000). An organic LED display exhibiting pure RGB colors. *Synthetic Materials* 111-112, 1–6.
- Furukawa, A., S. Sasaki, M. Hoshi, A. Matsuzono, K. Moritoh, and T. Baba (2004). High-power single-mode vertical-cavity surface-emitting lasers with triangular holey structure. *Applied Physics Letters* 85(22), 5161–5163.
- Gabor, D. (1948). A new microscopic principle. *Nature* 161, 777–778.
- Gabrielli, L., J. Cardenas, C. Poitras, and M. Lipson (2009). Silicon nanostructure cloak operating at optical frequencies. *Nature Photonics* 3, 461–463.
- Galas, J., J. Torres, M. Belotti, Q. Kou, and Y. Chen (2005). Microfluidic tunable dye laser with integrated mixer and ring resonator. *Applied Physics Letters* 86, 264101.
- Gale, M. (1998). Replication. In H. Herzig (Ed.), *Micro-Optics*, Chapter 6, pp. 153–177. London: Taylor & Francis.
- Gale, M., R. Kunz, and H. Zappe (1995). Polymer and III-V transducer platforms for integrated optical sensors. *Optical Engineering* 34(8), 2396–2406.
- Gale, M., M. Rossi, J. Pedersen, and H. Schütz (1994). Fabrication of continuous-relief micro-optical elements by direct laser writing in photoresist. *Optical Engineering* 33(11), 3556–3566.
- Gale, M., M. Rossi, H. Schütz, P. Ehbets, H.-P. Herzig, and D. Prongué (1993). Continuous-relief diffractive optical elements for two-dimensional array generation. *Applied Optics* 32(14), 2526–2533.
- Gambin, Y., O. Legrand, and S. Quake (2006). Microfabricated rubber microscope using soft solid immersion lenses. *Applied Physics Letters* 88(174102-1 – 17402-3).
- Gan, F. and F. Kärtner (2005). High-speed silicon electrooptic modulator design. *IEEE Photonics Technology Letters* 17(5), 1007–1009.
- Gascoyne, P., J. Vykoukal, J. Schwartz, T. Anderson, D. Vykoukal, K. Current, C. McConaghy, F. Becker, and C. Andrews (2004). Dielectrophoresis-based programmable fluidic processors. *Lab on a Chip* 4, 299–309.
- Geerlings, E., M. Rattunde, J. Schmitz, G. Kaufel, J. Wagner, B. Bläsi, D. Kallweit, and H. Zappe (2008). Widely tunable micro-mechanical external-cavity diode laser emitting around 2.1 μm . *IEEE Journal of Selected Topics in Quantum Electronics* 44(11), 1071–1075.
- Geerlings, E., M. Rattunde, J. Schmitz, G. Kaufel, H. Zappe, and J. Wagner (2006). Widely tunable GaSb-based external cavity diode laser emitting at around 2.3 μm .

- IEEE Photonics Technology Letters* 18(18), 1913–1915.
- Genov, D., S. Zhang, and X. Zhang (2009). Mimicking celestial mechanics in metamaterials. *Nature Physics* 5, 687–692.
- Gianchandani, Y., O. Tabata, and H. Zappe (Eds.) (2007). *Comprehensive Microsystems*, Amsterdam. Elsevier.
- Gianfrani, L., G. Gagliardi, G. Pesce, and A. Sasso (1997). High-sensitivity detection of NO₂ using a 740 nm semiconductor diode laser. *Applied Physics B* 64(4), 487–491.
- Glasenapp, C., W. Mönch, H. Krause, and H. Zappe (2007). Biochip reader with dynamic holographic excitation and hyperspectral fluorescence detection. *Journal of Biomedical Optics* 12(1), 014038–1–014038–8.
- Gmachl, C., F. Capasso, D. Sivco, and A. Cho (2001). Recent progress in quantum cascade lasers and applications. *Reports on Progress in Physics* 64, 1533–1601.
- Godwin, W. (1797). *The Enquirer: Reflections on Education, Manners and Literature in a Series of Essays*. London: G.G. and R. Robinson.
- Gombert, A., B. Bläsi, C. Bühler, P. Nitz, and J. Mick (2004). Some application cases and related manufacturing techniques for optically functional microstructures on large areas. *Optical Engineering* 43(11), 2525–2533.
- Goodman, J. (2004). *Introduction to Fourier Optics* (3rd ed.). Colorado, USA: Roberts and Company.
- Goos, F. and H. Hänchen (1947). Ein neuer und fundamentaler Versuch zur Totalreflexion. *Annalen der Physik* 436, 333–346.
- Goos, F. and H. Lindberg-Hänchen (1949). Neumessung des Strahlversetzungseffektes bei Totalreflexion. *Annalen der Physik* 440, 251–252.
- Gowar, J. (1993). *Optical Communication Systems* (2nd ed.). New York: Prentice Hall.
- Grabarnik, S., R. Wolffenbittel, A. Emadi, and M. Loktev (2007). Planar double-grating microspectrometer. *Optics Express* 15(6), 3581–3588.
- Green, M., J. Zhao, A. Wang, P. Reece, and M. Gal (2001). Efficient silicon light-emitting diodes. *Nature* 412, 805–808.
- Griffis, T., J. Baker, S. Sargent, B. Tanner, and J. Zhang (2004). Measuring field-scale isotopic CO₂ fluxes with tunable diode laser absorption spectroscopy and micrometeorological techniques. *Agricultural and Forest Meteorology* 124, 15–29.
- Grunwald, R., S. Woggon, R. Ehlert, and W. Reinecke (1997). Thin-film microlens arrays with non-spherical elements. *Pure and Applied Optics* 6, 663–671.
- Gu, L., W. Jiang, X. Chen, L. Wang, and R. .Chen (2007). High speed silicon photonic crystal waveguide modulator for low voltage operation. *Applied Physics Letters* 90, 071105–1 – 071105–3.
- Gustavsson, J., A. Haglund, J. Vuküsić, J. Bengtsson, P. Jedrasik, and A. Larsson (2005). Efficient and individually controllable mechanisms for mode and polarization selection in VCSELs, based on a common, localized, sub-wavelength surface grating. *Optics Express* 13(17), 6626–6634.
- Haas, K. and K. Rose (2003). Hybrid inorganic/organic polymers with nanoscale building blocks: precursors, processing, properties and applications. *Reviews on Advanced Materials Science* 5, 47–52.

- Hadley, G. and R. Smith (1995). Full-vector waveguide modeling using an iterative finite-difference method with transparent boundary conditions. *IEEE Journal of Lightwave Technology* 13(3), 465–469.
- Haglund, A., J. Gustavsson, J. Bengtsson, P. Jedrasik, and A. Larsson (2006). Design and evaluation of fundamental-mode and polarization-stabilized VCSELs with a sub-wavelength surface grating. *IEEE Journal of Quantum Electronics* 42(3), 231–240.
- Hane, K. and M. Sasaki (2007). Micro-mirrors. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 3, Chapter 3.01, pp. 1–63. Amsterdam: Elsevier.
- Hänsch, T., M. Pernier, and A. Schawlow (1971, January). Laser action of dyes in gelatin. *IEEE Journal of Quantum Electronics January*, 45–46.
- Hariharan, P. (2003). *Optical Interferometry* (2nd ed.). San Diego: Academic Press.
- Hartmann, D. M., O. Kibar, and S. C. Esener (2001). Optimization and theoretical modeling of polymer microlens arrays fabricated with the hydrophobic effect. *Applied Optics* 40(16), 2736–2746.
- Haruna, M., M. Takahashi, K. Wakahayashi, and H. Nishihara (1990). Laser beam lithographed micro-Fresnel lenses. *Applied Optics* 29(34), 5120–5126.
- Haruna, M., S. Yoshida, H. Toda, and H. Nishihara (1987). Laser-beam writing system for optical integrated circuits. *Applied Optics* 26(21), 4587–4592.
- Hashimoto, T., Y. Nakasuga, Y. Yamada, H. Terui, M. Yanagisawa, Y. Akahori, Y. Tohmori, K. Kato, and Y. Suzuki (1998). Multichip optical hybrid integration technique with planar lightwave circuit platform. *IEEE Journal of Lightwave Technology* 16, 1249–1258.
- Hashizume, J. and F. Koyama (2004). Plasmon enhanced optical near-field probing of metal nanoaperture surface emitting laser. *Optics Express* 12(25), 6391–6396.
- Hattori, H., C. Seassal, E. Touraille, P. Rojo-Romeo, X. Letartre, G. Hollinger, P. Viktorovitch, L. Di Cioccio, M. Zussy, L. El Melhaoui, and J. Fedeli (2006). Heterogeneous integration of microdisk lasers on silicon strip waveguides for optical interconnects. *IEEE Photonics Technology Letters* 18(1), 223–225.
- Hayes, R. and B. Feenstra (2003). Video-speed electronic paper based on electrowetting. *Nature* 425, 383–385.
- Hecht, E. (2002). *Optics* (4th International ed.). Reading, MA: Addison-Wesley.
- Hecht, J. (1999). *City of Light*. Oxford: Oxford University Press.
- Heckele, M. and W. Schomburg (2004). Review on micro molding of thermoplastic polymers. *Journal of Micromechanics and Microengineering* 14, R1–R14.
- Heikenfeld, J., K. Zhou, E. Kreit, B. Raj, S. Yang, B. Sun, A. Milarcik, L. Clapp, and R. Schwartz (2009). Electrofluidic displays using Young–Laplace transposition of brilliant pigment dispersions. *Nature Photonics* 3, 292–296.
- Helbo, B., A. Kristensen, and A. Menon (2003). A micro-cavity fluidic dye laser. *Journal of Micromechanics and Microengineering* 13, 307–311.
- Held, G. (2001, March/April). On the road to OC-768. *IT-Professional*, 46–48.
- Helou, Z. E., B. Erba, S. Churassy, G. Wannous, M. Néri, R. Bacis, and E. Boursey (2006). Design and performance of a low-temperature-multi-pass-cell for absorbance measurements of atmospheric gases. application to ozone. *Journal of Quantitative*

- Spectroscopy and Radiative Transfer* 101(1), 119 – 128.
- Heng, X., D. Erickson, L. Baugh, Z. Yaqoob, P. Sternberg, D. Psaltis, and C. Yang (2006). Optofluidic microscopy—a method for implementing a high resolution optical microscope on a chip. *Lab on a Chip* 6, 1274–1276.
- Henry, C., G. Blonder, and R. Kazarinov (1989). Glass waveguides on silicon for hybrid optical packaging. *IEEE Journal of Lightwave Technology* 7(10), 1530–1539.
- Herzig, H. (1998). Design of refractive and diffractive micro-optics. In H. Herzig (Ed.), *Micro-Optics*, Chapter 1, pp. 1–30. London: Taylor & Francis.
- Herzig, H., E. Kley, M. Cummer, and L. Wittig (2005). Micro-optics. In M. Motamedi (Ed.), *MOEMS*, Bellingham, WA, pp. 75–120. SPIE.
- Herzig, H., D. Prongué, and R. Dändliker (1990). Design and fabrication of highly efficient fan-out elements. *Japanese Journal of Applied Physics* 29(7), L1307–L1309.
- Hickernell, F. (1988). Optical waveguide on silicon. *Solid State Technology* 31(11), 83–88.
- Hiltunen, J., K. Kautio, J. Makinen, P. Karioja, and S. Kauppinen (2002). Passive multimode fiber-to-edge-emitting laser alignment based on a multilayer LTCC substrate. *Proceedings of the 2002 IEEE Electronic Components and Technology Conference*, 815–820.
- Himeno, A., K. Kato, and T. Miya (1998). Silica-based planar lightwave circuits. *IEEE Journal of Selected Topics in Quantum Electronics* 4(6), 913 – 924.
- Hirschman, K., L. Tsybeskov, S. Duttagupta, and P. Fauchet (1996). Silicon-based visible light-emitting devices integrated into microelectronic circuits. *Nature* 384, 338–341.
- Hoffmann, H., W. Jochs, and G. Westenberger (1990). Dispersion formula for the thermo-optic coefficient of optical glasses. In A. Marker (Ed.), *Properties and Characteristics of Optical Glass II*, Volume 1327 of *Proceedings of SPIE*, pp. 219–230.
- Hofstetter, D., J. Faist, M. Beck, A. Müller, and U. Oesterle (1999). Demonstration of high-performance 10.16 μm quantum cascade distributed feedback lasers fabricated without epitaxial regrowth. *Applied Physics Letters* 75(5), 665–667.
- Hofstetter, D., B. Maisenhölder, and H. Zappe (1998). Quantum well intermixing for fabrication of lasers and photonic integrated circuits. *IEEE Journal of Selected Topics in Quantum Electronics* 4(4), 794–802.
- Hofstetter, D., H. Zappe, and R. Dändliker (1995). Monolithically integrated optical displacement sensor in GaAs/AlGaAs. *Electronics Letters* 31(24), 2121–2122.
- Hofstetter, D., H. Zappe, and R. Dändliker (1996). A monolithically integrated double Michelson interferometer for optical displacement measurement with direction determination. *IEEE Photonics Technology Letters* 8(10), 1370–1372.
- Hofstetter, D., H. Zappe, and R. Dändliker (1997). Optical displacement measurement with GaAs/AlGaAs-based monolithically integrated Michelson interferometers. *IEEE Journal of Lightwave Technology* 15(4), 663–670.
- Hofstetter, D., H. Zappe, J. Epler, and J. Söchtig (1994). Single-growth-step GaAs/AlGaAs distributed Bragg reflector lasers with holographically-defined recessed gratings. *Electronics Letters* 30(22), 1858–1859.

- Hohlfeld, D., M. Epmeier, and H. Zappe (2003). A thermally-tunable, silicon-based optical filter. *Sensors and Actuators A* 103(1–2), 93–99.
- Hohlfeld, D. and H. Zappe (2004). All-dielectric tunable optical filter based on the thermo-optic effect. *Journal of Optics A: Pure and Applied Optics* 6(6), 504–511.
- Hohlfeld, D. and H. Zappe (2007). Thermal and optical characterization of silicon-based tunable optical thin-film filters. *IEEE Journal of Microelectromechanical Systems* 16(3), 500–510.
- Hojo, J., Y. Naito, H. Mori, K. Foujikawa, N. Kato, T. Wakyama, E. Komatsu, and M. Itasaka (1991). A 1/3-in 510 (H) x 492 (V) CCD image sensor with mirror image function. *IEEE Transactions on Electron Devices* 38(5), 954–957.
- Hokari, R. and K. Hane (2008). A varifocal micromirror with pure parabolic surface using bending moment drive. *Proceedings of the 2008 IEEE/LEOS International Conference on Optical MEMS and Nanophotonics*, 92–93.
- Homola, J., S. Yee, and G. Gauglitz (1999). Surface plasmon resonance sensors: review. *Sensors & Actuators B* 54, 3–15.
- Hongbin, Y., Z. Guangya, C. Siong, L. Feiwen, W. Shouhua, and Z. Mingsheng (2008). An electromagnetically driven lamellar grating based Fourier transform microspectrometer. *Journal of Micromechanics and Microengineering* 18, 055016.
- Honma, M., T. Nose, and S. Sato (1999). Optimization of device parameters for minimizing spherical aberration and astigmatism in liquid crystal microlenses. *Optical Review* 6(2), 139–143.
- Hornbeck, L. (1993). Current status of the digital micromirror device (DMD) for projection television applications. In *Proceedings of the IEEE Electron Devices Meeting (IEDM 93)*, pp. 381–384.
- Hornbeck, L. (1998). Current status and future applications for DMDTM-based projection displays. In *Proceedings of the Fifth International Display Workshop (IDW 98)*, pp. 713–716.
- Hoshino, K. and I. Shimoyama (2003). Analysis of elastic micro optical components under large deformation. *Journal of Micromechanics and Microengineering* 13, 149–154.
- Houde-Walter, S., J. Inman, A. Dent, and G. Greaves (1993). Sodium and silver environments and ion-exchange processes in silicate and aluminosilicate glasses. *The Journal of Physical Chemistry* 97, 9330–9336.
- Huang, M., Y. Zhou, and C. Chang-Hasnain (2007). A surface-emitting laser incorporating a high-index-contrast subwavelength grating. *Nature Photonics* 1, 119–122.
- Huang, M. Y., Y. Zhou, and C. Chang-Hasnain (2008a). A nanoelectromechanical tunable laser. *Nature Photonics* 2(3), 180–184.
- Huang, P.-H., T.-C. Huang, Y.-T. Sun, and S.-Y. Yang (2008b). Fabrication of large area resin microlens arrays using gas-assisted ultraviolet embossing. *Optics Express* 16(5), 3041–3048.
- Hug, M., D. Rieser, P. Manns, and G. Kleer (2005). Investigations on process parameters influencing the quality of optical lenses formed by non-isothermal embossing of inorganic glasses. In A. Duparre, R. Geyl, L. Wang, and A. Friesem (Eds.), *Optical Fabrication, Testing, and Metrology II*, Volume 5965 of *Proceedings of SPIE*, pp.

- 59651X.
- Hunsperger, R. (2002). *Integrated Optics: Theory and Technology* (5th ed.). Berlin: Springer-Verlag.
- Hunt, H. and J. Wilkinson (2008). Optofluidic integration for microanalysis. *Microfluidics and Nanofluidics* 4, 53–79.
- Hutley, M. (1982). *Diffraction Gratings*. London: Academic Press.
- Hutley, M. (1997). Refractive lenslet arrays. In H. Herzig (Ed.), *Micro-Optics*, Chapter 5, pp. 127–152. London: Taylor & Francis.
- Hwang, S., D. Seo, J. An, M. Kim, W. Choi, S. Cho, S. Lee, H.-H. Park, and H. Cho (2006). Parallel optical transmitter module using angled fibers and a V-grooved silicon optical bench for VCSEL array. *IEEE Transactions on Advanced Packaging* 29(3), 457–462.
- Hybertsen, M. (1994). Absorption and emission of light in nanoscale silicon structures. *Physical Review Letters* 72(10), 1514–1517.
- Iacona, F., G. Franzò, and C. Spinella (2000). Correlation between luminescence and structural properties of Si nanocrystals. *Journal of Applied Physics* 87(3), 1295–1303.
- Iga, K., Y. Kokobun, and M. Oikawa (1984). *Fundamentals of Microoptics*. New York: Academic Press.
- Iga, K., M. Oikawa, S. Misawa, J. Banno, and Y. Kokubun (1982). Stacked planar optics: an application of the planar microlens. *Applied Optics* 21(19), 3456–3460.
- Iizuka, K. (1983). *Engineering Optics* (2nd ed.). Berlin: Springer-Verlag.
- Iizuka, K. (2002a). *Elements of Photonics*, Volume I. New York: Wiley.
- Iizuka, K. (2002b). *Elements of Photonics*, Volume II. New York: Wiley.
- Inoue, K. and K. Ohtaka (2004). *Photonic Crystals: Physics, Fabrication and Applications*. Berlin: Springer-Verlag.
- Intonti, F., S. Vignolini, V. Türeci, M. Colocci, P. Bettotti, L. Pavesi, S. L. Schweizer, R. Wehrspohn, and D. Wiersma (2006). Rewritable photonic circuits. *Applied Physics Letters* 89(21), 211117.
- Ippolito, S., B. Goldberg, and M. Ünlü (2001). High spatial resolution subsurface microscopy. *Applied Physics Letters* 78(26), 4071–4073.
- Ishikawa, K., J. Zhang, and A. Tuantranont (2003). An integrated micro-optical system for VCSEL-to-fiber active alignment. *Sensors & Actuators A* 103, 109–115.
- Jackson, J. (1998). *Classical Electrodynamics* (3rd ed.). New York: Wiley.
- Jalali, B. and S. Fathpour (2006). Silicon photonics. *IEEE Journal of Lightwave Technology* 24(12), 4600–4615.
- Jamieson, T. (1981). Thermal effects in optical systems. *Optical Engineering* 20(2), 156–160.
- Jenkins, A. (2008). The road to nanophotonics. *Nature Photonics* 2, 258–260.
- Jeong, K. and L. Lee (2005). A novel microfabrication of a self-aligned vertical comb drive on a single SOI wafer for optical MEMS applications. *Journal of Micromechanics and Microengineering* 15, 277–281.
- Jeong, K.-H., G. L. Liu, N. Chronis, and L. P. Lee (2004). Tunable microdoublet lens array. *Optics Express* 12(11), 2494–2500.

- Ji, C., M. Choi, S. Kim, K. Song, J. Bu, and H. Nam (2007). Electromagnetic two-dimensional scanner using radial magnetic field. *IEEE Journal of Microelectromechanical Systems* 16(4), 989–996.
- Joannopoulos, J., R. Meade, and J. Winn (1995). *Photonic Crystals: Molding the Flow of Light*. Princeton, NJ: Princeton University Press.
- Johnson, S., S. Fan, P. Villeneuve, J. D. Joannopoulos, and L. A. Kolodziejski (1999). Guided modes in photonic crystal slabs. *Physical Review B* 60(8), 5751–5758.
- Jones, R. (1941). New calculus for the treatment of optical systems. I. Description and discussion of the calculus. *Journal of the Optical Society of America* 31, 488–493.
- Jorgensen, C. and V. Yarberry (2001). A 3D geometry modeler for the SUMMiT V MEMS designer. *Proceedings of Modeling and Simulation of Microsystems 2001*, 594–597.
- Jung, I. W., S. Kim, and O. Solgaard (2009). High-reflectivity broadband photonic crystal mirror MEMS scanner with low dependence on incident angle and polarization. *IEEE Journal of Microelectromechanical Systems* 18(4), 924 – 932.
- Kaino, T. (2000). Waveguide fabrication using organic nonlinear optical materials. *Journal of Optics A* 2, R1–R7.
- Kallioniemi, I., T. Ammer, and M. Rossi (2000). Optimization of continuous-profile blazed gratings using rigorous diffraction theory. *Optics Communications* 177, 15–24.
- Kallweit, D. and H. Zappe (2006). Fabrication of bulk-Si micromirrors with an integrated tilt sensing mechanism. *Journal of Micromechanics and Microengineering* 16(2), 463–469.
- Kang, J., J. Moon, S. Lee, S. Park, S. Jang, S. Yang, and S.-M. Yang (2008). Thermoresponsive hydrogel photonic crystals by three-dimensional holographic lithography. *Advanced Materials* 20, 3061–3065.
- Kang, Y., J. Walish, T. Gorishnyy, and E. Thomas (2007). Broad-wavelength-range chemically tunable block-copolymer photonic gels. *Nature Materials* 6, 957–960.
- Kärkkäinen, A. H. O., J. M. Tamkin, J. D. Rogers, D. R. Neal, O. E. Hormi, G. E. Jabbour, J. T. Rantala, and M. R. Descour (2002). Direct photolithographic deforming of organomodified siloxane films for micro-optics fabrication. *Applied Optics* 41(19), 3988–3998.
- Karlsson, R. (2004). SPR for molecular interaction analysis: a review of emerging application areas. *Journal of Molecular Recognition* 17(3), 151–161.
- Katigbak, A., J. Strother, and J. Lin (2009). Compact silicon slot waveguide polarization splitter. *Optical Engineering* 48(8), 080503–1–3.
- Kato, H., T. Matsushita, A. Takayama, M. Egawa, and M. Inoue (2002). Properties of one-dimensional magnetophotonic crystals for use in optical isolator devices. *IEEE Transactions on Magnetics* 38(5), 3246–3248.
- Kawachi, M. (1996). Recent progress in silica-based planar lightwave circuits on silicon. *IEE Proceedings-Optoelectronics* 143(5), 257–262.
- Kedzierski, J., S. Berry, and B. Abedian (2009). New generation of digital microfluidic devices. *IEEE Journal of Microelectromechanical Systems* 18(4), 845 – 851.

- Keldysh, L. (1958). The effect of a strong electric field on the optical properties of insulating crystals. *Soviet Physics JETP* 34, 788–790.
- Kermene, V., A. Saviot, M. Vampouille, B. Colombeau, C. Froehly, and T. Dohnalik (1992). Flattening of spatial laser beam profile with low losses and minimal beam divergence. *Optics Letters* 17(12), 859.
- Kidger, M. (2002). *Fundamental Optical Design*. Bellingham, WA: SPIE Optical Engineering Press.
- Kidger, M. (2004). *Intermediate Optical Design*. Bellingham, WA: SPIE Optical Engineering Press.
- Kim, H., J. Ge, J. Kim, S. Choi, H. Lee, H. Lee, W. Park, Y. Yin, and S. Kwon (2009). Structural colour printing using a magnetically tunable and lithographically fixable photonic crystal. *Nature Photonics* 3, 534–540.
- Kim, J., M. Serpe, and L. Lyon (2004). Hydrogel microparticles as dynamically tunable microlenses. *Journal of the American Chemical Society* 126(31), 9512–9513.
- Kim, J.-H., H.-K. Lee, B. Kim, J.-W. Jeon, J.-B. Yoon, and E. Yoon (2003). A high fill-factor micro-mirror stacked on a crossbar torsion spring for electrostatically-actuated two-axis operation in a large-scale optical switch. *Proceedings of the 2003 IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, 259–262.
- Kim, S. H., K.-D. Lee, J.-Y. Kim, M.-K. Kwon, and S.-J. Park (2007). Fabrication of photonic crystal structures on light emitting diodes by nanoimprint lithography. *Nanotechnology* 18(5), 055306.
- Kim, T., S.-S. Lee, Y. Yee, J.-U. Bu, C.-G. Park, and M.-H. Ha (2002). Large tilt angle electrostatic force actuated micro-mirror. *IEEE Photonics Technology Letters* 14(11), 1569–1571.
- Kimerling, L. (2000). Silicon microphotronics. *Applied Surface Science* 159-160, 8–13.
- King, C., L. Lin, and M. Wu (1996). Out-of-plane refractive microlens fabricated by surface micromachining. *Photonics Technology Letters* 8(10), 1349–1351.
- Kinoshita, M., T. Kobayashi, M. Yagi, and T. Ikeda (2007). Fabrication of liquid crystal microlens arrays using dye-doped polymerizable liquid crystals. *Journal of Photopolymer Science and Technology* 20(1), 91–92.
- Kippenberg, T., J. Kalkman, A. Polman, and K. Vahala (2006). Demonstration of an erbium-doped microdisk laser on a silicon chip. *Physical Review A* 74, 051802(R).
- Kirchhoff, G. and R. Bunsen (1860). Chemical analysis by observation of spectra. *Annalen der Physik und der Chemie* 110, 161–189.
- Kittel, C. (2005). *Solid State Physics* (8th ed.). New York: Wiley.
- Kleer, G., W. Döll, M. Hock, P. Manns, and E. Schäffer (2003). Composite coating materials for the hot embossing of microoptical components in inorganic glasses. *Glass Science and Technology* 76, 7–12.
- Klein, M. (1979). Einstein and the development of quantum physics. In A. French (Ed.), *Einstein: A Centenary Volume*, Chapter 5, pp. 138. Cambridge, MA: Harvard University Press.
- Klein, R. and E. Voges (1994). Integrated-optic ammonia sensor. *Analytical and Bio-analytical Chemistry* 349, 394–398.

- Kneipp, K., Y. Wang, H. Kneipp, L. Perelman, I. Itzkan, R. R. Dasari, and M. S. Feld (1997). Single molecule detection using surface-enhanced Raman scattering (SERS). *Physical Review Letters* 78(9), 1667–1670.
- Knittel, J., H. Richter, M. Hain, S. Somalingam, and T. Tschudi (2005). Liquid crystal lens for spherical aberration compensation in a Blu-ray disc system. *IEE Proceedings - Science, Measurement and Technology* 152(1), 15 – 18.
- Koch, T. and U. Koren (1991). Semiconductor photonic integrated circuits. *IEEE Journal of Quantum Electronics* 27(3), 641–653.
- Kögel, B., H. Halbritter, M. Lackner, M. Schwarzott, M. Maute, M.-C. Amann, F. Winter, and P. Meissner (2006). Micromechanically widely tunable VCSEL for absorption spectroscopy at around 1.55 μm . *Proceedings of the 2006 IEEE/LEOS International Conference on Optical MEMS*, 7–8.
- Kong, S. and R. Wolffenbuttel (2005). Spectral performance of a micromachined infrared spectrum analyzer in silicon. *IEEE Transactions on Instrumentation and Measurement* 54(1), 264–267.
- Koos, C., P. Vorreau, T. Vallaitis, P. Dumon, W. Bogaerts, R. Baets, B. Esembeson, I. Biaggio, T. Michinobu, F. Diederich, W. Freude, and J. Leuthold (2009). All-optical high-speed signal processing with silicon–organic hybrid slot waveguides. *Nature Photonics* 3(4), 216—219.
- Kopitkovas, G., L. Urech, and T. Lippert (2006). Fabrication of micro-optics in polymers and in UV transparent materials. In J. Perriere, E. Millon, and E. Fogarassy (Eds.), *Recent Advances in Laser Processing of Materials*, pp. 105–136. Amsterdam: Elsevier.
- Korvink, J. and O. Paul (Eds.) (2006). *MEMS: A practical guide to design, analysis and applications*, Norwich, NY. William Andrew.
- Kotera, H., I. Kanno, T. Inoue, and T. Suzuki (2005). Deformable micro-mirror for adaptive optics composed of piezoelectric PZT thin films. *Proceedings of the 2005 IEEE/LEOS International Conference on Optical MEMS*, 31–32.
- Kotlyar, M., T. Karle, M. Settle, L. O’Faolain, and T. Krauss (2004). Low-loss photonic crystal defect waveguides in InP. *Applied Physics Letters* 84(18), 3588–3590.
- Kou, Q., I. Yesilyurt, and Y. Chen (2006). Collinear dual-color laser emission from a microfluidic dye laser. *Applied Physics Letters* 88, 091101.
- Koyama, F. (2006). Recent Advances of VCSEL Photonics. *IEEE Journal of Lightwave Technology* 24(12), 4502–4513.
- Koyama, K., M. Yoshita, M. Baba, T. Suemoto, and H. Akiyama (1999). High collection efficiency in fluorescence microscopy with a solid immersion lens. *Applied Physics Letters* 75(12), 1667–1669.
- Kraus, J. and K. Carver (1992). *Electromagnetics* (4th ed.). New York: McGraw Hill.
- Krause, E. (1880). *Erasmus Darwin*. New York: D. Appleton and Company.
- Krause, H., W. Mönch, and H. Zappe (2006). Replicated, high aspect-ratio micro-optical components fabricated from inorganic sol-gel materials. *Applied Optics* 45(20), 4843–4849.
- Krauss, T. (2003). Planar photonic crystal waveguide devices for integrated optics. *physica status solidi (a)* 197(3), 688–702.

- Kress, B. and P. Meyrueis (2000). *Digital Diffractive Optics*. Chichester, England: Wiley.
- Kress, B. and P. Meyrueis (2009). *Applied Digital Optics*. Chichester, England: Wiley.
- Kress, B., D. Zaleta, W. Daschner, K. Urquhart, R. Stein, and S. H. Lee (1993, August). Diffractive optics fabricated by direct write methods with an electron beam. In *NASA Marshall Space Flight Center, Conference on Binary Optics: An Opportunity for Technical Exchange p 195-205 (SEE N94-17331 03-74)*, pp. 195–205.
- Kreuzberger, T., A. Harnisch, M. Helgert, L. Erdmann, and R. Brunner (2009). Sol-gel process to cast quartz glass microlens arrays. *Microelectronic Engineering* 86(4-6), 1173–1175.
- Krishnamoorthy, A., L. Chirovsky, W. Hobson, R. Leibengath, S. Hui, G. Zydzik, K. Goossen, J. Wynn, B. Tseng, J. Lopata, J. Walker, J. Cunningham, and L. D'Asaro (1999). Vertical-cavity surface-emitting lasers flip-chip bonded to gigabit-per-second CMOS circuits. *IEEE Photonics Technology Letters* 11(1), 128–130.
- Krogmann, F., W. Mönch, A. Werber, and H. Zappe (2007). Tunable micro-fluidic micro-lenses. In K. Buschow, R. Cahn, M. Flemings, P. Veysiere, E. Kramer, and S. Mahajan (Eds.), *The Encyclopedia of Materials: Science and Technology*. Amsterdam: Pergamon.
- Krogmann, F., W. Mönch, and H. Zappe (2006). A MEMS-based variable micro-lens system. *Journal of Optics A: Pure and Applied Optics* 8, S330–S336.
- Krogmann, F., W. Mönch, and H. Zappe (2008a). Electrowetting for tunable micro-optics. *IEEE Journal of Microelectromechanical Systems* 17(6), 1501–1512.
- Krogmann, F., H. Qu, W. Mönch, and H. Zappe (2008b). Push/pull actuation using opto-electrowetting. *Sensors and Actuators A* 141, 499–505.
- Krogmann, F., R. Shaik, L. Lasinger, W. Mönch, and H. Zappe (2008c). Reconfigurable liquid micro-lenses with high positioning accuracy. *Sensors and Actuators A* 143, 129 – 135.
- Kroll, M., J. McClintock, and O. Ollinger (1987). Measurement of gaseous oxygen using laser diode spectroscopy. *Applied Physics Letters* 51(18), 1465–1467.
- Krupenkin, T., S. Yang, and P. Mach (2003). Tunable liquid microlens. *Applied Physics Letters* 82(3), 316–318.
- Ksendzov, A. and Y. Lin (2005). Integrated optics ring-resonator sensors for protein detection. *Optics Letters* 30(24), 3344–3346.
- Kuiper, S. and B. Hendriks (2004). Variable-focus liquid lens for miniature cameras. *Applied Physics Letters* 85(7), 1128–1130.
- Kuiper, S., B. H. W. Hendriks, J. F. Suijver, S. Deladi, and I. Helwigen (2007). Zoom camera based on liquid lenses. In D. L. Dickensheets, B. P. Gogoi, and H. Schenk (Eds.), *MOEMS and Miniaturized Systems VI*, Volume 6466 of *Proceedings of SPIE*, pp. 64660F. SPIE.
- Kuo, S.-M. and C.-H. Lin (2008). The fabrication of non-spherical microlens arrays utilizing a novel SU-8 stamping method. *Journal of Micromechanics and Microengineering* 18, 125012 (7pp).
- Kuo, Y., Y. Lee, Y. Ge, S. Ren, J. Roth, T. Kamins, D. Miller, and J. Harris (2006). Quantum-confined Stark effect in Ge/SiGe quantum wells on Si for optical modula-

- tors. *IEEE Journal of Selected Topics in Quantum Electronics* 12(6), 1503–1513.
- Kuwata-Gonokami, M., R. Jordan, A. Dodabalapur, H. Katz, M. Schilling, and R. Slusher (1995). Polymer microdisk and microring lasers. *Optics Letters* 20(20), 2093–2095.
- Kuznetsov, M. (1985). Radiation losses in dielectric waveguide Y-branch structures. *IEEE Journal of Lightwave Technology* LT-3(3), 674–677.
- Lal, S., S. Link, and N. Halas (2007). Nano-optics from sensing to waveguiding. *Nature Photonics* 1, 641–648.
- Lambeck, P. (2006). Integrated optical sensors for the chemical domain. *Measurement Science and Technology* 17, R93–R116.
- Lang, T., D. Genon-Catalot, P. Dandrea, I. S. Dupont, and P. Benech (1998). Integrated optical displacement sensor with four quadrature phase-shifted output signals. *Journal of Optics* 29, 135–140.
- Lang, W. (1996). Silicon microstructuring technology. *Materials Science and Engineering: R: Reports* 17(1), 1–55.
- Lao, Y., B. Sun, K. Zhou, and J. Heikenfeld (2008). Ultra-high transmission electrowetting displays enabled by integrated reflectors. *IEEE Journal of Display Technology* 4(2), 120–122.
- Larson, A., L. Sandström, S. Höjer, H. Ahlberg, and B. Broberg (1997). Evaluation of distributed Bragg reflector lasers for high-sensitivity near-infrared gas analysis. *Optical Engineering* 36(1), 117–123.
- Lawrence, M. (1993). Lithium niobate integrated optics. *Reports on Progress in Physics*, 363–429.
- Lee, C. and J. Yeh (2008). Development and evolution of MOEMS technology in variable optical attenuators. *Journal of Micro/Nanolithography, MEMS and MOEMS* 7(2), 021003.
- Lee, H., W. Hwang, M. Oh, H. Park, T. Zyung, and J. Kim (1997). High performance electro-optic polymer waveguide device. *Applied Physics Letters* 71(26), 3779–3781.
- Lee, J., B. Hong, W. Kim, S. Min, Y. Kim, M. V. Jouravlev, R. Bose, K. S. Kim, I.-C. Hwang, L. J. Kaufman, C. W. Wong, and P. K. K. S. Kim (2009). Near-field focusing and magnification through self-assembled nanoscale spherical lenses. *Nature* 460, 498–501.
- Leger, J. (1998). Laser beam shaping. In H. Herzig (Ed.), *Micro-Optics*, Chapter 9, pp. 223–257. London: Taylor & Francis.
- Leich, A., V. Hurm, J. Sohn, T. Feltgen, W. Bronner, K. Köhler, H. Walcher, J. Rosenzweig, and M. Schlechtweg (2002). 65 GHz bandwidth optical receiver combining a flip-chip mounted waveguide photodiode and GaAs-based HEMT distributed amplifier. *Electronics Letters* 38(25), 1706–1707.
- Leisher, P. O., A. J. Danner, and K. D. Choquette (2006). Single-mode 1.3- μm photonic crystal vertical-cavity surface-emitting laser. *IEEE Photonics Technology Letters* 18(20), 2156–2158.
- Lesem, L., P. Hirsch, and J. Jordan (1969, March). The kinoform: a new wavefront reconstruction device. *IBM Journal of Research and Development*, 150–155.

- Levy, M., R. Osgood, H. Hegde, F. Cadieu, R. Wolfe, and V. Fratello (1996). Integrated optical isolators with sputter-deposited thin-film magnets. *IEEE Photonics Technology Letters* 8(7), 903–905.
- Levy, U. and R. Shamai (2008). Tunable optofluidic devices. *Microfluidics and Nanofluidics* 4, 97–105.
- Lewis, G. (1926). The conservation of photons. *Nature* 118(2981), 874–875.
- Li, G. and P. Yu (2003). Optical intensity modulators for digital and analog applications. *IEEE Journal of Lightwave Technology* 21(9), 2010–2030.
- Li, H. and W. Huck (2002). Polymers in nanotechnology. *Current Opinion in Solid State and Materials Science* 6(1), 3–8.
- Li, H. and K. Iga (Eds.) (2003). *Vertical-Cavity Surface-Emitting Laser Devices*. Berlin: Springer-Verlag.
- Li, X., C. Antoine, D. Lee, J. Wang, and O. Solgaard (2006a). Tunable blazed gratings. *IEEE Journal of Microelectromechanical Systems* 15(3), 597–604.
- Li, Z., Z. Zhang, T. Emery, A. Scherer, and D. Psaltis (2006b). Single mode optofluidic distributed feedback dye laser. *Optics Express* 14(2), 696–701.
- Li, Z., Z. Zhang, A. Scherer, and D. Psaltis (2006c). Mechanically tunable optofluidic distributed feedback dye laser. *Optics Express* 14(22), 10494–10499.
- Liang, T. and H. Tsang (2005). Integrated polarization beam splitter in high index contrast silicon-on-insulator waveguides. *IEEE Photonics Technology Letters* 17(2), 393–395.
- Liao, L., A. Liu, D. Rubin, J. Basak, Y. Chetrit, H. Nguyen, R. Cohen, N. Izhaky, and M. Paniccia (2007). 40 Gbit/s silicon optical modulator for high-speed applications. *Electronics Letters* 43(22), 1196–1197.
- Lin, L., S. Lee, K. Pister, and M. Wu (1994). Micro-machined three-dimensional micro-optics for integrated free-space optical system. *IEEE Photonics Technology Letters* 6(12), 1445–1447.
- Lindberg, D. (1983). *Studies in the History of Medieval Optics*, Volume CB 186 of *Collected Studies Series*. London: Variorum Reprints.
- Lindlein, N., J. Lamprecht, and J. Schwider (2004). Interferometric measurement of microlenses including cylindrical lenses. In J. Jahns and K.-H. Brenner (Eds.), *Microoptics: From Technology to Applications*, pp. 179–213. Springer-Verlag.
- Lippmann, G. (1875). Relations entre les phénomènes électriques et capillaires. *Annales des Chimie et des Physique* 5, 494–549.
- Lipson, M. (2005). Guiding, modulating, and emitting light on silicon—challenges and opportunities. *IEEE Journal of Lightwave Technology* 23(12), 4222–4238.
- Lipson, M. (2009). Silicon photonics: the optical spice rack. *Electronics Letters* 45(12), 576–578.
- Lipson, S., H. Lipson, and D. Tannhauser (1995). *Optical Physics* (3rd ed.). Cambridge: Cambridge University Press.
- Little, B., S. Chu, P. Absil, J. Hryniewicz, F. Johnson, F. Seiferth, D. Gill, V. Van, O. King, and M. Trakalo (2004). Very high-order microring resonator filters for WDM applications. *IEEE Photonics Technology Letters* 16(10), 2263–2265.

- Little, B., S. Chu, H. Haus, J. Foresi, and J. Laine (1997). Microring resonator channel dropping filters. *IEEE Journal of Lightwave Technology* 15(6), 998–1005.
- Liu, A., L. Liao, D. Rubin, H. Nguyen, and B. Ciftcioglu (2007). High-speed optical modulation based on carrier depletion in a silicon waveguide. *Optics Express* 15(2), 660–668.
- Liu, A. and X. Zhang (2007). A review of MEMS external-cavity tunable lasers. *Journal of Micromechanics and Microengineering* 17, R1–R13.
- Liu, A., B. Zhao, F. Chollet, Q. Zou, A. Asundi, and H. Fujita (2000). Micro-optomechanical grating switches. *Sensors and Actuators A* 86, 127–134.
- Liu, C. (2006). *Foundations of MEMS*. Upper Saddle River, NJ, USA: Pearson Prentice Hall.
- Liu, C., V. Kamaev, and Z. Vardeny (2005). Efficiency enhancement of an organic light-emitting diode with a cathode forming two-dimensional periodic hole array. *Applied Physics Letters* 86, 143501.
- Liu, C., J. Park, and J. Choi (2008). A planar lens based on the electrowetting of two immiscible liquids. *Journal of Micromechanics and Microengineering* 18, 035023.
- Liu, G., J. Kim, Y. Lu, and L. Lee (2006). Optofluidic control using photothermal nanoparticles. *Nature Materials* 5, 27–32.
- Liu, H., M. Yan, P. Shum, H. Ghafouri-Shiraz, and D. Liu (2004). Design and analysis of anti-resonant reflecting photonic crystal VCSEL lasers. *Optics Express* 12(18), 4269–4274.
- Liu, J.-M. (2005). *Photonic Devices*. Cambridge: Cambridge University Press.
- Liu, Q., L. Liu, S. Lau, H. Zappe, and J. Epler (1996). Investigation of optical losses in photoelastic and ridge waveguides in GaAs–AlGaAs heterostructures. *IEEE Photonics Technology Letters* 8(6), 806–808.
- Lo, D., J. Parris, and J. Lawless (1993). Laser and fluorescence properties of dye-doped sol-gel silica from 400 nm to 800 nm. *Applied Physics B: Lasers and Optics* 56(6), 385–390.
- Loewen, E. and E. Popov (1997). *Diffraction Gratings and Applications*. Boca Raton, FL, USA: CRC Press.
- Lohmann, A. (1989). Scaling laws for lens systems. *Applied Optics* 28(23), 4996–4998.
- Lohmann, A. and D. Paris (1967). Binary Fraunhofer holograms, generated by computer. *Applied Optics* 6, 1739–1748.
- Lohmann, A. and R. Syms (2003). External cavity laser with a vertically etched silicon blazed grating. *IEEE Photonics Technology Letters* 15(1), 120 – 122.
- Londoño, C., W. Plummer, and P. Clark (1993). Athermalization of a single-component lens with diffractive optics. *Applied Optics* 32(13), 2295–2302.
- López, C. A. and A. H. Hirsra (2008). Fast focusing using a pinned-contact oscillating liquid lens. *Nature Photonics* 2(10), 610–613.
- López-Higuera, J. (Ed.) (2002). *Handbook of Optical Fibre Sensing Technology*, New York. Wiley.
- Lotsch, H. (1968). Reflection and refraction of a beam of light at a plane interface. *Journal of the Optical Society of America* 58, 551–561.

- Lu, Z., D. Lockwood, and J. Baribeau (1996). Visible light emitting Si/SiO₂ superlattices. *Solid State Electronics* 40(1-8), 197–201.
- Lutz, G. (1999). *Semiconductor Radiation Detectors*. Berlin: Springer-Verlag.
- Lutzenberger, B., D. Dickensheets, and T. Kaiser (2003). Large area molded silicon nitride micro mirrors. *IEEE Photonics Technology Letters* 15(10), 1407–1409.
- Lynch, D. and W. Livingston (2001). *Color and Light in Nature* (2nd ed.). Cambridge: Cambridge University Press.
- Lytle, J., G. Wilkerson, and J. Jaramillo (1979). Wideband optical transmission properties of seven thermoplastics. *Applied Optics* 18(11), 1842–1848.
- Lyubchanskii, I., N. Dadoenkova, M. Lyubchanskii, E. Shapovalov, and T. Rasing (2003). Magnetic photonic crystals. *Journal of Physics D* 36, 277–287.
- Ma, H., A. Jen, and L. Dalton (2002). Polymer-based optical waveguides: materials, processing, and devices. *Advanced Materials* 14(19), 1339–1365.
- Macleod, H. (2001). *Thin-Film Optical Filters* (3rd ed.). Bristol: IOP Publishing.
- Madelung, O. (1978). *Introduction to Solid State Theory*. Berlin: Springer-Verlag.
- Mader, D., P. Waibel, A. Seifert, and H. Zappe (2009). Multichamber tunable liquid microlenses with active aberration correction. In *Proceedings of the 2009 IEEE International Conference on Micro Electro Mechanical Systems (MEMS 09)*, pp. 991–994. IEEE.
- Madou, M. (1997). *Fundamentals of Microfabrication*. Boca Raton, FL, USA: CRC Press.
- Maeda, S., K. Abe, K. Yamamoto, O. Tohyama, and H. Ito (1996). Active endoscope with SMA (shape memory alloy) coil springs. *Proceedings of the 1996 IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, 290–295.
- Mahler, L., A. Tredicucci, F. Beltram, C. Walther, J. Faist, B. Witzigmann, H. Beere, and D. Ritchie (2009). Vertically emitting microdisk lasers. *Nature Photonics* 3(1), 46–49.
- Maier, S. and H. Atwater (2005). Plasmonics: localization and guiding of electromagnetic energy in metal/dielectric structures. *Journal of Applied Physics* 98, 011101.
- Maiman, T. (1960). Stimulated optical radiation in ruby. *Nature* 187, 493–494.
- Maisenhölder, B., H. Zappe, R. Kunz, P. Riel, M. Moser, and J. Edlinger (1997). A GaAs/AlGaAs refractometer platform for integrated optical sensing applications. *Sensors and Actuators B* 39(1–3), 324–329.
- Mait, J. (1997). Fourier array generators. In H.-P. Herzig (Ed.), *Micro-Optics*, Chapter 11, pp. 293–323. London: Taylor & Francis.
- Malacara, D. and S. DeVore (1992). Interferogram evaluation and wavefront testing. In D. Malacara (Ed.), *Optical Shop Testing* (2nd ed.), Chapter 13, pp. 455–500. New York: Wiley.
- Malacara, D., M. Servín, and Z. Malacara (2005). *Interferogram Analysis for Optical Testing*. London: Taylor & Francis.
- Malitson, I. (1965). Interspecimen comparison of the refractive index of fused silica. *Journal of the Optical Society of America* 55, 1205–1209.

- Malshe, A. and J. O'Conner (2005). MEMS and MOEMS packaging. In M. Motamedi (Ed.), *MOEMS*, Bellingham, WA, pp. 515–544. SPIE.
- Mancebo, T. and S. Bará (1998). Interferometric monitoring of surface shaping processes in microlenses produced by melting photoresist. *Journal of Modern Optics* 45(5), 1029–1037.
- Mansfield, S. and G. Kino (1990). Solid immersion microscope. *Applied Physics Letters* 57(24), 2615–2617.
- Manzardo, O., H. Herzig, C. Marxer, and N. de Rooij (1999). Miniaturized time-scanning Fourier transform spectrometer based on silicon technology. *Optics Letters* 24(23), 1705–1707.
- Mao, X., J. Waldeisen, B. Juluri, and T. Huang (2007). Hydrodynamically tunable optofluidic cylindrical microlens. *Lab on a Chip* 7, 1303–1308.
- Marcatili, E. (1969). Dielectric rectangular waveguide and directional coupler for integrated optics. *Bell System Technical Journal* 48, 2071–2102.
- Marcatili, E. and S. Miller (1969). Improved relations describing directional control in electromagnetic wave guidance. *Bell System Technical Journal* 48, 2161–2188.
- Marcou, J. (1997). *Plastic Optical Fibres*. New York: Wiley.
- Marcuse, D. (1974). *Theory of Dielectric Optical Waveguides*. New York: Academic Press.
- Marks, R., D. Mathine, G. Peyman, J. Schwiegerling, and N. Peyghambarian (2009). Adjustable fluidic lenses for ophthalmic corrections. *Optics Letters* 34(4), 515–517.
- Markus, K., D. Koester, A. Cowen, R. Mahadevan, V. Dhuler, D. Roberson, and L. Smith (1995). MEMS infrastructure: the multiuser MEMS processes (MUMPs). In K. W. Markus (Ed.), *Micromachining and Microfabrication Process Technology*, Volume 2639 of *Proceedings of SPIE*, pp. 54–63.
- Marom, D. (2007). Optical communications. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 3, Chapter 3.07, pp. 219–266. Amsterdam: Elsevier.
- März, R. (2004). Modeling of photonic waveguides and circuits. In J. Jahns and K.-H. Brenner (Eds.), *Microoptics: From Technology to Applications*, pp. 57–75. Springer-Verlag.
- Mason, B., J. Barton, G. Fish, L. Coldren, and S. DenBaars (2000). Design of sampled grating DBR lasers with integrated semiconductor optical amplifiers. *IEEE Photonic Technology Letters* 12(7), 762–764.
- Masuda, S., S. Fujioka, M. Honma, T. Nose, and S. Sato (1996). Dependence of the optical properties on the device and material properties in liquid crystal microlenses. *Japanese Journal of Applied Physics* 35, 4688–4672.
- Masuda, S., S. Takahashi, T. Nose, S. Sato, and H. Ito (1997). Liquid-crystal microlens with a beam-steering function. *Applied Optics* 36(20), 4772–4778.
- Maxwell, G., B. Manning, M. Nield, M. Hadow, C. Ford, M. Clements, S. Lucas, P. Townley, R. McDougall, S. Oliver, R. Cecil, L. Johnston, A. Poustie, R. Webb, I. Lealman, L. Rivers, J. King, S. Perrin, R. Moore, I. Reid, and D. Scrase (2002). Very low coupling loss, hybrid-integrated all-optical regenerator with passive assembly. *Proceedings of the 28th European Conference on Optical Communication*

- (*ECOC 2002*) 5, 1–2.
- Maxwell, J. (1861). On physical lines of force. *Scientific Papers 1*, 488–489.
- McCall, S., A. Levi, R. Slusher, S. Pearton, and R. Logan (1992). Whispering-gallery mode microdisk lasers. *Applied Physics Letters 60*(3), 289–291.
- McMurdy, J., G. Crawford, and G. Jay (2006). Monolithic microspectrometer using tunable ferroelectric liquid crystals. *Applied Physics Letters 89*, 081105–1 – 081105–3.
- Mekis, A., S. Fan, and J. Joannopoulos (1998). Bound states in photonic crystal waveguides and waveguide bends. *Physical Review B 58*(8), 4809–4817.
- Menz, W., J. Mohr, and O. Paul (2001). *Microsystem Technology*. Weinheim: Wiley-VCH.
- Merenda, F., J. Rohner, J. Fournier, and R. Salathé (2007). Miniaturized high-NA focusing-mirror multiple optical tweezers. *Optics Express 15*(10), 6075–6086.
- Merz, P., H. Quenzer, H. Bernt, B. Wanger, and M. Zoberbier (2003). A novel micro-machining technology for structuring borosilicate glass substrates. In *Proceedings of the 12th International Conference on Solid State Sensors, Actuators and Microsystems (Transducers 2003)*, pp. 258–261.
- Mescheder, U., C. Estan, G. Somogyi, and M. Freudenreich (2005). MOEMS focusing device with ideal parabolic deflection shape. In *Proceedings of the 13th International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers 2005)*, pp. 1000–1005.
- Messerschmidt, B., B. L. McIntyre, and S. N. Houde-Walter (1996). Desired concentration-dependent ion exchange for micro-optic lenses. *Applied Optics 35*(28), 5670–5676.
- Mias, S. and H. Camon (2008a). A review of active optical devices: I. Amplitude modulation. *Journal of Micromechanics and Microengineering 18*(8), 083001.
- Mias, S. and H. Camon (2008b). A review of active optical devices: II. Phase modulation. *Journal of Micromechanics and Microengineering 18*(8), 083002.
- Michelson, A. (1927). Measurement of the velocity of light between Mount Wilson and Mount San Antonio. *The Astrophysical Journal 65*(1), 1–13.
- Michelson, A. and E. Morley (1887). On the relative motion of the earth and the luminiferous ether. *American Journal of Science 34*(203), 333–345.
- Miler, M. and M. Skalsky (1980). Chirped and curved grating coupler focusing both outgoing beam and guided wave. *Optics Communications 33*(1), 13–16.
- Miller, D., D. Chemla, T. Damen, A. Gossard, W. Wiegmann, T. Wood, and C. Burrus (1984). Band-edge electroabsorption in quantum well structures: The quantum-confined Stark effect. *Physical Review Letters 53*(22), 2173–2176.
- Miller, L. (1969a). Controlled collapse reflow chip joining. *IBM Journal of Research and Development 13*, 239–250.
- Miller, O., J. McLeod, and W. Sherwood (1951). Thin sheet plastic fresnel lenses of high aperture. *Journal of the Optical Society of America 41*(11), 807–815.
- Miller, S. (1969b). Integrated optics - An introduction. *Bell System Technical Journal 48*, 2059–2069.

- Milne, J., J. Dell, A. Keating, and L. Faraone (2009). Widely tunable MEMS-based Fabry-Perot filter. *IEEE Journal of Microelectromechanical Systems* 18(4), 905–913.
- Miyajima, H., N. Asaoka, T. Isokawa, M. Ogata, and Y. Aoki (2003). A MEMS electromagnetic optical scanner for a commercial confocal laser scanning microscope. *IEEE Journal of Microelectromechanical Systems* 12(3), 243–251.
- Miyashita, T., K. Hamanaka, M. Kato, S. Ishihara, H. Sato, E. Sato, and T. Morokuma (2004). Wavefront aberration measurement technology for microlens using the mach-zehnder interferometer provided with a projected aperture. In W. Osten and E. Novak (Eds.), *Interferometry XII: Applications*, Volume 5532 of *Proceedings of SPIE*, pp. 117–127.
- Möller, S. and S. Forrest (2002). Improved light out-coupling in organic light emitting diodes employing ordered microlens arrays. *Journal of Applied Physics* 91(5), 3324–3327.
- MONA (2008). A European Roadmap for Photonics and Nanotechnologies. Technical report, European Union - MONA Consortium.
- Monat, C., P. Domachuk, and B. Eggleton (2007). Integrated optofluidics: A new river of light. *Nature Photonics* 1, 106–114.
- Mönch, W., J. Dehnert, E. Jaufmann, and H. Zappe (2006a). Flory-Huggins swelling of polymer Bragg mirrors. *Applied Physics Letters* 89(16), 164104.
- Mönch, W., J. Dehnert, O. Prucker, J. Rühle, and H. Zappe (2006b). Tunable Bragg filters based on polymer swelling. *Applied Optics* 45(18), 4284–4290.
- Mönch, W. and H. Zappe (2004). Fabrication and testing of micro-lens arrays by all-liquid techniques. *Journal of Optics A: Pure and Applied Optics* 6(4), 330–337.
- Moore, D. T. (1980). Gradient-index optics: a review. *Applied Optics* 19(7), 1035–1038.
- More, F. D., M. D. Giulio, S. Pietralunga, A. Zappettini, L. Nasi, V. Rigato, and M. Martinelli (2003). Sputtered stoichiometric TeO₂ glass films: Dispersion of linear and nonlinear optical properties. *Journal of Applied Physics* 94(3), 1654–1661.
- Morgan, R., M. Hibbs-Brenner, R. Walterson, J. Lehman, T. Marta, S. Bounnak, E. Kalweit, T. Akinwande, and J. Nohava (1995). Producibile GaAs-based MOVPE-grown vertical-cavity top-surface emitting lasers with record performance. *Electronics Letters* 31(6), 462–464.
- Morimune, T., H. Kajii, and Y. Ohmori (2006). Photoresponse properties of a high-speed organic photodetector based on copper-phthalocyanine under red light illumination. *IEEE Photonics Technology Letters* 18(24), 2662–2664.
- Morris, G. and K. McIntyre (1997). Optical system design with diffractive optics. In J. Turunen and F. Wyrowski (Eds.), *Diffractive Optics*, Chapter 3, pp. 81–104. Berlin: Akademie Verlag.
- Mortherier, G. and P. Vankwikelberge (1997). *Handbook of Distributed Feedback Laser Diodes*. Norwood, MA, USA: Artech House.
- Mosher, L., C. Waits, B. Morgan, and R. Ghodssi (2009). Double-exposure grayscale photolithography. *IEEE Journal of Microelectromechanical Systems* 18(2), 308–315.
- Motamedi, M. (Ed.) (2005). *MOEMS: Micro-Opto-Electro-Mechanical Systems*, Bellingham, WA. SPIE Optical Engineering Press.

- Mouroulis, P., W. Smith, and R. Johnson (Eds.) (2008). *Current Developments in Lens Design and Optical Engineering IX*, Volume 7060, Bellingham, WA. SPIE.
- Mugele, F. and J.-C. Baret (2005). Electrowetting: from basics to applications. *Journal of Physics: Condensed Matter* 17(28), R705–R774.
- Muhlschlegel, P., H. Eisler, O. Martin, B. Hecht, and D. Pohl (2005). Resonant optical antennas. *Science* 308, 1607–1609.
- Mulkens, J., D. Flagello, B. Streefkerk, and P. Graeupner (2004). Benefits and limitations of immersion lithography. *Journal of Microlithography, Microfabrication and Microsystems* 3(1), 104–114.
- Müller, C., A. Falcou, N. Reckefuss, M. Rojahn, V. Wiederhirn, P. Rudati, H. Frohne, O. Nuyken, H. Becker, and K. Meerholz (2003). Multi-colour organic light-emitting displays by solution processing. *Nature* 421, 829–833.
- Mulligan, J. (1998). Who were Fabry and Pérot? *American Journal of Physics* 66(9), 797–802.
- Murata, H. (1996). *Handbook of Optical Fibers and Cables* (2nd ed.). Boca Raton, FL, USA: CRC Press.
- Murphy, E. (1999). *Integrated optical circuits and components*. Boca Raton, FL, USA: CRC Press.
- Nagali, V., S. Chou, D. Baer, R. Hanson, and J. Segall (1996). Tunable diode-laser absorption measurements of methane at elevated temperatures. *Applied Optics* 35(21), 4026–4032.
- Nagarajan, R., C. Joyner, R. Schneider, J. Bostak, T. Butrie, A. Dentai, V. Dominic, P. Evans, M. Kato, M. Kauffman, D. Lambert, S. Mathis, A. Mathur, R. Miles, M. Mitchell, M. Missey, S. Murthy, A. Nilsson, F. Peters, S. Pennypacker, J. Pleumeekers, R. Salvatore, R. Schlenker, R. Taylor, H.-S. Tsai, M. Van Leeuwen, J. Webjorn, M. Ziari, D. Perkins, J. Singh, S. Grubb, M. Reffle, D. Mehuys, F. Kish, and D. Welch (2005). Large-scale photonic integrated circuits. *IEEE Journal of Selected Topics in Quantum Electronics* 11(1), 50–65.
- Najafi, S. (1992). *Introduction to Glass Integrated Optics*. Norwood, MA, USA: Artech House.
- Nakamura, S. and S. Chichibu (Eds.) (2000). *Introduction to Nitride Semiconductor Blue Lasers and Light Emitting Diodes*, London. Taylor & Francis.
- Nakamura, S., S. Pearton, and G. Fasol (2000). *The Blue Laser Diode* (2nd ed.). Berlin: Springer-Verlag.
- Nakazawa, M., E. Yoshida, T. Yamamoto, E. Yamada, and A. Sahara (1998). TDM single channel 640 Gbit/s transmission experiment over 60 km using 400 fs pulse train and walk-off free, dispersion flattened non-linear optical loop mirror. *Electronics Letters* 34(9), 907–908.
- Namjou, K., S. Cai, E. Whittaker, J. Faist, C. Gmachl, F. Capasso, D. Sivco, and A. Cho (1998). Sensitive absorption spectroscopy with a room-temperature distributed-feedback quantum-cascade laser. *Optics Letters* 23(3), 219–221.
- Narasimhamurty, T. S. (1981). *Photoelastic and Electro-optic Properties of Crystals*. New York: Plenum.

- Naulleau, P., E. Anderson, E. Gullikson, and J. Bokor (2001). Fabrication of high-efficiency multilayer-coated binary blazed gratings in the EUV regime. *Optics Communications* 200, 27–34.
- Naumov, A., M. Loktev, I. Guralnik, and G. Vdovin (1998). Liquid-crystal adaptive lenses with modal control. *Optics Letters* 23(13), 992–994.
- Newton, I. (1740). *Opticks* (1979 Reprint ed.). New York: Dover Publications.
- Neyer, A., B. Wittmann, and M. Johnck (1999). Plastic-optical-fiber-based parallel optical interconnects. *IEEE Journal of Selected Topics in Quantum Electronics* 5(2), 193–200.
- Ng, W., M. Lourenco, R. Gwilliam, S. Ledain, G. Shao, and K. Homewood (2001). An efficient room-temperature silicon-based light-emitting diode. *Nature* 410, 192–194.
- Nikolov, I. and C. Ivanov (2000). Optical plastic refractive measurements in the visible and near-infrared regions. *Applied Optics* 39(13), 2067–2070.
- Nilsson, D., S. Balslev, and A. Kristensen (2005). A microfluidic dye laser fabricated by nanoimprint lithography in a highly transparent and chemically resistant cyclo-olefin copolymer (COC). *Journal of Micromechanics and Microengineering* 15, 296–300.
- Nishi, Y. and R. Doering (Eds.) (2000). *Handbook of Semiconductor Manufacturing Technology*, New York. Marcel Dekker.
- Nishihara, H., M. Haruna, and T. Suhara (1989). *Optical Integrated Circuits*. New York: McGraw Hill.
- Noda, S. (2006). Recent progresses and future prospects of two- and three-dimensional photonic crystals. *IEEE Journal of Lightwave Technology* 24(12), 4554–4567.
- Noda, S., K. Tomoda, N. Yamamoto, and A. Chutinan (2000). Full three-dimensional photonic bandgap crystals at near-infrared wavelengths. *Science* 289, 604–606.
- Noginov, M., G. Zhu, A. Belgrave, and R. Bakker (2009). Demonstration of a spaser-based nanolaser. *Nature* 460, 1110–1112.
- Nomura, I., K. Kishino, T. Ebisawa, Y. Sawafuji, R. Ujihara, K. Tasai, H. Nakamura, T. Asatsuma, and H. Nakajima (2009). Photopumped green lasing on BeZnSeTe double heterostructures grown on InP substrates. *Applied Physics Letters* 94, 021104–1–021104–3.
- Nose, T., S. Masuda, and S. Sato (1991). Optical properties of a hybrid-aligned liquid crystal microlens. *Molecular Crystals and Liquid Crystals* 199, 27–35.
- Nose, T. and S. Sato (1989). A liquid crystal microlens obtained with a non-uniform electric field. *Liquid Crystals* 5(5), 1425–1433.
- Novotny, L. and B. Hecht (2006). *Principles of Nano-optics*. Cambridge: Cambridge University Press.
- Obayya, S. S. A., B. M. A. Rahman, and H. A. El-Mikati (2000). New full-vectorial numerically efficient propagation algorithm based on the finite element method. *IEEE Journal of Lightwave Technology* 18(3), 409–415.
- Oh, K. and C. Ahn (2007). Magnetic actuation. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 2, Chapter 2.02, pp. 39–68. Amsterdam: Elsevier.
- Oh, M., M. Lee, and H. Lee (1999). Polymeric waveguide polarization splitter with a buried birefringent polymer. *IEEE Photonics Technology Letters* 11(9), 1144–1146.

- Oikawa, M. (2004). The integrated microlens: The development of vertical and horizontal integration. In J. Jahns and K.-H. Brenner (Eds.), *Microoptics: From Technology to Applications*, pp. 215–224. Springer-Verlag.
- Okamoto, H., M. Wada, Y. Sakai, T. Hirono, Y. Kawaguchi, Y. Kondo, Y. Kadota, K. Kishi, and Y. Itaya (1998). A narrow beam 1.3- μm -super luminescent diode integrated with a spot-size converter and a new type rear absorbing region. *IEEE Journal of Lightwave Technology* 16(10), 1881–1887.
- Okamoto, K., J. Kashiwagi, T. Tanaka, and M. Kubota (2009). Nonpolar m-plane InGaN multiple quantum well laser diodes with a lasing wavelength of 499.8 nm. *Applied Physics Letters* 94, 071105–1 – 071105–3.
- Okamoto, K., I. Niki, A. Scherer, Y. Narukawa, T. Mukai, and Y. Kawamaki (2005). Surface plasmon enhanced spontaneous emission rate of InGaN/GaN quantum wells probed by time-resolved photoluminescence spectroscopy. *Applied Physics Letters* 87, 071102.
- Okano, H., A. Hiruta, H. Komano, T. Nishio, R. Takahashi, T. Kanaya, and S. Tsuji (1998). Hybrid integrated optical WDM transceiver module for FTTH systems. *Hitachi Cable Review* 17, 23–30.
- Ollier, E. (2002). Optical MEMS devices based on moving waveguides. *IEEE Journal of Selected Topics in Quantum Electronics* 8(1), 155–162.
- Ollier, E., P. Philippe, C. Chabrol, and P. Mottier (1999). Micro-opto-mechanical vibration sensor integrated on silicon. *IEEE Journal of Lightwave Technology* 17(1), 26–29.
- Onoe, H., K. Matsumoto, and I. Shimoyama (2007). Three-dimensional sequential self-assembly of microscale objects. *Small* 3(8), 1383–1389.
- Onoe, H., A. Nakai, E. Iwase, and K. Matsumoto (2009). Temperature-controlled transfer and self-wiring for multi-color light-emitting diode arrays. *Journal of Micromechanics and Microengineering* 19, 075015.
- Oppliger, Y., P. Sixt, J. Stauffer, J. Mayor, P. Regnault, and G. Voirin (1994). One-step 3D shaping using a gray-tone mask for optical and microelectronic applications. *Microelectronic Engineering* 23(1-4), 449–454.
- Oulton, R., V. Sorger, D. Genov, D. Pile, and X. Zhang (2008). A hybrid plasmonic waveguide for subwavelength confinement and long-range propagation. *Nature Photonics* 2, 496–500.
- Owa, S. and H. Nagasaka (2004). Advantage and feasibility of immersion lithography. *Journal of Microlithography, Microfabrication and Microsystems* 3(1), 97–103.
- Ozaktas, H., H. Urey, and A. Lohmann (1994). Scaling of diffractive and refractive lenses for optical computing and interconnections. *Applied Optics* 33(17), 3782–3789.
- Ozbay, E. (2006). Plasmonics: merging photonics and electronics at nanoscale dimensions. *Science* 311, 189–193.
- Painter, O., R. Lee, A. Scherer, A. Yariv, J. O’Brien, P. D. Dapkus, and I. Kim (1999). Two-dimensional photonic band-gap defect mode laser. *Science* 284.
- Papoulis, A. (1962). *The Fourier Integral and its Applications*. New York: McGraw Hill.

- Park, W. and J.-B. Lee (2004). Mechanically tunable photonic crystal structure. *Applied Physics Letters* 85(21), 4845–4847.
- Patel, N., S. Cinà, and J. Burroughes (2002). High-efficiency organic light-emitting diodes. *IEEE Journal of Selected Topics in Quantum Electronics* 8(2), 346–361.
- Pearce, C. (1983). Crystal growth and wafer preparation. In S. Sze (Ed.), *VLSI Technology*, Chapter 2, pp. 51–92. McGraw Hill.
- Pearsall, T. (1982). *GaInAsP Alloy Semiconductors*. Chichester, England: Wiley.
- Pelletier, N., B. Bêche, N. Tahani, J. Zyss, L. Camberlein, and E. Gaviot (2007). SU-8 waveguiding interferometric micro-sensor for gage pressure measurement. *Sensors & Actuators A* 135, 179–184.
- Pendry, J. (2000). Negative refraction makes a perfect lens. *Physical Review Letters* 85(18), 3966–3969.
- Pennings, E., G. Khoe, M. Smit, and T. Staring (1996). Integrated-optic versus microoptic devices for fiber-optic telecommunication systems: a comparison. *IEEE Journal of Selected Topics in Quantum Electronics* 2(2), 151–164.
- Petersen, K. (1977). Micromechanical light modulator array fabricated on silicon. *Applied Physics Letters* 31(8), 521–523.
- Petersen, K. (1980). Silicon torsional scanning mirror. *IBM Journal of Research and Development* 24(5), 631–637.
- Petersen, K. (2005). A new age for MEMS. In *Digest of Technical Papers: The 13th International Conference on Solid-State Sensors, Actuators and Microsystem (Transducers 2005)*, Volume 1, pp. 1–4.
- Pettinger, B. (2006). Tip-enhanced Raman spectroscopy (TERS). *Topics in Applied Physics* 103, 217–240.
- Pettit, G. and W. Turner (1965). Refractive index of InP. *Journal of Applied Physics* 36, 2081.
- Petzoldt, J. and W. Pannhorst (1991). Chemistry and structure of glass-ceramic materials for high precision optical applications. *Journal of Non-Crystalline Solids* 129, 191–198.
- Pietarinen, J., V. Kalima, T. T. Pakkanen, and M. Kuittinen (2008). Improvement of UV-moulding accuracy by heat and solvent assisted process. *Microelectronic Engineering* 85(2), 263–270.
- Pisignano, D., L. Persano, G. Gigli, P. Visconti, T. Stomeo, M. D. Vittorio, G. Barbarella, L. Favaretto, and R. Cingolani (2004). Planar organic photonic crystals fabricated by soft lithography. *Nanotechnology* 15(7), 766–770.
- Pitt, C. (1973). Sputtered-glass optical waveguides. *Electronics Letters* 9(17), 401–403.
- Platt, B. and R. Shack (2001). History and principles of Shack-Hartmann wavefront sensing. *Journal of Refractive Surgery* 17, S573–S577.
- Plaza, J., A. Llobera, C. Dominguez, J. Esteve, I. Salinas, J. Garcia, and J. Berganzo (2004). BESOI-based integrated optical silicon accelerometer. *IEEE Journal of Microelectromechanical Systems* 13(2), 355–364.
- Poffo, L., P. Lemaître-Auger, P. Benech, and P. Benech (2005). Improvement of a glass integrated optical displacement sensor interferometer using acousto-optic effect. *Measurement Science and Technology* 16, 1341–1348.

- Polman, A., G. Van den Hoven, J. Custer, J. Shin, R. Serna, and P. Alkemade (1995). Erbium in crystal silicon: Optical activation, excitation, and concentration limits. *Journal of Applied Physics* 77(3), 1256–1262.
- Popov, E., B. Bozkov, M. Sabeva, and D. Maystre (1995). Blazed holographic grating efficiency - numerical comparison with different profiles. *Optics Communications* 117, 413–416.
- Popovic, Z., R. Sprague, and G. N. Connell (1988). Technique for monolithic fabrication of microlens arrays. *Applied Optics* 27(7), 1281–1284.
- Pregla, R. (2004). Modeling of waveguide-optical devices by the method of lines. In J. Jahns and K.-H. Brenner (Eds.), *Microoptics: From Technology to Applications*, pp. 77–98. Springer-Verlag.
- Prieto, F., B. Sepúlveda, A. Calle, A. Llobera, C. Domínguez, A. Abad, A. Montoya, and L. Lechuga (2003). An integrated optical interferometric nanodevice based on silicon technology for biosensor applications. *Nanotechnology* 14, 907–912.
- Prongué, D., H. P. Herzig, R. Dändliker, and M. T. Gale (1992). Optimized kinoform structures for highly efficient fan-out elements. *Applied Optics* 31(26), 5706–5711.
- Psaltis, D., S. Quake, and C. Yang (2006). Developing optofluidic technology through the fusion of microfluidics and optics. *Nature* 442, 381–386.
- Que, L. (2007). Thermal actuation. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 2, Chapter 2.03, pp. 69–100. Amsterdam: Elsevier.
- Quilliet, C. and B. Berge (2001). Electrowetting: a recent outbreak. *Current Opinion in Colloid & Interface Science* 6, 34–39.
- Råde, L. and B. Westergren (1999). *Mathematics Handbook* (4th ed.). Berlin: Springer-Verlag.
- Rajic, S., J. Corbeil, and P. Datskos (2003). Feasibility of tunable MEMS photonic crystal devices. *Ultramicroscopy* 97, 473–479.
- Rakic, A. and M. Majewski (2003). Cavity and mirror design for vertical-cavity surface-emitting lasers. In H. Li and K. Iga (Eds.), *Vertical-Cavity Surface-Emitting Laser Devices*, Chapter 8, pp. 259–302. Springer-Verlag.
- Ralls, K., T. Courtney, and J. Wulf (1976). *Introduction to Materials Science and Engineering*. New York: Wiley.
- Rao, P., X. Wang, and A. Theuwissen (2008). CCD structures implemented in standard 0.18 μm CMOS technology. *Electronics Letters* 44(8), 548–549.
- Rao, Y. (1997). In-fibre Bragg grating sensors. *Measurement Science and Technology* 8, 355–375.
- Rasch, A., E. Handrich, G. Spahlinger, M. Hafen, S. Voigt, and M. Weingärtner (2004). Fiber-Optic Gyros and MEMS Accelerometers. In J. Jahns and K.-H. Brenner (Eds.), *Microoptics: From Technology to Applications*, pp. 275–288. Springer-Verlag.
- Rauch, T., M. Böberl, S. Tedde, J. Fürst, M. Kovalenko, G. Hesser, U. Lemmer, W. Heiss, and O. Hayden (2009). Near-infrared imaging with quantum-dot-sensitized organic photodiodes. *Nature Photonics* 3, 332–336.
- Ray, S. (2002). *Applied Photographic Optics* (3rd ed.). Oxford: Focal Press.

- Rayleigh, Lord (1874). On the manufacture and theory of diffraction gratings. *Philosophical Magazine* 47, 81–93, 193–205.
- Razavi, B. (2002). *Design of Integrated Circuits for Optical Communications*. New York: McGraw Hill.
- Reed, G. and A. Knights (2004). *Silicon Photonics*. Chichester, England: Wiley.
- Reichelt, S., J. Fiala, A. Werber, K. Förster, C. Heilmann, R. Klemm, and H. Zappe (2008, February). Development of an implantable optical pulse oximeter. *IEEE Transactions on Biomedical Engineering* 55(2), 581–588.
- Reichelt, S. and H. Zappe (2005). Combined Twyman-Green and Mach-Zehnder interferometer for micro-lens testing. *Applied Optics* 44(27), 5786–5792.
- Reichelt, S. and H. Zappe (2007). Design of spherically corrected, achromatic variable-focus liquid lenses. *Optics Express* 15(21), 14146–14154.
- Reineke, S., F. Lindner, G. Schwartz, and N. Seidler (2009). White organic light-emitting diodes with fluorescent tube efficiency. *Nature* 459, 234–239.
- Reinitzer, F. (1888). Beiträge zur Kenntniss des Cholesterins. *Monatshefte für Chemie* 9(1), 421–441.
- Remenyi, J., P. Koppa, L. Domjan, and E. Loerincz (2003). Phase modulation configuration of a liquid crystal display. In G. C. Righini and A. Consortini (Eds.), *19th Congress of the International Commission for Optics: Optics for the Quality of Life*, Volume 4829 of *Proceedings of SPIE*, pp. 815–816. SPIE.
- Ren, H., Y. Fan, and S. Wu (2003, Jan). Tunable Fresnel lens using nanoscale polymer-dispersed liquid crystals. *Applied Physics Letters* 83(8), 1515–1517.
- Ren, H. and S.-T. Wu (2008). Tunable-focus liquid microlens array using dielectrophoretic effect. *Optics Express* 16(4), 2646–2652.
- Revin, D., J. Cockburn, M. Steer, R. Airey, M. Hopkinson, A. Krysa, L. Wilson, and S. Menzel (2007). InGaAs / AlAsSb / InP quantum cascade lasers operating at wavelengths close to 3 μm . *Applied Physics Letters* 90, 021108.
- Rhoderick, E. and R. Williams (1988). *Metal-semiconductor contacts*. Oxford: Clarendon Press.
- Ridley, B. (1988). *Quantum Processes in Semiconductors*. Oxford: Clarendon Press.
- Ried, J., J. Shewchun, B. Garside, and E. Balik (1978). High sensitivity pollution detection employing tunable diode lasers. *Applied Optics* 17(2), 300–307.
- Righini, G., S. Pelli, M. Brenci, and M. Ferrari (2001). Active optical waveguides based on Er- and Er/Yb-doped silicate glasses. *Journal of Non-Crystalline Solids* 284, 223–229.
- Rinne, S. A., F. García-Santamaría, and P. V. Braun (2008). Embedded cavities and waveguides in three-dimensional silicon photonic crystals. *Nature Photonics* 2(1), 52–56.
- Rogalski, A. (2000). *Infrared detectors*. London: Taylor & Francis.
- Rogers, A. (2001). *Understanding Optical Fiber Communications*. Boston: Artech House.
- Roichman, Y. and D. Grier (2005). Holographic assembly of quasicrystalline photonic heterostructures. *Optics Express* 13(14), 5434–5439.
- Römer, H. (2006). *Theoretical Optics: An Introduction*. Weinheim: Wiley-VCH.

- Romero, L. A. and F. M. Dickey (1996). Lossless laser beam shaping. *Journal of the Optical Society of America A* 13(4), 751–760.
- Ronchi, V. (1970). *The Nature of Light: An Historical Survey*. Cambridge, MA: Harvard University Press.
- Rong, H., A. Liu, R. Jones, O. Cohen, D. Hak, R. Nicolaescu, A. Fang, and M. Paniccia (2005). An all-silicon Raman laser. *Nature* 493, 292–294.
- Rong, H., S. Xu, O. Cohen, O. Raday, M. Lee, V. Sih, and M. Paniccia (2008). A cascaded silicon Raman laser. *Nature Photonics* 2(3), 170–174.
- Rooman, C., M. Kuijk, R. Vounckx, and P. Heremans (2005). Reflective-refractive microlens for efficient light-emitting-diode-to-fiber coupling. *Optical Engineering* 44(9), 095005–1 – 095005–5.
- Rothman, L. S., A. Barbe, D. C. Benner, L. R. Brown, C. Camy-Peyret, M. R. Carleer, K. Chance, C. Clerbaux, V. Dana, V. M. Devi, A. Fayt, J. M. Flaud, R. R. Gamache, A. Goldman, D. Jacquemart, K. W. Jucks, W. J. Lafferty, J. Y. Mandin, S. T. Massie, V. Nemtchinov, D. A. Newnham, A. Perrin, C. P. Rinsland, J. Schroeder, K. M. Smith, M. A. H. Smith, K. Tang, R. A. Toth, J. V. Auwera, P. Varanasi, and K. Yoshino (2003). The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001. *Journal of Quantitative Spectroscopy and Radiative Transfer* 82(1–4), 5 – 44.
- Rudmann, H. and M. Rossi (2004). Design and fabrication technologies for ultraviolet replicated micro-optics. *Optical Engineering* 43(11), 2575–2582.
- Ruffieux, P., T. Scharf, H. Herzig, and R. Völkel (2006). On the chromatic aberration of microlenses. *Optics Express* 14(11), 4687–4694.
- Russell, P. S. J. (2006). Photonic-crystal fibers. *IEEE Journal of Lightwave Technology* 24(12), 4729–4749.
- Sadiku, M. (2001). *Elements of Electromagnetics* (3rd ed.). Oxford: Oxford University Press.
- Sagberg, H., M. Lacolle, I. Johansen, O. Løvhaugen, R. Belicov, O. Solgaard, and A. Sudbø (2004). Micromechanical gratings for visible and near-infrared spectroscopy. *IEEE Journal of Selected Topics in Quantum Electronics* 10(3), 604–613.
- Saleh, B. and M. Teich (1991). *Fundamentals of Photonics*. New York: Wiley.
- Sander, D. and J. Müller (2001). Selffocussing phase transmission grating for an integrated optical microspectrometer. *Sensors & Actuators A* 88, 1–9.
- Sanders, A., D. Routenberg, B. Wiley, Y. Xia, E. R. Dufresne, and M. Reed (2006). Observation of plasmon propagation, redirection, and fan-out in silver nanowires. *Nano Letters* 6(8), 1822–1826.
- Sanghera, J. and I. Aggarwal (1988). *Infrared Fiber Optics*. Boca Raton, FL, USA: CRC Press.
- Sargent, E. (2009). Infrared photovoltaics made by solution processing. *Nature Photonics* 3, 325–331.
- Sasaki, M., M. Tabata, T. Haga, and K. Hane (2006). Piezoresistive rotation angle sensor integrated in micromirror. *Japanese Journal of Applied Physics* 45, 3789–3793.
- Sasian, J. and R. Chipman (1993). Staircase lens: a binary and diffractive field curvature corrector. *Applied Optics* 32(1), 60–66.

- Sato, S. (1979). Liquid-crystal lens-cells with variable focal length. *Japanese Journal of Applied Physics* 18, 1679–1684.
- Schift, H. (2008). Nanoimprint lithography: An old story in modern times? A review. *Journal of Vacuum Science & Technology B* 26(2), 458–480.
- Schmitt, K., B. Schirmer, C. Hoffmann, A. Brandenburg, and P. Meyrueis (2007). Interferometric biosensor based on planar optical waveguide sensor chips for label-free detection of surface bound bioreactions. *Biosensors and Bioelectronics* 22, 2591–2597.
- Schnitzer, I., E. Yablonovitch, C. Caneau, T. Gmitter, and A. Scherer (1993). 30% external quantum efficiency from surface textured, thin-film light-emitting diodes. *Applied Physics Letters* 63(16), 2174–2176.
- Schubert, A., J. Edelmann, and T. Burkhardt (2006). Micro structuring of borosilicate glass by high-temperature micro-forming. *Microsystem Technologies* 12, 790–795.
- Schubert, C., J. Berger, S. Diez, H. Ehrke, R. Ludwig, U. Feiste, C. Schmidt, H. Weber, G. Toptchiyski, S. Randel, and K. Petermann (2002). Comparison of interferometric all-optical switches for demultiplexing applications in high-speed OTDM systems. *IEEE Journal of Lightwave Technology* 20(4), 618–624.
- Schwarzer, H. and S. Teiwes (1997). Diffractive elements for optical image processing. In J. Turunen and F. Wyrowski (Eds.), *Diffractive Optics*, Chapter 9, pp. 243–274. Berlin: Akademie Verlag.
- Seeger, K. (1999). *Semiconductor Physics* (7th ed.). Berlin: Springer-Verlag.
- Segré, G. (2007). *Faust in Copenhagen*. New York: Viking.
- Sellmeier, W. (1872). Zur Erklärung der abnormalen Farbenfolge im Spectrum einiger Substanzen. *Annalen der Physik und Chemie* 219(6), 272–282.
- Selvaganapathy, P. (2007). Polymers. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 1, Chapter 1.04, pp. 75–106. Amsterdam: Elsevier.
- Senturia, S. (2001). *Microsystem Design*. Dordrecht: Kluwer Academic Publishers.
- Senturia, S. (2002). Diffractive MEMS: the polychromator and related devices. *Proceedings of the 2002 IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, 5–6.
- Séviigné, Marquise de (1823). *Lettres de Madame de Séviigné, de sa famille, et de ses amis*. Paris: Chez Dalibon.
- Shalaev, V. (2007). Optical negative-index metamaterials. *Nature Photonics* 1, 41–48.
- Shannon, R. (1997). *The Art and Science of Optical Design*. Cambridge: Cambridge University Press.
- Shao, Y., D. Dickensheets, and P. Himmer (2004). 3-D MOEMS mirror for laser beam pointing and focus control. *IEEE Journal of Selected Topics in Quantum Electronics* 10(3), 528–535.
- Shen, Y., C. Friend, Y. Jiang, D. Jakubczyk, J. Swiatkiewicz, and P.N.Prasad (2000). Nanon photonics: interactions, materials and applications. *Journal of Physical Chemistry B* 104, 7577–7587.
- Shi, L., C. Du, X. Dong, Q. Deng, and X. Luo (2007). Effective formation method for an aspherical microlens array based on an aperiodic moving mask during exposure.

- Applied Optics* 46(34), 8346–8350.
- Shibaguchi, T. and H. Funato (1992). Lead-lanthanum zirconate-titanate (PLZT) electrooptic variable focal-length lens with stripe electrodes. *Japanese Journal of Applied Physics* 31(9B), 3196–3200.
- Shieh, H.-P., Y.-P. Huang, and K.-O. Chien (2005). Micro-optics for liquid crystal displays applications. *IEEE/OSA Journal of Display Technology* 1(1), 62–76.
- Shimizu, H. and Y. Nakano (2006). Fabrication and characterization of an InGaAsP/InP active waveguide optical isolator with 14.7 dB/mm TE mode nonreciprocal attenuation. *IEEE Journal of Lightwave Technology* 24(1), 38–43.
- Shinar, J. (2003). *Organic Light Emitting Devices*. Berlin: Springer-Verlag.
- Shinoda, M., K. Nakamura, M. Yabe, N. Watanabe, T. Satoh, N. Hirai, T. Fujita, K. Kime, and Y. Ishida (1996). Optical pick-up for DVD. *IEEE Transactions on Consumer Electronics* 42(3), 808–813.
- Shuto, Y., H. Sato, S. Yanagi, M. Ohno, and S. Tohno (1999). Optical characteristics and reliability of plastic split alignment sleeves for single-mode optical fiber connectors. *IEEE Journal of Selected Topics in Quantum Electronics* 5(5), 1418–1425.
- Siegman, A. (1986). *Lasers*. Sausalito, CA: University Science Books.
- Silfvast, W. (1996). *Laser Fundamentals*. Cambridge: Cambridge University Press.
- Simmons, J. H. and K. S. Potter (2000). *Optical Materials*. San Diego: Academic Press.
- Simonov, A. N., O. Akhzar-Mehr, and G. Vdovin (2005). Light scanner based on a viscoelastic stretchable grating. *Optics Letters* 30(9), 949–951.
- Sinclair, M., M. Butler, S. Kravitz, and W. Zubrzycki (1997). Synthetic infrared spectra. *Optics Letters* 22(13), 1036–1038.
- Singh, J., T. Gan, A. Agarwal, M., and S. Liw (2005). 3D free space thermally actuated micromirror device. *Sensors and Actuators A* 123-124, 468–475.
- Sinziger, S. and J. Jahns (2003). *Microoptics* (2nd ed.). Weinheim: Wiley-VCH.
- Sinzinger, S. (2007). Diffractive and defractive micro-optics. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 3, Chapter 3.02, pp. 65–99. Amsterdam: Elsevier.
- Skogen, E., J. Raring, G. Morrison, C. Wang, V. Lal, M. Mašanović, and L. Coldren (2005). Monolithically integrated active components: a quantum-well intermixing approach. *IEEE Journal of Selected Topics in Quantum Electronics* 11(2), 343–355.
- Smith, F. and T. King (2000). *Optics and Photonics*. New York: Wiley.
- Smith, G. and D. Atchison (1997). *The Eye and Visual Optical Instruments*. Cambridge: Cambridge University Press.
- Smith, H. (1996). A proposal for maskless, zone-plate-array nanolithography. *Journal of Vacuum Science & Technology B* 14(6), 4318–4322.
- Smith, R. (1978). *Semiconductors* (2nd ed.). Cambridge: Cambridge University Press.
- Smith, W. (1990). *Modern Optical Engineering* (2nd ed.). Boston: McGraw Hill.
- Soldano, L. and E. Pennings (1995). Optical multi-mode interference devices based on self-imaging: principles and applications. *IEEE Journal of Lightwave Technology* 13(4), 615–627.
- Solgaard, O. (2009). *Photonic Microsystems*. New York: Springer-Verlag.

- Solgaard, O., D. Lee, K. Yu, U. Krishnamoorthy, K. Li, and J. Heritage (2003, March). Microoptical phased arrays for spatial and spectral switching. *IEEE Communications Magazine*, 96–102.
- Somekh, S., E. Garmire, A. Yariv, H. Garvin, and R. Hunsperger (1973). Channel optical waveguide directional couplers. *Applied Physics Letters* 22(5), 46–47.
- Song, B., Y. Lee, D. Jang, J. Lee, and J. Lee (2007). High-performance optical pick-up actuator with singlet objective lens for BD/DVD/CD compatible drive. *Microsystem Technologies* 13, 1253–1260.
- Song, B., S. Noda, T. Asano, and Y. Akahane (2005). Ultra-high-Q photonic double-heterostructure nanocavity. *Nature Materials* 4, 207–210.
- Song, C., N. Nguyen, S. Tan, and A. Asundi (2009a). A micro optofluidic lens with short focal length. *Journal of Micromechanics and Microengineering* 19, 085012.
- Song, W., A. E. Vasdekis, Z. Li, and D. Psaltis (2009b). Low-order distributed feedback optofluidic dye laser with reduced threshold. *Applied Physics Letters* 94(5), 051117.
- Song, W., A. E. Vasdekis, Z. Li, and D. Psaltis (2009c). Optofluidic evanescent dye laser based on a distributed feedback circular grating. *Applied Physics Letters* 94(16), 161110.
- Soref, R. and B. Bennett (1987). Electrooptical effects in silicon. *IEEE Journal of Quantum Electronics* QE-23(1), 123–129.
- Soref, R. and J. Lorenzo (1986, Jan). All-silicon active and passive guided-wave components for $\lambda = 1.3$ and $1.6 \mu\text{m}$. *IEEE Journal of Quantum Electronics* QE-22(6), 873–879.
- Soukoulis, C., S. Linden, and M. Wegener (2007). Negative refractive index at optical wavelengths. *Science* 315, 47–49.
- Soutar, C. and K. Lu (1994). Determination of the physical properties of an arbitrary twisted-nematic liquid crystal cell. *Optical Engineering* 33(8), 2704–2712.
- Splett, A. and K. Petermann (1994). Low-loss single-mode optical waveguides with large cross-section in standard epitaxial silicon. *IEEE Photonics Technology Letters* 6(3), 425–427.
- Srinivasan, V., V. Pamula, and R. Fair (2004). An integrated digital microfluidic lab-on-a-chip for clinical diagnostics on human physiological fluids. *Lab on a Chip* 4, 310–315.
- Steele, R. (2008, February). Diode lasers track long-term trend. *Laser Focus World*, 59–69.
- Stegeman, G. and R. Stolen (1989). Waveguides and fibers for nonlinear optics. *Journal of the Optical Society of America B* 6(4), 652–662.
- Steier, W., A. Chen, S. Lee, S. Garner, H. Zhang, V. Chuyanov, L. Dalton, F. Wang, A. Ren, C. Zhang, G. Todorova, A. Harper, H. Fetterman, D. Chen, A. Udupa, D. Bhattacharya, and B. Tsap (1999). Polymer electro-optic devices for integrated optics. *Chemical Physics* 245, 487–506.
- Stern, M. (1997). Binary optics fabrication. In H. Herzig (Ed.), *Micro-Optics*, Chapter 3, pp. 53–85. London: Taylor & Francis.
- Stern, M., M. Holz, S. Medeiros, and R. Knowlden (1991). Fabricating binary optics: process variables critical to optical efficiency. *Journal of Vacuum Science & Technol-*

- ogy *B* 9(6), 3117–3121.
- Stewart, G. and B. Culshaw (1994). Optical waveguide modelling and design for evanescent field chemical sensors. *Optical and Quantum Electronics* 26, S249–S259.
- Stigwall, J. and J. Bengtsson (2004). Design of array of diffractive optical elements with inter-element coherent fan-outs. *Optics Express* 12(23), 5675–5683.
- Stockham, A. and J. Smith (2006). Tolerancing microlenses using ZEMAX. In E. Johnson, G. Nordin, and T. Suleski (Eds.), *Micromachining Technology for Micro-Optics and Nano-Optics IV*, Volume 6110 of *Proceedings of SPIE*, pp. 61100U.
- Stover, J. (1995). *Optical Scattering* (2nd ed.), Volume PM 24. Bellingham, WA: SPIE Optical Engineering Press.
- Strandman, C. and Y. Bäcklund (1998). Passive and fixed alignment of devices using flexible silicon elements formed by selective etching. *Journal of Micromechanics and Microengineering* 8(1998), 39–44.
- Strassner, M., J. Daleiden, N. Chitica, D. Keiper, B. Stålnacke, S. Greek, and K. Hjort (2000). III–V semiconductor material for tunable Fabry–Perot filters for coarse and dense WDM systems. *Sensors and Actuators A* 85, 249–255.
- Streubel, K., N. Linder, R. Wirth, and A. Jaeger (2002). High brightness AlGaInP light-emitting diodes. *IEEE Journal of Selected Topics in Quantum Electronics* 8(2), 321–332.
- Su, G., C. Chiu, and F. Jiang (2005). Vertical micromirrors integrated with electromagnetic microactuators for two-dimensional optical matrix switches. *IEEE Photonics Technology Letters* 17(9), 1860–1862.
- Sugihwo, F., M. Larson, and J. Harris (1998). Micromachined widely tunable vertical cavity laser diodes. *IEEE Journal of Microelectromechanical Systems* 7(1), 48–55.
- Sugiura, N. and S. Morita (1993). Variable-focus liquid-filled optical lens. *Applied Optics* 32(22), 4181–4186.
- Suleski, T., B. Baggett, W. Delaney, C. Koehler, and E. Johnson (1999). Fabrication of high-spatial-frequency gratings through computer-generated near-field holography. *Optics Letters* 24(9), 602–604.
- Suleski, T. and R. Te Kolste (2005). Fabrication trends for free-space microoptics. *IEEE Journal of Lightwave Technology* 23(2), 633–646.
- Suleski, T. J. and D. C. O’Shea (1995). Gray-scale masks for diffractive-optics fabrication: I. Commercial slide imagers. *Applied Optics* 34(32), 7507–7517.
- Sun, Q., Y. Wang, L. Li, D. Wang, T. Zhu, J. Xu, C. Yang, and Y. Li (2007). Bright, multicoloured light-emitting diodes based on quantum dots. *Nature Photonics* 1(12), 717–722.
- Sun, Y. and S. Forrest (2008). Enhanced light out-coupling of organic light-emitting devices using embedded low-index grids. *Nature Photonics* 2(8), 483–487.
- Süsskind, C. (1988). Heinrich Hertz: A short life. *IEEE Transactions on Microwave Theory and Techniques* 36(5), 802–805.
- Suzuki, K., K. Takiguchi, and K. Hotate (2000). Monolithically integrated resonator microoptic gyro on silica planar lightwave circuit. *IEEE Journal of Lightwave Technology* 18(1), 66–72.

- Switkes, M., R. Kunz, M. Rothschild, R. Sinta, M. Yeung, and S.-Y. Baek (2003). Extending optics to 50 nm and beyond with immersion lithography. *Journal of Vacuum Science & Technology B* 21(6), 2794–2799.
- Syms, R. (1988). Advances in channel waveguide lithium niobate integrated optics. *Optical and Quantum Electronics* 20, 189–213.
- Syms, R. and A. Lohmann (2003). MOEMS tuning element for a Littrow external cavity laser. *IEEE Journal of Microelectromechanical Systems* 12(6), 921–928.
- Takahashi, H., S. Suzuki, K. Kato, and I. Nishi (1990). Arrayed-waveguide grating for wavelength division multi/demultiplexer with nanometre resolution. *Electronics Letters* 26(2), 87–88.
- Tanaka, N., N. Nakamura, Y. Matsunaga, S. Manabe, H. Tango, and O. Yoshida (1997). A low driving voltage CCD with single layer electrode structure for area image sensor. *IEEE Transactions on Electron Devices* 44(11), 1869–1874.
- Tanaka, Y., Y. Komma, Y. Shimizu, T. Shimazaki, J. Murata, and S. Mizuno (2004). Lens design of compatible objective lens for Blu-ray Disc and Digital Versatile Disk with diffractive optical element and phase steps. *Japanese Journal of Applied Physics* 43, 4742–4745.
- Tang, C. and S. VanSlyke (1987). Organic electroluminescent diodes. *Applied Physics Letters* 51(12), 913–915.
- Tang, S. K. Y., C. A. Stan, and G. M. Whitesides (2008). Dynamically reconfigurable liquid-core liquid-cladding lens in a microfluidic channel. *Lab on a Chip* 8, 395–401.
- Tang, Y., W. Ni, C. Sotomayor Torres, and G. Hansson (1995). Fabrication and characterisation of Si – Si_{0.7}Ge_{0.3} quantum dot light emitting diodes. *Electronics Letters* 31(16), 1385–1386.
- Testorf, M. and S. Sinziger (1995). Evaluation of microlens properties in the presence of high spherical aberration. *Applied Optics* 34(28), 6431–6437.
- Thibos, L., R. Applegate, J. Schwiegerling, and R. Webb (2002). Standards for reporting the optical aberrations of eyes. *Journal of Refractive Surgery* 18(5), S652–S660.
- Tien, P. (1971). Light waves in thin films and integrated optics. *Applied Optics* 10(11), 2395–2413.
- Tomlinson, W. J. (1980). Applications of GRIN-rod lenses in optical fiber communication systems. *Applied Optics* 19(7), 1127–1138.
- Toshiyoshi, H. (2007). Electrostatic actuation. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 2, Chapter 2.01, pp. 1–38. Amsterdam: Elsevier.
- Totschnig, G., M. Lackner, R. Shau, J. Roskopf, M. Amann, and F. Winter (2003). High-speed vertical-cavity surface-emitting laser (VCSEL) absorption spectroscopy of ammonia (NH₃) near 1.54 μm. *Applied Physics B* 76, 603–608.
- Totsu, K. and M. Esashi (2005). Gray-scale photolithography using maskless exposure system. *Journal of Vacuum Science & Technology B* 23(4), 1487–1490.
- Tran, A., Y. Lo, Z. Zhu, D. Haronian, and E. Mozdy (1996). Surface micromachined Fabry-Perot tunable filter. *IEEE Photonics Technology Letters* 8(3), 393–395.
- Traut, S., M. Rossi, and H. Herzig (2000). Replicated arrays of hybrid elements for application in a low-cost micro-spectrometer array. *Journal of Modern Optics* 47(13),

- 2391–2397.
- Trupke, T., J. Zhao, A. Wang, R. Corkish, and M. Green (2003). Very efficient light emission from bulk crystalline silicon. *Applied Physics Letters* 82(18), 2996–2998.
- Tsuchiya, T. (2007). Silicon and related materials. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 1, Chapter 1.01, pp. 1–24. Amsterdam: Elsevier.
- Tsuchizawa, T., K. Yamada, H. Fukuda, T. Watanabe, J. Takahashi, M. Takahashi, T. Shoji, E. Tamechika, S. Itabashi, and H. Morita (2005). Microphotonic devices based on silicon microfabrication technology. *IEEE Journal of Selected Topics in Quantum Electronics* 11(1), 232–240.
- Tsuji, S. and H. Kawaguchi (2005). Colored thin films prepared from hydrogel microspheres. *Langmuir* 21, 8439–8442.
- Tsurumi, N., W. Gao, A. Kimura, and S. Kiyono (2005). Replication of a sinusoidal grid surface by hot embossing and UV-casting. In F. Janabi-Sharifi (Ed.), *Optomechatronic Systems Control*, Volume 6052 of *Proceedings of SPIE*, pp. 60520P1–8. SPIE.
- Turunen, J. (1998). Diffraction theory of microrelief gratings. In H. Herzig (Ed.), *Micro-Optics*, Chapter 2, pp. 31–52. London: Taylor & Francis.
- Turunen, J. and F. Wyrowski (Eds.) (1997). *Diffraction Optics*. Berlin: Akademie Verlag.
- Tyson, R. (2002). Bit-error rate for free-space adaptive optics laser communications. *Journal of the Optical Society of America A* 19(4), 753–758.
- Uemukai, M., S. Nozu, and T. Suhara (2005). High-efficiency InGaAs QW distributed Bragg reflector laser with curved grating for squeezed light generation. *IEEE Journal of Selected Topics in Quantum Electronics* 11(5), 1143–1147.
- Urbach, F. (1953). The long-wavelength edge of photographic sensitivity and of the electronic absorption of solids. *Physical Review* 92, 1324.
- Urey, H. and D. Dickensheets (2005). Display and imaging systems. In M. Motamedi (Ed.), *MOEMS*, Bellingham, WA, pp. 369–452. SPIE.
- Vahala, K. (2003). Optical microcavities. *Nature* 424, 839–846.
- Valentine, J., S. Zhang, T. Zentgraf, E. Ulin-Avila, D. A. Genov, G. Bartal, and X. Zhang (2008). Three-dimensional optical metamaterial with a negative refractive index. *Nature* 455(7211), 376–379.
- Valette, S. (1988). State of the art of integrated optics technology at LETI for achieving passive optical components. *Journal of Modern Optics* 35(6), 993–1005.
- Valette, S., S. Renard, J. Jadot, P. Gidon, and C. Erbeia (1990). Silicon-based integrated optics technology for optical sensor applications. *Sensors and Actuators A* A21–A23, 1087–1091.
- van Damm, M., D. Le Mignant, and B. Macintosh (2004). Performance of the Keck observatory adaptive-optics system. *Applied Optics* 34(29), 5458–5467.
- van den Boom, H., W. Li, P. van Bennekom, I. Tafur Monroy, and G.-D. Khoe (2001). High-capacity transmission over polymer optical fiber. *IEEE Journal of Selected Topics in Quantum Electronics* 7(3), 461–470.
- Van Kessel, P., L. Hornbeck, R. Meier, and M. Douglass (1998). A MEMS-based projection display. *Proceedings of the IEEE* 86(8), 1687–1704.

- Vdovin, G. and L. Sarro (1994). Flexible reflecting membranes micromachined in silicon. *Semiconductor Science and Technology* 9, 1570–1572.
- Vedantam, S., H. Lee, J. Tang, J. Conway, M. Staffaroni, and E. Yablonovitch (2009). A plasmonic dimple lens for nanoscale focusing of light. *Nano Letters* 9(10), 3447–3452.
- Veldkamp, W. B. and E. J. V. Allen (1983). Binary holographic LO beam multiplexer for IR imaging detector arrays. *Applied Optics* 22(10), 1497–1507.
- Verrill, V. (1982). The effects of diamond wear on the production and properties of ruled diffraction gratings. *Journal of Physics E* 15, 516–519.
- Veselago, V. (1968). Electrodynamics of substances with simultaneously negative electrical and magnetic permeabilities. *Soviet Physics Uspekhi* 10(4), 5–13.
- Vezenov, D., B. Mayers, R. Conroy, G. M. Whitesides, P. T. Snee, Y. Chan, D. G. Nocera, and M. G. Bawendi (2005a). A low-threshold, high-efficiency microfluidic waveguide laser. *Journal of the American Chemical Society* 127, 8952–8953.
- Vezenov, D., B. Mayers, D. Wolfe, and G. Whitesides (2005b). Integrated fluorescent light source for optofluidic applications. *Applied Physics Letters* 86, 041104.
- Viheriälä, J., M.-R. Viljanen, J. Kontio, T. Leinonen, J. Tommila, M. Dumitrescu, T. Niemi, and M. Pessa (2009). Soft stamp UV-nanoimprint lithography for fabrication of laser diodes. *Journal of Micro/Nanolithography, MEMS and MOEMS* 8(3), 033004.
- Vineich, B., L. Evdokimovich, A. Safronov, and S. Smirnov (2004). Application of deformable mirrors in industrial CO₂ lasers. *Quantum Electronics* 34(4), 333–340.
- Vlasov, Y., W. Green, and F. Xia (2008). High-throughput silicon nanophotonic wavelength-insensitive switch for on-chip optical networks. *Nature Photonics* 2, 242–246.
- Völkel, R., H. Herzig, P. Nussbaum, and R. Dändliker (1996). Microlens array imaging system for photolithography. *Optical Engineering* 35(11), 3323–3330.
- Wada, O., T. Sakurai, and T. Nakagami (1986). Recent progress in optoelectric integrated circuits (OEIC's). *IEEE Journal of Quantum Electronics* QE-22(6), 805–821.
- Wale, M. and M. Goodwin (1992). Flip-chip bonding optimizes opto-ICs. *IEEE Circuits and Devices* November, 25–31.
- Walker, R. (1985). Simple and accurate loss measurement technique for semiconductor waveguides. *Electronics Letters* 21(13), 581–583.
- Walker, S. and J. Jahns (1990). Array generation with multilevel phase gratings. *Journal of the Optical Society of America A* 7(8), 1509–1513.
- Wallrabe, U., J. Korvink, and J. Mohr (2007). LIGA. In Y. Gianchandani, O. Tabata, and H. Zappe (Eds.), *Comprehensive Microsystems*, Volume 1, Chapter 1.11, pp. 293–340. Amsterdam: Elsevier.
- Wallrabe, U., C. Solf, J. Mohr, and J. Korvink (2005). Miniaturized Fourier transform spectrometer for the near infrared wavelength regime incorporating an electromagnetic linear actuator. *Sensors & Actuators A* 123-124, 459–467.
- Wang, S., J. Hurst, F. Ma, R. Sidhu, X. Sun, X. Zheng, A. Holmes, A. Huntington, L. Coldren, and J. Campbell (2002). Low-noise impact-ionization-engineered avalanche photodiodes grown on InP substrates. *IEEE Photonics Technology Let-*

- ters 14(12), 1722–1724.
- Wang, Y., Y. Kanamori, and K. Hane (2009). Pitch-variable blazed grating consisting of freestanding silicon beams. *Optics Express* 17(6), 4419–4426.
- Weigl, B., M. Grabherr, C. Jung, R. Jager, G. Reiner, R. Michalzik, D. Sowada, and K.-J. Ebeling (1997). High-performance oxide-confined GaAs VCSEL's. *IEEE Journal of Selected Topics in Quantum Electronics* 3(2), 409–415.
- Weinert, A. (1999). *Plastic Optical Fibers*. Erlangen and Munich: Publicis MCD Verlag.
- Weis, R. and T. Gaylord (1985). Lithium niobate: summary of physical properties and crystal structure. *Applied Physics A* 37, 191–203.
- Welch, D., F. Kish, R. Nagarajan, C. Joyner, R. Schneider, V. Dominic, M. Mitchell, S. Grubb, T.-K. Chiang, D. Perkins, and A. Nilsson (2006). The realization of large-scale photonic integrated circuits and the associated impact on fiber-optic communication systems. *IEEE Journal of Lightwave Technology* 24(12), 4674–4683.
- Weldon, V., J. O’Gorman, J. Pérez-Camacho, and J. Hegarty (1996). Oxygen sensing using single-frequency GaAs-AlGaAs DFB laser diodes and VCSELs. *Electronics Letters* 32(3), 219–221.
- Weldon, V., J. O’Gorman, P. Phelan, J. Hegarty, and T. Tanbun-Ek (1995). H₂S and CO₂ gas sensing using DFB laser diodes emitting at $\lambda=1.57\ \mu\text{m}$. *Sensors & Actuators B* 29(1–3), 101–107.
- Welstand, R., J. Zhu, W. Chen, A. Clawson, and S. Pappert (1999). Combined Franz-Keldysh and quantum-confined Stark effect waveguide modulator for analog signal transmission. *IEEE Journal of Lightwave Technology* 17(3), 497–502.
- Werber, A. and H. Zappe (2005). Tunable microfluidic microlenses. *Applied Optics* 44(16), 3238–3245.
- Werber, A. and H. Zappe (2006a). Thermo-pneumatically actuated, membrane-based micro-mirror devices. *Journal of Micromechanics and Microengineering* 16(12), 2524–2531.
- Werber, A. and H. Zappe (2006b). Tunable, membrane-based, pneumatic micro-mirror. *Journal of Optics A: Pure and Applied Optics* 8, S313–S317.
- Werber, A. and H. Zappe (2008, October). Tunable pneumatic micro-optics. *IEEE Journal of Microelectromechanical Systems* 17(5), 1218–1227.
- White, J. and W. A. Fraser (1949). Method of making optical elements. U.S. Patent 2,464,738.
- Wiederhecker, G., L. Chen, A. Gondarenko, and M. Lipson (2009). Controlling photonic structures using optical forces. *Nature* 462, 633–636.
- Wierer, J., A. David, and M. Megens (2009). III-nitride photonic-crystal light-emitting diodes with high extraction efficiency. *Nature Photonics* 3, 163–169.
- Wilbur, J., A. Kumar, E. Kim, and G. Whitesides (1994). Microfabrication by micro-contact printing of self-assembled monolayers. *Advanced Materials* 6(7/8), 600–604.
- Wilbur, J. and G. Whitesides (1999). Self-assembly and self-assembled monolayers in micro-and nanofabrication. In G. Timp (Ed.), *Nanotechnology*, Chapter 8, pp. 331–369. New York: Springer-Verlag.

- Williams, B., S. Kumar, Q. Hu, and J. Reno (2005). Operation of terahertz quantum-cascade lasers at 164 K in pulsed mode and at 117 K in continuous-wave mode. *Optics Express* 13(9), 3331–3339.
- Wilmsen, C., H. Temkin, and L. Coldren (Eds.) (2001). *Vertical-Cavity Surface-Emitting Lasers*, Cambridge. Cambridge University Press.
- Wippermann, F. C., P. Schreiber, A. Brauer, and P. Craen (2007). Bifocal liquid lens zoom objective for mobile phone applications. In M. M. Blouke (Ed.), *Sensors, Cameras, and Systems for Scientific/Industrial Applications VIII*, Volume 6501 of *Proceedings of SPIE*, pp. 650109. SPIE.
- Wirth, R., C. Karnutsch, S. Kugler, and K. Streubel (2001). High-efficiency resonant-cavity LEDs emitting at 650 nm. *IEEE Photonics Technology Letters* 13(5), 421–423.
- Wolfe, D., R. Conroy, P. Garstecki, B. Mayers, M. A. Fischbach, K. E. Paul, M. Prentiss, and G. M. Whitesides (2004). Dynamic control of liquid-core/liquid-cladding optical waveguides. *Proceedings of the National Academy of Sciences of the USA* 101(34), 12434–12438.
- Wolffenbuttel, R. (2004). State-of-the-art in integrated optical microspectrometers. *IEEE Transactions on Instrumentation and Measurement* 53(1), 197–202.
- Wollaston, W. (1802). A method of examining refractive and dispersive powers, by prismatic reflection. *Philosophical Transactions of the Royal Society of London* 92, 365–380.
- Wolters, G. (2005). Albert Einstein und Ernst Mach. In J. Renn (Ed.), *Albert Einstein: Ingenieur des Universums*, pp. 113. Weinheim: Wiley-VCH.
- Wood, A. P. (1992). Design of infrared hybrid refractive–diffractive lenses. *Applied Optics* 31(13), 2253–2258.
- Wood, R. (1902). On a remarkable case of uneven distribution of light in a diffraction grating spectrum. *Proceedings of the Physical Society of London* 18(1), 269–275.
- Wood, V., M. Panzer, J. Chen, M. Bradley, J. Halpert, M. Bawendi, and V. Bulović (2009). Inkjet-printed quantum dot-polymer composites for full-color ac-driven displays. *Advanced Materials* 21, 2151–2155.
- Wooten, E., K. Kissa, A. Yi-Yan, E. Murphy, D. Lafaw, P. Hallemeier, D. Maack, D. Attanasio, D. Fritz, G. McBrien, and D. Bossi (2000). A review of lithium niobate modulators for fiber-optic communications systems. *IEEE Selected Topics in Quantum Electronics* 6(1), 69–82.
- Wu, B., C. Chen, G. Ding, D. Zhang, and Y. Cui (2004). Hybrid-integrated Michelson fiber optic accelerometer. *Optical Engineering* 43(2), 313–318.
- Wu, M. (1997). Micromachining for optical and optoelectronic systems. *Proceedings of the IEEE* 85(11), 1833–1856.
- Wu, S. and H. Frankena (1993). Integrated optical sensors using micromechanical bridges and cantilevers. In M. Tabib-Azar and D. Polla (Eds.), *Integrated Optics and Microstructures*, Volume 1793 of *Proceedings of SPIE*, pp. 83–89. SPIE.
- Wu, S. and E. Glytsis (2003). Holographic grating formation in photopolymers: analysis and experimental results based on a nonlocal diffusion model and rigorous coupled-wave analysis. *Journal of the Optical Society of America B* 20(6), 1177–1188.

- Wyant, J. and K. Creath (1992). Basic wavefront aberration theory for optical metrology. In *Applied Optics and Optical Engineering*, Volume X, Chapter 1, pp. 2–53. Academic Press.
- Wyrowski, F. (1993). Design theory of diffractive elements in the paraxial domain. *Journal of the Optical Society of America A* 10(7), 1553–1561.
- Wysocki, G., R. Curl, F. Tittel, R. Maulini, J. Bulliard, and J. Faist (2005). Widely tunable mode-hop free external cavity quantum cascade laser for high resolution spectroscopic applications. *Applied Physics B* 81, 769–777.
- Xia, Y. and G. Whitesides (1998). Soft lithography. *Annual Review of Materials Science* 28, 153–184.
- Xiong, Z., G. Peng, B. Wu, and P. Chu (1999). Highly tunable Bragg gratings in single-mode polymer optical fibers. *IEEE Photonics Technology Letters* 11(3), 352–354.
- Xu, X., M. Mihnev, A. Taylor, and S. Forrest (2009). Organic photodetector arrays with indium tin oxide electrodes patterned using directly transferred metal masks. *Applied Physics Letters* 94, 043313.
- Xu, Y., J. Singh, C. Premachandran, A. Khairyanto, K. Chen, N. Chen, C. Sheppard, and M. Olivo (2008). Design and development of a 3D scanning MEMS OCT probe using a novel SiOB package assembly. *Journal of Micromechanics and Microengineering* 18, 125005.
- Yalcinkaya, A., H. Urey, D. Brown, and T. Montague (2006). Two-axis electromagnetic microscanner for high resolution displays. *IEEE Journal of Microelectromechanical Systems* 15(4), 786–794.
- Yan, R., D. Gargas, and P. Yang (2009). Nanowire photonics. *Nature Photonics* 3, 569–576.
- Yano, M., F. Yamagishi, and T. Tsuda (2005). Optical MEMS for Photonic Switching—Compact and Stable Optical Crossconnect Switches for Simple, Fast, and Flexible Wavelength Applications in Recent Photonic Networks. *IEEE Journal of Selected Topics in Quantum Electronics* 11(2), 383–394.
- Yariv, A. (1973). Coupled-mode theory for guided-wave optics. *IEEE Journal of Quantum Electronics* QE-9(9), 919–933.
- Yariv, A. (1991). *Optical Electronics* (4th ed.). Philadelphia: Saunders College Publishing.
- Yasufuku, S. (1992). Progress in materials application for optical disks in Japan. *IEEE Electrical Insulation Magazine* 8(1), 7–15.
- Ye, J., R. Zentel, S. Arpiainen, J. Ahopelto, F. Jonsson, S. G. Romanov, and C. M. S. Torres (2006). Integration of self-assembled three-dimensional photonic crystals onto structured silicon wafers. *Langmuir* 22, 7378–7383.
- Yeralan, S., J. Gunther, D. Ritums, R. Cid, and M. Popovich (2002). Switchable Bragg grating devices for telecommunications applications. *Optical Engineering* 41(8), 1774–1779.
- Yin, L., V. Vlasko-Vlasov, J. Pearson, J. M. Hiller, J. Hua, U. Welp, D. E. Brown, and C. W. Kimball (2005). Subwavelength focusing and guiding of surface plasmons. *Nano Letters* 5(7), 1399–1402.

- Yoshikawa, T., H. Sugimoto, H. Hotta, K. Tada, K. Kabayashi, F. Miyasaka, and K. Asakawa (1995). 650 nm AlGaInP visible light laser diode with dry-etched mesa stripe. *Japanese Journal of Applied Physics* 34(1), 1273–1278.
- Yoshikuni, Y. (2002). Semiconductor arrayed waveguide gratings for photonic integrated devices. *IEEE Journal of Selected Topics in Quantum Electronics* 8(6), 1102–1114.
- Young, T. (1804). Experimental demonstration of the general law of the interference of light. *Philosophical Transactions of the Royal Society of London* 94, 1–16.
- Yu, F. and X. Yang (1997). *Optical Engineering*. Cambridge: Cambridge University Press.
- Yu, K., D. Lee, U. Krishnamoorthy, N. Park, and O. Solgaard (2006). Micromachined Fourier transform spectrometer on silicon optical bench platform. *Sensors & Actuators: A* 130-131, 523–530.
- Yu, W., X. Yuan, N. Ngo, W. Que, W. Cheong, and V. Koudriachov (2002). Single-step fabrication of continuous surface relief micro-optical elements in hybrid sol-gel glass by laser direct writing. *Optics Express* 10(10), 443–448.
- Yuan, X., W. Yu, M. He, J. Bu, W. Cheong, H. Niu, and X. Peng (2005). Soft-lithography-enabled fabrication of large numerical aperture refractive microlens array in hybrid SiO₂ – TiO₂ sol-gel glass. *Applied Physics Letters* 86, 114102–1–3.
- Zaets, W. and K. Ando (1999). Optical waveguide isolator based on nonreciprocal loss/gain of amplifier covered by ferromagnetic layer. *IEEE Photonics Technology Letters* 11(8), 1012–1014.
- Zappe, H. (1995). *Introduction to Semiconductor Integrated Optics*. Boston: Artech House.
- Zappe, H. (1999). Semiconductor optical sensors. In H. Baltes, W. Göpel, and J. Hesse (Eds.), *Sensors Update*, Volume 5, Chapter 1, pp. 1–44. Weinheim: Wiley–VCH.
- Zappe, H. (2004). *Laser Diode Microsystems*. Berlin: Springer-Verlag.
- Zappe, H., H. Arnot, and R. Kunz (1994a). Technology and devices for hybrid and monolithic integrated optical sensors. *Sensors and Actuators A* 41(1–3), 141–144.
- Zappe, H., H. Arnot, and R. Kunz (1994b). Technology for III-V-based integrated optical sensors. *Sensors and Materials* 6(5), 261–270.
- Zappe, H., M. Hess, M. Moser, R. Hövel, K. Gulden, H.-P. Gauggel, and F. Monti di Sopra (2000). Narrow-linewidth vertical-cavity surface-emitting lasers for oxygen detection. *Applied Optics* 39(15), 2475–2479.
- Zeitner, U., P. Schreiber, and W. Karthe (2004). Modeling of free-space micro-optics. In J. Jahns and K.-H. Brenner (Eds.), *Microoptics: From Technology to Applications*, pp. 43–55. Springer-Verlag.
- Zernike, F. (1934). Beugungstheorie des Schneidverfahrens und seiner verbesserten Form, der Phasenkontrastmethode. *Physica* 1, 689–704.
- Zhang, D.-Y., N. Justis, V. Lien, Y. Berdichevsky, and Y.-H. Lo (2004a). High-performance fluidic adaptive lenses. *Applied Optics* 43(4), 783–787.
- Zhang, D.-Y., N. Justis, and Y.-H. Lo (2004b). Fluidic adaptive lens of transformable lens type. *Applied Physics Letters* 84(21), 4194–4196.

- Zhang, D.-Y., N. Justis, and Y.-H. Lo (2004c). Fluidic zoom-lens-on-a-chip with wide field-of-view tuning range. *IEEE Photonics Technology Letters* 16(10), 2356–2358.
- Zhang, D.-Y., N. Justis, and Y.-H. Lo (2004d). Integrated fluidic adaptive zoom lens. *Optics Letters* 29(24), 2855–2857.
- Zhang, D.-Y., V. Lien, Y. Berdichevsky, J. Choi, and Y.-H. Lo (2003). Fluidic adaptive lens with high focal length tunability. *Applied Physics Letters* 82(19), 3171–3172.
- Zhang, X., A. Liu, H. Cai, A. Yu, and C. Lu (2007). Retro-axial VOA using parabolic mirror pair. *IEEE Photonics Technology Letters* 19(9), 692–694.
- Zheng, B., J. Michel, F. Ren, L. Kimerling, D. Jacobson, and J. Poate (1994). Room-temperature sharp line electroluminescence at $\lambda=1.54\ \mu\text{m}$ from an erbium-doped, silicon light-emitting diode. *Applied Physics Letters* 64(21), 2842–2844.
- Zheng, X., V. Kaman, S. Yuan, Y. Xu, O. Jerphagnon, A. Keating, R. Anderson, H. Poulsen, B. Liu, J. Sechrist, C. Pularla, R. Helkey, D. Blumenthal, and J. Bowers (2003). Three-Dimensional MEMS Photonic Cross-Connect Switch Design and Performance. *IEEE Journal of Selected Topics in Quantum Electronics* 9(2), 571–578.
- Zhou, K., J. Heikenfeld, K. Dean, E. Howard, and M. R. Johnson (2009). A full description of a simple and scalable fabrication process for electrowetting displays. *Journal of Micromechanics and Microengineering* 19, 065029.
- Zhu, L., Y. Huang, and A. Yariv (2005). Integrated microfluidic variable optical attenuator. *Optics Express* 13(24), 9916–9921.
- Zia, R., J. Schuller, A. Chandran, and M. Brongersma (2006). Plasmonics: the next chip-scale technology. *Materials Today* 9(7-8), 20–27.
- Ziman, J. (1972). *Principles of the Theory of Solids* (2nd ed.). Cambridge: Cambridge University Press.
- Zimmermann, H. (2000). *Integrated Silicon Opto-Electronics*. Berlin: Springer-Verlag.

Index

- Abbe
 - diagram, 84
 - number, 71–72, 84, 85, 246
 - number, diffractive lens, 305
- Abbe, Ernst, 10, 71
- aberrations, lens, *see* lenses
- absorption, 77–81, 86
 - coefficient, 77–79, 401, 439, 458
 - in waveguides, 361–364
 - of photons, 397
 - spectroscopy, 457–461
- achromatic lenses, 246, 308, 514
- adaptive optics, xii, 189–193
- afocal, 229, 234
- Ag
 - for metamaterials, 559
 - for plasmonics, 555, 558
 - for reflection, 171–172
- air
 - refractive index, 26, 165
- Airy
 - disk, 187, 219, 220, 282–283
- Airy, George, 8, 10, 187
- Al
 - for plasmonics, 555
 - for reflection, 170–171
 - mole fraction, 90
- Al₂O₃, 158, 415
- AlGaAs, 90, 92, 351, 360, 407, 414
- AlGaInP, 90, 407, 408
 - laser characteristics, 417–418
- AlGaN, 415
- Alhazen (Ibn-al-Haitham), 4, 7, 9
- alignment
 - active, 492–493
 - passive, 493
- aluminum, *see* Al
- Ampère’s law, 20
- Ampère, André Marie, 20
- amplification
 - optical, 398–399
- anamorphic optics, 317, 361, 422
- angular frequency, 25
- anti-reflection coating, 154, 316, 356
- ants
 - fried, 213
- aperture, 222, 239
 - circular, 279–283
 - diffraction from, 266–270
 - dual slit, 274–275
 - multiple slits, 275–276
 - rectangular, 277–278
 - single slit, 270–272
 - waveguide, 361
- aplanatic lenses, 237
- apochromatic lenses, 246
- apodization, 274
- APS, *see* photodetectors
- AR, *see* anti-reflection coating
- Arago, Dominique Francois, 8
- arrayed waveguide grating, 369, 383, 385
- aspheric lenses, 237, 251, 258, 259, 482, 486
- assembly, optical, 491–497
 - active alignment, 492–493
 - flip-chip, 495–496
 - passive alignment, 493
 - Si optical bench, 496–497
 - V-grooves, 493–496
- astigmatism, 46, 240–241
 - in laser diodes, 422
 - in microlenses, 474
 - waveguide far-field, 361
- athermal lenses, 306–308
- attenuation
 - in waveguides, 361–364
 - optical, 77–82, 401
- Au
 - for contacts, 495
 - for electroplating, 483
 - for plasmonics, 555, 558
 - for reflection, 172
- avalanche photodiodes, 431
- AWG, *see* arrayed waveguide grating
- BaF₂, 373
- back focal length, 226, 258

- Bacon, Roger, 9
 ball lenses, 224, 257–258, 497
 Balmer, Johann, 6
 band tails, 80, 362
 bandgap energy, 88, 91–92
 values, 92
 bandpass filter, 163
 bandwidth, 434
 barium borate, 77
 Bartholinus, Erasmus, 8
 Bayer filter, 438
 BD, *see* Blu-ray disc
 beam
 expander, 234–235
 optical, *see* optical beam
 shaping, 310
 Beer's law, 78, 398, 458
 Beer, August, 78
 Bessel function, 186, 280, 373
 recurrence relation, 280
 Bessel, Friedrich Wilhelm, 186
 $\text{Bi}_3\text{Fe}_5\text{O}_{12}$, 450
 biaxial materials, 73
 binary compound semiconductor, 89
 binary optics, 12, 304, 477–480
 Biot, Jean-Baptist, 20
 birefringence, 8, 72–75, 86, 95
 BK7, *see* glass
 blazed gratings, 11, 297–300
 Blu-ray disc, 283, 309, 454–457, 518
 Bohr, Niels, 6, 60
 Boltzmann, Ludwig, 396
 Bordeaux, 259
 Bragg
 filters, *see* filters
 for photonic crystals, 542–545
 mirrors, *see* mirrors
 stack, 312
 Bragg, William Henry, 162
 Bragg, William Lawrence, 162
 bread
 with butter, 228
 Brewster
 angle, 111, 117
 window, 111
 Brewster, David, 8, 111, 259
 Brillouin scattering, 82
 Brillouin, Léon, 82
 broad-area laser, *see* lasers
 Bunsen, Robert, 6
 butt coupling
 waveguides, 356

 CaCO_3 , 73
 CaF_2 , 83, 85, 257
 camera-on-a-chip, 438

 candela
 definition, 410
 carrots
 novel uses, 451
 CCD, *see* photodetectors
 CD, 454–457, 482
 CdSe, 413
 chalcogenides, 83, 373
 charge density, 19
 chirp
 in gratings, 296
 in modulators, 442
 chocolate
 optical properties, 525
 chromatic aberration, 243–247
 axial, 244
 correction, 245–247
 diffractive lenses, 304–305
 lateral, 244
 circle of least confusion, 240, 244
 circular polarization, *see* polarization
 Clausius, Rudolf, 67
 Clausius-Mossotti equation, 67
 CMOS, 194, 385, 438
 coherence, 126–128
 definition, 126
 length, 126, 142
 time, 126
 collimation, 42, 52, 210, 233
 coma, 239–240, 243
 in microlenses, 474
 compound semiconductors, 89–93, 405
 concave
 definition, 175, 208
 lens behavior, 210
 conductivity, 58
 confusion, maximum, 240
 conjugates, 214, 260
 convex
 definition, 175, 208
 lens behavior, 210
 coupled mode analysis
 in photonic crystals, 542–545
 coupling
 efficiency, 356
 waveguides, 356–359
 critical angle, 115–116, 327, 330
 crown glass, 72
 cryolite, 155
 cutoff
 in a waveguide, 325
 CVD, 467

 DBR laser, *see* lasers
 decibel, 79
 depth of field, *see* depth of focus

- depth of focus, 48, 51, 230
- Descartes, René, 5, 7, 27
- detectors, *see* photodetectors
- DFB laser, *see* lasers
- diamond, 85
- dielectric
 - constant, 19, 57
 - films, 93–94
- dielectrophoresis, 506, 509
- diffraction
 - apertures
 - circular, 279–283
 - dual slit, 274–275
 - multiple slits, 275–276
 - rectangular, 277–278
 - single slit, 270–272
 - binary optics, 12, 304, 477–480
 - definition, 265
 - diffractive lenses, 300–309
 - aberrations, 304–306
 - achromatic, 308
 - athermal, 306–308
 - binary optics, 12, 304, 477–480
 - hybrids, 306–309
 - kinofoms, 309, 475, 477
 - diffractive optical elements, 309–311, 475
 - applications, 310–311
 - beam shaping, 310
 - design, 309–310
 - fan-outs, 311
 - efficiency, 287
 - far field, 267, 269
 - Fraunhofer regime, 267, 269
 - Fresnel
 - regime, 267, 300–301
 - zone plates, 304, 478
 - Fresnel-Kirchhoff formula, 268–270
 - gratings, 11, 283–300
 - blazed, 11, 295, 297–300
 - chirped, 296
 - coupled mode analysis, 542–545
 - couplers for waveguides, 359
 - Dammann, 311
 - DBR/DFB lasers, 420–422
 - diffraction pattern, 288
 - dispersion, 291, 300
 - efficiency, 294
 - fabrication, 480–481
 - free spectral range, 292–293
 - ghosts, 294
 - grating equation, 284, 289, 299
 - holographic, 295, 316, 480–481
 - lamellar, 295
 - Littrow mount, 299–300
 - profiles, 295–296
 - resolution, 317
 - ruled, 295
 - stray light, 293
 - surface emission, 428
 - tunable, 518–521
 - VCSELs, 422–425
 - Wood's anomalies, 294
 - kinofom, 309, 475, 477
 - limit, 219, 265, 282
 - microlenses, 251
 - near-field, 267
 - orders, 284
 - diffractive lenses, *see* diffraction
 - diffractive optical elements, *see* diffraction
- Digital Micromirror Device, *see* DMD
- diopter, 210, 507
- dipole moment, 63
- Dirac, Paul, 6, 397
- direct bandgap semiconductor, 87, 88
- dispersion
 - diffractive microlenses, 304
 - fiber, 376–380
 - grating, 291, 300
 - material, 68, 71, 244
- displays
 - optofluidic, 534
- distortion, 242
- distributed feedback lasers, *see* lasers
- DMD, 14–15, 193–196
- DOE, *see* diffractive optical elements
- DOF, *see* depth of focus
- Dollond, John, 10
- Doppler shift, 127
- Drude model, 65
- Drude, Paul, 65
- Ducasse, Alain, 56
- DVD, 283, 309, 454–457, 482
- e-beam, 475, 545
- echelette gratings, *see* blazed gratings
- Edison, Thomas, 10
- effective
 - focal length, 226, 258
 - index, 313, 331–332
- eigenvalue equation, 47, 120, 337
- Einstein, Albert, 6
- elastomers, 224, 485, 487
- electric
 - field, 18
 - flux density, 18, 72
- electro-optic effects, 76, 440–442, 522
- electroabsorption, 441
- electroactive polymers, 519
- electromagnetic
 - field
 - energy density, 32
 - power density, 33

- model for waveguides, 337–350
- waves, 30–36
 - boundary conditions, 106–107, 151–152
 - one-dimensional, 27
 - polarization, 36–40, 107
 - spectrum, 34–36
 - three-dimensional, 40
 - transverse, 27–28
- electrorefraction, 440
- electrowetting, 504–510
 - displays, 534
 - on dielectrics, 506
 - opto-electrowetting, 509
- elliptical polarization, *see* polarization
- end-fire coupling
 - waveguides, 357
- energy
 - bands, 87–89, 542–545
 - conversion to wavelength, 35
 - in electromagnetic fields, 32–34
- epitaxial crystal growth, 91
- erbium-doped fiber amplifiers, 429
- etalon, *see* Fabry-Perot
- Euclid, 4, 7
- Euler, Leonhard, 147
- evanescent field, 118–122, 341–344
- external cavity laser, *see* lasers
- external incidence, 109, 111, 115
- extraordinary axis, 73
- f-number, *see* lenses
- fabrication
 - binary optics, 477–480
 - continuous-profile optics, 475–477
 - direct-write lithography, 475
 - e-beam, 475, 545
 - electroplating, 483
 - embossing, 259, 484–485
 - etching, 467
 - flip-chip, 495–496
 - glass molding, 486
 - gray-scale lithography, 476–477
 - hybrid integration, 491–492
 - injection molding, 259, 484
 - interference lithography, 12, 295, 480–481, 547
 - ion exchange, 489–491
 - layer deposition, 467
 - LIGA, 483
 - lithography, 12, 467, 474–481
 - microcontact printing, 253, 472–474
 - MOEMS, 488–489
 - monolithic integration, 491–492
 - nano-imprint lithography, 485, 528
 - near-field holography, 481
 - optical assembly, 491–497
 - PDMS, 485, 487, 510, 511
 - photonic crystals, 545–548
 - photoresist reflow, 12, 469–472
 - Pyrex, 85, 135, 197, 221, 467, 508, 510
 - replication, 482–487
 - self-assembly, 468–474, 545–548, 552
 - semiconductor processes, 466–468
 - Si optical bench, 496–497
 - soft lithography, 472, 528, 530
 - sol-gels, 486–487, 553
 - SU-8, 193, 486, 511
 - UV casting, 485
 - V-grooves, 493–496
- Fabry, Georges, 145
- Fabry-Perot
 - etalon, 145
 - interferometer, *see* interferometers
 - laser, *see* lasers
 - resonances
 - for waveguide loss measurement, 363
 - laser, 400–403
 - parasitic, 357
- fan-outs, 311
- far-field
 - broad-area lasers, 419
 - diffraction, 267, 269
 - edge-emitting lasers, 421
 - mirror, 185
 - VCSELs, 424
 - waveguide, 361
- Faraday effect, 97, 449
- Faraday's law, 21
- Faraday, Michael, 5, 21
- Fermat's principle, 102–103, 204, 206
- Fermat, Pierre de, 5, 7, 102
- Fermi, Enrico, 397
- Fermi-Dirac distribution, 397
- fiber, optical, 13, 369–380
 - attenuation, 375–376
 - Bragg gratings (FBG), 391, 518
 - dispersion, 376–380
 - dispersion-flattened, 379
 - dispersion-shifted, 379
 - field distribution, 373–374
 - graded-index, 370
 - gyroscope, 144
 - hollow-core, 371, 553
 - modes, 374
 - multimode, 370
 - photonic crystal, 371, 553
 - plastic, 372, 386
 - polarization-maintaining, 371
 - single-mode, 370
- Fick, Adolf Eugen, 10
- field
 - curvature, 241–242
- fill factor, 253

- filters
 - Bragg, 163–164, 519–521
 - optofluidic, 533–534
 - tunable, 163–164, 519–521
- finesse, 147
- Fizeau interferometer, *see* interferometers
- Fizeau, Armand Hippolyte, 8, 140
- flint glass, 72
- flip-chip, 495–496
- focal length, 208–211, 224
 - thick lenses, 226
- focal plane, 215
- FOG, *see* gyroscope
- Foucault, Jean, 8
- Fourier transform, 127, 185, 270, 273–274, 279, 288, 310
- Fourier, Jean Baptiste Joseph, 5, 127
- Franklin, Benjamin, 10
- Franz, Walter, 441
- Franz-Keldysh effect, 441, 442, 446
- Fraunhofer
 - diffraction, 185, 267
 - doublet, 246, 247, 308, 514
 - lines, 71, 244, 247
 - regime, 269
- Fraunhofer, Joseph von, 6, 8, 11, 185, 284
- free spectral range
 - Fabry-Perot interferometer, 148
 - grating, 292–293
- free-electron absorption, 81
- free-space optics, 321
- Freiburg im Breisgau, 82, 141
- frequency
 - conversion to wavelength, 35
 - doubling, *see* second harmonic generation
- Fresnel
 - diffraction, 267, 300–301
 - equations, 107–109
 - lenses, 12, 258–260, 301–304, 475, 477
 - fabrication, 483, 485
 - tunable, 517, 522
 - reflection, 107–109, 408
 - zone plates, 12, 304, 478, 517, 522
- Fresnel, Jean Augustin, 5, 8, 108, 258
- Fresnel-Kirchhoff formula, 268–270
- fringe
 - definition, 134, 138
- fused silica, 467
- GaAs, 56, 87, 89, 116, 388
 - bandgap, 88
 - electro-optic effect, 76, 440–443
 - for lasers, 404, 415, 423
 - for LEDs, 407, 408
 - for photodetectors, 431
 - gain in, 399
 - heterostructures, 414
- Gabor, Dennis, 11
- gain
 - optical, 399–400
- Galilean telescope, 229
- Galilei, Galileo, 9, 229
- gallium arsenide, *see* GaAs
- GaN, 90, 407, 415
- GaP, 408, 549
- GaSb, 317
- Gauss' law, 19
- Gauss, Johann Carl Friedrich, 10, 19
- Gaussian beam, 46–52, 210
 - collimation, 233
 - definition, 48
 - depth of focus, 48, 51, 230
 - expansion, 234–235
 - focussing, 232–233
 - from fiber, 374
 - intensity, 50
 - magnification, 231
 - radius of curvature, 48, 49
 - through lenses, 230–235
 - waist, 48, 49
 - width, 48, 230
- Gaussian lens formula, 211, 227
- Ge
 - for photodetectors, 431
- geometric optics
 - definition, 203
- GGG (gadolinium gallium garnet), 98, 450
- ghosts, 294
- gimbals (for micromirrors), 183, 198
- glass, 73, 83–87
 - BK7, 71, 84, 109, 110, 113, 117, 121, 154, 156, 158, 246, 257, 343, 348, 352, 358, 467, 470, 479
 - chalcogenide, 83, 373
 - crown, 72, 246
 - flint, 72, 246
 - fluorozirconate, 373
 - for optical fiber, 372
 - halide, 83
 - HEBS, 476
 - optical absorption, 79, 376
 - Pyrex, 85, 135, 197, 220, 221, 467, 508
 - SF11, 307
 - SF59, 450
 - SF6, 85, 358
 - sputtered, 85
 - thermal expansion, 85, 86
 - ULE, 86
 - Zerodur, 86
- gold, *see* Au
- Goos, Fritz, 347
- Gouy phase shift, 48, 49

- Gouy, Louis Georges, 48
 Granger, Hermione, 559
 grating
 coupling
 waveguides, 359
 light valve, 311
 gratings, *see* diffraction
 Green, Alfred, 139
 Grimaldi, Francesco Maria, 5, 7
 GRIN lenses, 255–257, 491, 517
 group velocity, 30, 349
 gyroscope
 integrated optical, 388
 optical fiber (FOG), 144
- Hänchen, Hilda (Lindberg-), 347
 half-wave plate, 75
 Hall, Chester Moor, 10
 handedness, *see* polarization
 Hartmann, Johannes, 191
 HEBS glass, 476
 Helmholtz equation, 46, 120, 337
 paraxial, 47
 Helmholtz, Hermann von, 6, 46
 Hertz, Heinrich, 6, 25
 heterostructure, 414, 431, 435
 HgCdTe, 90, 433
 holograms, 12, 309, 311, 453, 480
 holography, 11
 near-field, 481
 Hook, Robert, 5, 7, 9
 Huygens' principle, 266–268
 Huygens, Christiaan, 5, 8, 9, 266
 hydrogels, 522
- IBM, 468
 II-VI semiconductors, 90, 416, 431
 III-V semiconductors, 89–93, 466
 bandgap, 91–92, 405
 cleaving, 149
 for optical telecommunications, 383, 385, 491
 for waveguides, 351
 light emission, 405
 monolithic integration, 491
 image
 real, 173, 180, 213
 virtual, 173, 180, 213
 immersion optics, 224
 impedance, 32, 108, 151
 for optical multilayers, 159–164
 implanted waveguide, 354
 incoherent light sources, 126
 indirect bandgap semiconductor, 88, 405
 indium tin oxide, *see* ITO, 508, 517
 InGaAs, 90, 385
 for photodetectors, 431
 InGaAsP, 407, 415, 421, 428, 451, 549
 InGaN, 90, 407, 415
 InP, 76, 90, 382, 383, 385, 415
 integrated optics, 13
 applications, 380
 definition, 321, 380
 for sensors, 386–391
 for telecommunications, 386
 intensity
 modulation, *see* modulators
 optical, 78
 interference, 128–135
 constructive, 131
 destructive, 131
 filters, 160–164
 intensities, 131–134
 lithography, 480–481, 547
 path length difference, 133–134
 superposition of fields, 129–130
 visibility, 134
 wavelength difference, 131–132
 interferometers, 136–150
 Fabry-Perot, 145–150, 400–403, 552
 Fizeau, 140
 Mach-Zehnder, 141–142, 253, 382, 390, 446
 Michelson, 137–139, 388
 Sagnac, 143–145, 200, 388
 Twyman-Green, 139, 253
 Young, 136–137, 390
 internal incidence, 115–122
 inverse centimeters, 36
 ion-exchange waveguide, 354
 isolator
 optical, 97, 382, 448–451
 isotropic materials, 73
 ITO, 411, 413
- Jones matrices, 40
 Jones, R. Clark, 40
- kaleidoscope, 8
 KDP (KH_2PO_4), 73, 76, 77, 443
 Keats, John, 7
 Keldysh, Leonid Veniaminovich, 441
 Kepler, Johannes, 7, 9, 229
 Keplerian telescope, 229
 Kerr effect, 76, 440, 441
 Kerr, John, 76
 kinoform, 12, 309, 475, 477
 Kirchhoff, Gustav, 6, 269
 KNbO_3 , 443
 Kramers, Hendrik Anthony, 60
 Kramers-Kronig relation, 60, 68, 441
 Kronecker, Leopold, 44
 Kronig, Ralph, 60
 KTP (KTiOPO_4), 416, 443

- Lambert, Johann Heinrich, 78
lamellar gratings, 295
Laplace, Pierre-Simon, 5
lasers
 active region, 414
 Ar, 36, 480
 broad-area, 418–419
 characteristics of
 DFB, 421
 Fabry-Perot, 417–418
 VCSEL, 424
 DBR, 312, 383, 388, 419–422
 DFB, 313, 377, 382, 420–422, 427, 460, 523, 529–531
 diode, 12, 33, 56, 90, 92, 233, 238, 315, 317, 413–428
 disk, 427–428, 538
 distributed feedback, 419–422
 dye, 527–531
 external cavity, 315–317
 VCSEL, 425
 Fabry-Perot, 383, 416–419, 527–529
 fiber, 425
 HeCd, 52, 475, 480
 HeNe, 26, 75, 111, 128, 131, 154, 156, 187, 267, 289, 290, 332, 390
 lead-salt, 460
 materials, 404–405
 mode-hops, 313
 monomode emission, 420, 528
 multimode emission, 418
 Nd:YAG, 77, 97, 310, 481, 523, 528, 547
 optofluidic, 527–531
 pointers, 416
 principles of, 396–405
 quantum cascade, 426–427
 Raman, 429
 resonance, 400–404
 ring, 145
 Si, 429
 single mode, 313
 spaser, 557
 threshold, 401
 threshold current, 417
 tunable, 421, 523
 VCSEL, 163, 313–315, 422–425, 458
 polarization control, 424
 tunable, 425
 using photonic crystals, 550
LaSFN9, 257
lateral
 definition, 352
lattice constant, 91–92
LCD, *see* liquid crystals, displays
LEDs, 56, 178, 405–413
 as a point source, 42
 Burrus structure, 408
 emission efficiency, 407–409
 emission from substrate, 116
 materials, 404–405
 OLEDs, 411–413
 power emission, 410
 quantum dots, 413, 538
 resonant cavity, 411
 spectrum, 409–410
 structures, 406–408
 superluminescent (SLED), 411
 using photonic crystals, 549
 using plasmons, 557
Leeuwenhoek, Anton van, 9
lenses
 aberrations, 235–247
 astigmatism, 240–241
 chromatic, 243–247
 coma, 239–240, 243
 diffractive, 304–306
 distortion, 242
 field curvature, 241–242
 spherical, 237–238
 achromatic, 246, 308, 514
 aperture, 222
 aplanatic, 237
 apochromatic, 246
 aspheric, 237, 251, 258, 259, 482, 486
 athermal, 306–308
 ball lenses, 224, 257–258, 497
 chromatic aberration, 243–247, 304–305
 combinations of, 228–230
 concave, 204, 208, 210
 convex, 204, 208, 210
 diffraction limited, 219, 265
 diffractive, 300–309
 aberrations, 304–306
 achromatic, 308
 athermal, 306–308
 binary optics, 12, 304, 477–480
 hybrids, 306–309
 kinofoms, 309, 475, 477
 diopter, 210
 elliptical, 205
 f/#, 224–225, 233
 focal length, 208–211, 224
 Fresnel lenses, 258–260, 483, 485
 Fresnel zone plates, 304, 478
 Fresnel-like lenses, 301–304, 475, 477
 Gaussian beams, 230–235
 GRIN lenses, 255–257, 491, 517
 hyperbolic, 205
 liquid, 504–510, 522, 531–533
 liquid crystal, 453, 517–518
 magnification, 217–218
 membrane, 510–516

- microlenses, 250–254
 - arrays, 253
 - profile measurement, 140
 - scaling, 250
 - sizes, 251
- modulation transfer function, 220
- multiple, 228–230
- numerical aperture, 222–224
- principle planes, 225, 228
- resolution, 220
- Seidel aberrations, 235–243
- sign conventions, 208
- spherical, 205
- spot diagram, 219
- Strehl ratio, 219, 282
- thick, 225–227
- thin, 209–212
- tunable
 - liquid, 504–510, 522, 531–533
 - liquid crystal, 517–518
 - membrane, 510–516
- zoom, 229
- Lensmaker's formula, 209
- LiF, 83
- liftoff, 468
- LIGA, 483
- light bulb, 33
- light-emitting diode, *see* LEDs
- light-guide, 322
- LiNbO₃, 56, 73, 76, 97, 352, 382, 443, 522
- line spread function, 219
- linear polarization, *see* polarization
- linewidth, 127, 399
- Lippersley, Hans, 9
- Lippmann, Gabriel, 505
- liquid crystals, 451–453
 - displays, 452
 - lenses, 453, 517–518
 - nematic, 451
 - on silicon, 453
 - switching, 452
- liquid lenses, 504–510, 522, 531–533
- Lister, Joseph Jackson, 10
- LiTaO₃, 443, 522
- lithium niobate, *see* LiNbO₃
- lithium triborate, 77
- lithography, *see* fabrication
- Littrow mount, 299–300
- Littrow, Joseph Johann von, 299
- loss
 - optical, *see* absorption
- love songs, 296
- LPCVD, 467
- LSF, *see* line spread function
- lumen
 - definition, 410
- Mach, Ludwig, 141
- Mach-Zehnder interferometer, *see* interferometers
- magnetic
 - field, 18
 - flux density, 18
- magneto-optics, 97–98, 449–451
 - data storage, 455
- magnification
 - Gaussian beam, 231
 - lenses, 217–218
 - mirrors, 179
- Maiman, Theodore, 10
- Malus, Etienne Louis, 8
- marginal rays, 239
- mass and spring model, 62
- material usage, 55–56
- materials
 - ceramics, 493
 - compound semiconductors, 89–93
 - dielectrics, 93–94, 467
 - elastomers, 472
 - electroactive polymers, 519
 - glasses, 83–87, 467, 489, 527
 - hydrogels, 522
 - LiNbO₃, 97, 489
 - liquid crystals, 451–453, 517–518
 - metals, 467
 - PDMS, 252, 485, 487, 510, 511, 528
 - polymers, 94–96, 484
 - Pyrex, 85, 135, 197, 221, 467, 508, 510
 - semiconductors, 87, 93
 - silicon, 89, 467, 488, 493–497
 - sol-gels, 96–97, 486–487
 - SU-8, 193, 486, 511, 527, 547
 - Teflon, 507
- Maxwell's equations, 19–23
- Maxwell, James Clerk, 5, 17, 23
- Maxwell-Boltzmann distribution, 396
- mean square value, 434
- measurement
 - displacement, 139
 - refractive index, 139, 142
 - surface deformation, 140
- MEH-PPV, 411
- membrane lenses, 510–516
- MEMS, xi, 13, 14, 145, 164, 182, 199, 253, 315, 465, 488–489
 - fabrication, 466–468
 - MOEMS, 13, 488–489
 - tunable lenses, 507–510
- meridional plane, 240
- metals
 - for laser diodes, 417
 - for mirrors, 170–172
 - reflectivity, 66
- metamaterials, 558–559

- MgAl₂O₄, 415
MgF₂, 83, 154–156, 158, 171, 559
Michelson interferometer, *see* interferometers
Michelson, Albert, 6, 8, 11, 137, 284
micro-electromechanical systems, *see* MEMS
micro-total-analysis systems (μ TAS), 525
microcontact printing, 253
microlenses, *see* lenses
micromirrors, *see* mirrors
microsystems, xi–xii
Mie scattering, 82
Mie, Gustav, 82
mirrors
 Bragg, 160–164, 419–425, 519–521
 cleaning, 170
 imaging, 179–180
 magnification, 179
 metals for, 170–172
 micromirrors, 180–189
 adaptive, 189–193
 aperture, 185
 applications, 193–199
 deformable, 192–193
 diffraction from, 185–189
 optical properties, 185–189
 parabolic, 176
 planarity, 185
 pneumatic, 515
 resolution, 189
 roughness, 185
 scanning, 182–185
 structure, 181–182
 mirror equation, 177, 178
 non-planar, 175–180
 parabolic, 175–176
 planar, 172–175
 imaging, 173
 rotating, 173–175
 spherical, 176–180
 surfaces, 170
 tunable, 163–164
MMI, *see* multimode interference coupler
modulation transfer function, 220–221, 512
modulators, 76, 439–448
 bulk, 444–445
 chirp, 442
 electro-optic, 142, 440–442
 Franz-Keldysh effect, 441, 442, 446
 integrated optical, 445–448
 intensity, 440
 Kerr effect, 441
 phase, 440
 plasma effect, 442
 Pockels effect, 440
 quantum well, 442
 quantum-confined Stark effect, 442, 446
MOEMS, *see* MEMS
monomode emission, lasers, 420, 528
Morley, Edward, 6
Mossotti, Ottaviano-Fabrizio, 67
MSM photodetector, 431
MTF, *see* modulation transfer function
Mulhouse, 78
multilayers
 dual-layer, 157–159
 half-wave, 155, 160
 impedance, 159–164
 materials, 155
 mirrors, 160–164
 optical, 150–164
 quarter-wave, 153–155, 159–160
multimode emission, lasers, 418
multimode interference coupler, 368, 385
multiple-lens systems, 228–230
NA, *see* numerical aperture
Na₃AlF₆, 155
nano-optics
 definition, 537
 metamaterials, 558–559
 photonic crystals, 539–553
 plasmonics, 554–558
 sub-wavelength optics, 486, 537, 538, 554, 558
nanophotonics, 537–539
nanorods, 424, 555
nanoshells, 558
nanospheres, 553, 555, 557
nanotubes, 539
nanowires, 539, 556–557
near-field
 diffraction, 267, 300–301
 holography, 481
 optics, 538
 waveguide, 360
Newton, Isaac, 5, 7, 10
NiGeAu, 417
noise-equivalent power, 434
non-linear materials, 75–77, 440–442
notation, 17
numerical aperture, 222–224
 in a waveguide or fiber, 358
numerical simulation
 ZEMAX, 220–221, 243, 246–247
OCT, *see* optical, coherence tomography
OEIC, *see* opto-electronic integrated circuits
OLED, *see* LEDs
olives
 may have pits, 451
on-axis rays, 237
opals, 541
OPD, *see* optical, path difference

- OPL, *see* optical, path length
- optical
- amplification, 398–399
 - assembly, 491–497
 - coherence tomography, 497
 - cross-connect, 198–199, 489
 - crossbars, 198
 - data storage, 454–457
 - fiber, *see* fiber, optical
 - isolator, 97, 382, 448–451
 - multilayers, *see* multilayers
 - path difference, 219
 - path length, 102, 206, 219
 - resolution, 282–283
 - sensors, 386–391, 457–461, 518
 - switching, 198–199, 311, 489
 - telecommunications, xi, 142, 372, 386, 539
 - thickness, 74
 - transitions, 396–400
 - tweezers, 536
- optical beam
- collimated, 42, 51, 52, 210
 - definition of, 42, 49
 - expander, 234–235
 - Gaussian, 46–52, 210
 - marginal rays, 239
 - on-axis, 237
 - principle rays, 239
 - skew rays, 219, 237, 240
- opto-electronic integrated circuits, 380
- opto-electrowetting, 509
- opto-thermal expansion coefficient, 306
- optofluidics, 525–536
- applications, 526
 - displays, 534
 - filters, 533–534
 - lasers, 527–531
 - microlenses, 504–516, 531–533
 - microscopy, 535–536
 - photonic crystals, 535
 - waveguides, 526–527
- ordinary axis, 73
- ormocer, 96
- overlap integral, 356
- OXC, *see* optical, cross-connect
- p-i-n junction
- in a photodiode, 430
- Pérot, Alfred, 145
- paper
- electronic, 534
- paraboloidal wavefront, *see* wavefront
- parametric amplification, 76
- paraxial, 46, 177
- approximation, 207, 215, 235
 - wave equation, 255
- Passard, Alain, 55
- PDMS, 252, 485, 487, 510, 511, 545
- PDOT, 412
- PECVD, 467
- PEDOT, 412
- permeability, 18
- negative, 558
- permittivity, 18, 57
- direction dependence, 72–75
 - negative, 558
 - of free space, 18
- Petzval, József Miksa, 241
- PFPE, 531
- phase modulation, *see* modulators
- phase shift
- total internal reflection, 117–118, 328
 - upon reflection, 110–111
- phase velocity, 30
- photodetectors, 430–436
- APS, 438
 - arrays, 436–438
 - avalanche photodiodes, 431
 - CCD, 436–438
 - dark current, 433
 - efficiency, 432–433
 - materials, 93
 - MSM, 431
 - noise, 433–435
 - noise-equivalent power, 434
 - organic, 435–436
 - photodiodes, 430–435
 - quantum dot, 436
 - responsivity, 432
 - signal-to-noise ratio, 434
- photon
- origin of name, 6
- photonic crystals, 539–553
- band structure, 542–545
 - coupled mode analysis, 542–545
 - fabrication, 481, 545–548
 - fiber, 371, 553
 - for improving LED efficiency, 408, 549
 - optofluidic, 535
 - structures, 539–541
 - tunable, 521–522, 535, 552–553
 - waveguides, 354, 548–549
 - woodpile structures, 547
- photonic integrated circuits, 380
- photonics
- definition, 321, 380
 - nanophotonics, 537–539
 - silicon, 354, 384–386, 428–429, 491, 496
- photoresist
- reflow lenses, 469–472
 - refractive index, 221, 252, 470
- PIC, *see* photonic integrated circuits

- pigeons
 - roasted, 52
- pitch
 - of a GRIN lens, 255
- planar lightwave circuits, 384, 491, 496
- planar wavefront, *see* wavefront
- Planck, Max, 6, 35
- plasma
 - effect, 442
 - frequency, 66, 171
- plasmonics, 554–558
 - for lasers, 424
 - light sources, 557
 - plasmons, 554–555
 - sensors, 557–558
 - waveguides, 555–557
- plastic optical fiber, *see* fiber, optical
- PLC, *see* planar lightwave circuits
- PLZT, 522
- PMMA, 95, 121, 227, 259, 260, 351, 372, 475, 484, 545
- pneumatic micro-optics, 510–516
- Pockels effect, 76, 440
- Pockels, Friedrich, 76
- POF, *see* fiber, optical, plastic
- Poincaré sphere, 39
- Poincaré, Jules Henri, 6, 39
- point source, 42, 266, 268
- point spread function, 219
- polarizability, 18, 57, 63, 66, 72
- polarization, 36–40
 - angle, 111, 114, 117
 - circular, 37, 39, 74
 - ellipse, 39
 - elliptical, 37
 - handedness, 37, 39
 - linear, 36, 39, 74
 - retarder, 73
 - rotation, 75, 449–450
- poly methyl methacrylate, *see* PMMA
- polycarbonate, 95, 351, 484
- polyimide, 351, 451
- polymers, 94–96
 - for LEDs, 411–413
 - for optical fiber, 372
 - for photodetectors, 435–436
 - for replication, 482–485
 - for waveguides, 351
- polysilicon, 192
- polystyrene, 95, 484
- polyvinyl chloride, *see* PVC
- population inversion, 399, 404
- position-sensitive detector, 197, 432
- Potter, Harry, 559
- power
 - density, 33
 - optical, 32, 78
 - reflection and transmission, 113–115
- Poynting vector, 32, 151
 - Gaussian beam, 50
 - time average, 33
- Poynting, John Henry, 32
- PPV, 411
- primary aberrations, 235–243
- principle
 - of least time, 102
 - planes, 225, 228
 - rays, 239
- propagation constant, 26, 439
 - complex, 58–59, 439
 - in a medium, 31
 - relation to wave vector, 26
- propagation vector, *see* wave vector
- PSF, *see* point spread function
- Ptolemy, 7
- PVC, 95, 343, 348, 358, 484
- Pyrex, 85, 135, 197, 221, 467, 508, 510
- Q (quality factor), 369, 551
- QCL, *see* lasers, quantum cascade
- quantum dots, 413, 429, 436, 538, 539
- quantum well, 415, 429, 442
 - intermixing, 383
- quantum-confined Stark effect, 442, 446
- quarter wave
 - film, 153–155
 - plate, 74
- quartz, 73, 83
- quaternary compound semiconductor, 90
- Rømer, Ole, 8
- Raman
 - laser, 429
 - scattering, 82, 557
- Raman, Chandrasekhara Venkata, 82
- ray tracing
 - by hand, 213–218
 - numerically, 218–221, 243, 246
- ray-optic model for waveguides, 322–337
- Rayleigh
 - limit, 219, 282
 - range, 49
 - scattering, 82, 86, 376
- Rayleigh, Lord, 11, 49
- real image, *see* image
- reciprocal optical systems, 359
- reflectance, 113
 - multilayer, 153
 - normal incidence, 114
- reflection, 103–104, 169–172
 - external incidence, 109, 111, 115
 - internal incidence, 115–122

- normal incidence, 114
- specular, 172, 286
- total internal, 116–118
- reflectivity, 109
 - metals, 66, 170–172
 - multilayer, 152–153
 - phase shift, 110–111
- refraction, 104–106
 - at a curved surface, 204–205
 - at a spherical surface, 205–208
 - negative, 558–559
- refractive index, 56–77
 - complex, 59–60, 401
 - definition, 31, 57
 - direction dependence, 72–75
 - extraordinary, 73, 97, 451
 - measurement of, 139
 - negative, 558
 - of air, 26, 165
 - of photoresist, 221, 252, 470
 - of water, 110, 165
 - ordinary, 73, 97, 451
 - temperature variation, 72, 442
 - values, 92
 - wavelength variation, 68–69
- replication, *see* fabrication
- resolution
 - grating, 317
 - lens, 220
 - optical, definition of, 282–283
 - spectrometer, 317–318
- resonance
 - atomic, 64
 - laser, 400–404
 - whispering gallery, 427
- responsivity, 432
- Reynolds number, 526
- Reynolds, Osborne, 526
- rhodamine, 528
- rib waveguide, 352
- RIBE, 468
- ridge waveguide, 352
- RIE, 468
- ring resonator, 369, 384, 387, 390, 427–428, 533, 538
- Rittenhouse, David, 11
- Rowland, Henry Augustus, 11
- sag height, 251
- sagittal plane, 240
- Sagnac interferometer, *see* interferometers
- Sagnac, Georges, 143
- salad
 - tossed, 419
- sapphire, 85, 257, 415
- Savart, Felix, 20
- sawtooth gratings, *see* blazed gratings
- scattering, 81–82
- Schottky, Walter, 431
- second harmonic generation, 76–77
- Seidel aberrations, 235–243
- Seidel, Philipp Ludwig von, 237
- Sellmeier formula, 69–70
- Sellmeier, W., 69
- semiconductor
 - absorption, 80
 - gain, 399–400
 - materials, 87–93
 - optical amplifier, 383
 - optical transitions, 395–400
 - organic, 411
 - photon absorption, 397
 - photon emission, 397
- sensors
 - μ TAS, 525
 - absorption spectroscopy, 457–461
 - chemical, 142, 388–391, 526
 - displacement, 388
 - fiber, 391, 518
 - integrated optical, 386–391
 - nanophotonic, 539
 - rotation, 144, 388
 - surface plasmon resonance, 389, 526, 554
 - TDLAS, 457–461
 - torque, 387
 - Young interferometer, 137, 390
- SERS, *see* surface-enhanced Raman spectroscopy
- SF11, *see* glass
- SF6, *see* glass
- Shack, Ronald, 191
- Shack-Hartmann sensor, 191
- sharks
 - hungry, 372
- SHG, *see* second harmonic generation
- Si, 56, 73, 89, 466, 467
 - amorphous, 520
 - as lens substrate, 470
 - bandgap, 88, 405
 - electro-optic effect, 76, 443
 - for photodetectors, 431, 436
 - for waveguides, 351
 - laser, 429
 - light emission, 428–429
 - optical bench, 384, 496–497
 - photonics, 354, 384–386, 428–429, 491, 496
 - plasma effect, 442
 - thermo-optic effect, 442
- Si₂O₃, 171
- Si₃N₄, 520
 - waveguide, 332, 333, 335, 340, 342, 344, 351
- SiC, 415
- side-mode suppression ratio, 421

- SiGe, 429
- sign conventions, 29
- signal-to-noise ratio, 434
- silica
 - fused, 467
- silicon, *see* Si
- silicon dioxide, *see* SiO₂
- silver, *see* Ag
- SiN, 192
- sinc function, 271
- SiO₂, 73, 83, 155, 428, 451, 467, 507, 508, 520
 - absorption, 79, 376
 - dispersion, 68, 70
 - waveguide, 332, 333, 335, 340, 342, 344, 351
- SiOB, *see* Si, optical bench
- skew rays, 219, 237, 240
- SLED, *see* LEDs
- Snell's law, 104, 204, 255
- Snell, Willebrord, 7, 104
- SOA, *see* semiconductor, optical amplifier
- SOI (silicon on insulator), 351, 545
- sol-gels, 96–97, 486–487
- spaser, 557
- speckles, 135
- spectrometers, 317–319
 - Fourier transform, 497
 - resolution, 317–318
- spectrum
 - electromagnetic, 34–36
 - supercontinuum, 538, 553
 - width (FWHM), 149
- specular reflection, 172, 286
- speed of light
 - from Maxwell's equations, 5, 25
 - historical, 6, 8–9
 - in a medium, 31, 102
- spherical
 - aberration, 237–238
 - in microlenses, 474
 - wavefront, *see* wavefront
- spin-on glass, 96
- spinel, 415
- spontaneous emission, 397
- spot diagram, 219
- SPR, *see* surface plasmon resonance
- Stark, Johannes, 442
- Steinheil achromatic doublet, 246
- step function, 273
- stimulated emission, 397
- stop-band, 163, 519
 - in VCSELs, 423
- stops, 239
- stray light, 293
- Strehl ratio, 219, 282
- Strehl, Karl, 219
- strip-loaded waveguide, 353
- Strutt, John William, 11, 49
- SU-8, 193, 486, 511, 527, 547
- sub-wavelength optics, *see* nano-optics
- supercontinuum spectrum, 538, 553
- surface plasmon resonance, 389, 526, 554–555
- surface-enhanced Raman spectroscopy, 557
- susceptibility, 61–67
 - atomic, 62–64
 - definition, 57
 - electronic, 65–66
 - non-linear, 75–77
 - plot, 67
- Ta₂O₅, 155, 390, 451
- tangential plane, 218, 240
- TDLAS, *see* sensors
- TE polarization, 73, 107
- Teflon, 507
- telescope, 229–230
- TEM, *see* transverse electromagnetic
- ternary compound semiconductor, 90
- thermal effects
 - athermal lenses, 306–308
 - in glass, 72
 - thermal coefficient of expansion, 72, 85, 86, 306
 - thermo-optic effect in glass, 72
 - thermo-optic effect in Si, 442, 520
- thick lenses, 225–227
- thin lens approximation, 209–212
- threshold current, for lasers, 417
- TiO₂, 155
- TiPtAu, 417
- TM polarization, 73, 107, 294
- total internal reflection, 116–118, 260
 - in a waveguide, 327, 347
 - in an LED, 408
 - phase shift, 117–118, 328
- transmittance, 113, 273
 - normal incidence, 114
- transmittivity, 109
- transverse
 - definition, 352
 - electromagnetic, 27
- tunable micro-optics
 - diffractive optics, 518
 - electrowetting, 504–510
 - Fresnel lenses, 517, 522
 - gratings, 518–521
 - liquid crystal, 517–518
 - liquid lenses, 504–510, 522, 531–533
 - membrane lenses, 510–516
 - photonic crystals, 521–522, 552–553
 - pneumatic, 510–516
- Twyman, Frank, 139
- Twyman-Green interferometer, *see* interferometers

- ULE glass, 86
- uniaxial materials, 73, 74, 451
- unit vector, 18
- Urbach tails, 80, 362

- V-grooves, 493–496, 507
- variable optical attenuator, 383, 489
- VCL, *see* VCSEL
- VCSEL, *see* lasers
- vector, 17
- velocity
 - group, 30, 349
 - modal, in a waveguide, 333, 348–350
 - phase, 30
- Verdet, Émile, 449
- vertical cavity lasers, *see* VCSEL
- virtual image, *see* image
- visibility
 - interference, 134
- visible wavelength range, 35, 293
- VOA, *see* variable optical attenuator

- water
 - refractive index, 110, 165
- wave
 - equation, 23–25
 - cylindrical coordinates, 373
 - eigenvalue form, 47, 120, 337
 - general solution, 25, 58
 - in a waveguide, 337–338
 - one-dimensional, 27
 - paraxial, 255
 - three-dimensional, 24
 - vector, 25, 87
 - in a medium, 31
 - in a waveguide, 323
 - relation to propagation constant, 26
 - velocity
 - from Maxwell's equations, 25
 - in a medium, 30–31, 102
- wavefront, 40–46, 103
 - correction, 190–193
 - definition, 41
 - Huygens' principle, 266–267
 - in a waveguide, 323
 - paraboloidal, 47
 - planar, 41, 48, 49
 - spherical, 42, 48, 49
- waveguides
 - arrayed waveguide grating, 369
 - asymmetric, 329–330, 335–337
 - channel, 350–354
 - implanted, 354
 - ion-exchange, 354
 - rib, 352
 - ridge, 352
 - strip-loaded, 353
 - characteristic equation, 324, 327–330, 346–347
 - characterization, 355–364
 - cladding, 326, 327
 - confinement factor, 344–345
 - core, 326, 327
 - coupling light in and out, 356–359
 - butt coupling, 356
 - end-fire coupling, 357
 - grating coupling, 359
 - curves, 364–365, 549
 - cutoff, 325
 - dielectric, 326–337
 - dimensions, 354
 - discrete propagation in, 324–326
 - effective index, 313, 331–332, 335
 - electric field, 339–345
 - electromagnetic model, 337–350
 - evanescent field, 341–344, 347
 - far-field, 361
 - fluidic, 526–527
 - Goos-Hänchen shift, 347–348
 - III-V, 351
 - isolators, 450–451
 - light-guide, 322
 - losses, 361–364
 - in nanowires, 556
 - in photonic crystal, 549
 - in plasmonic, 555
 - measurement, 363
 - mirror, 323–326
 - mode number, 324, 330
 - mode velocity, 333, 348–350
 - modeling, 355–356
 - modulators, 445–448
 - multimode interference coupler, 368
 - near-field, 360–361
 - photonic crystal, 354, 548–549
 - plasmonic, 555–557
 - polymer, 351
 - propagating and decaying solutions, 338–339
 - propagation constant, 330–331, 335, 340
 - proximity couplers, 366–368
 - PVC/BK7, 343, 348, 358
 - ray-optic model, 322–337
 - ring resonator, 369, 384, 387, 390, 533, 538
 - $\text{Si}_3\text{N}_4/\text{BK7}$, 335
 - $\text{SiO}_2/\text{Si}_3\text{N}_4$, 332, 333, 335, 340, 342, 344, 351
 - slab, 323, 326, 337, 541
 - slot, 354
 - solving, 332–337
 - stripe, 350
 - symmetric, 328–329
 - Y-junctions, 365–366
- wavelength, 26
 - conversion to frequency, 35

- in a medium, 31
- visible range, 35, 293
- wavelength-division multiplexing, 380, 381
- WDM, *see* wavelength division multiplexing
- whispering gallery modes, 427
- Witelo, 7
- Wollaston, William, 6
- Wood's anomalies, 294
- Wood, Robert W., 11, 294

- x-rays, 34, 141, 170, 304, 483

- YAG (yttrium aluminum garnet), 97, 410
- YIG (yttrium iron garnet), 97, 450
- Young
 - double-slit experiment, 274
 - equation, 504
 - interferometer, *see* interferometers
- Young, Thomas, 5, 8, 136, 274, 504
- YVO₄ (yttrium aluminum vanadate), 416

- Zehnder, Ludwig, 141
- ZEMAX, 220–221, 243, 246–247
- Zernike polynomials, 43–46, 191, 474
- Zernike, Frits, 43
- Zerodur, 86
- zig-zags, 331, 349
- ZnS, 413
- ZnSe, 90
- zone plates, *see* Fresnel zone plates
- ZrF₄, 373
- Zurich, 242, 431